

ISSN 1913-1844 (Print)
ISSN 1913-1852 (Online)

MODERN APPLIED SCIENCE

Vol. 8, No. 5 October 2014



CANADIAN CENTER OF SCIENCE AND EDUCATION

Editorial Board

Editor-in-Chief

Salam Al-Maliky, Ohio University, United States

Associate Editors

Carlos Bazan, San Diego State University, United States

Carolina Font Palma, University of Manchester, United Kingdom

Jill Smith, University of York, United Kingdom

Jin Zhang, University of California, United States

Editorial Assistant

Penny Han, Canadian Center of Science and Education, Canada

Editorial Board Members

Abdolmajid Maskooki	Julien Wist	Pauli A. A. Garcia
Afonso Severino Regateiro Francisco	Julio Javier Castillo	Peter Kusch
Ahmad Mujahid Ahmad Zaidi	Junjie Lu	Prabir Daripa
Alessandro Filisetti	Kenier Castillo	Prabir Sarker
Alhussein Assiry	Krishna Chetry	Qadir Bux alias Imran Latif
Anna Grana [†]	Lazaros Mavromatidis	Qiang Bai
Antonio Camarena-Ibarrola	Levent Kurt	Rajiv Pandey
Antonio Comi	Liang Yu	Ricardo Ondarza Rovira
Arvin Emadi	Lim Hwee San	Robello Samuel
Ashraf Maher Abdel Ghaffar	Li Zhenze	Rodica Luca
Atul Kumar Singh	Luigi Di Sarno	Saeed Doroudiani
Bakytzhan Kallemov	Luo Kun	Samendra Sherchan
Bayram Kizilkaya	Mahmoud Zarei	Sevgihan Yildiz Bircan
Chen Haisheng	Marek Brabec	Shang Yilun
Cheng Zhang	Martin Martinez García	Shivanand R. Wasan
Chi-Lun Gillan Huang	Mazanetz Michael Philip	Skrynyk Oleg
Christakis Constantinides	Meenu Vikram	S. M. Abrarov
Cristina Damian	Miguel A. Carvajal Rodriguez	Stavros Kourkoulis
Daniela Popescu	Miguel Miranda	Stefanos Dailianis
Danielly Albuquerque	Milan Vukićević	Sushil Kumar Kansal
Dinesh Sathyamoorthy	Mingxin Li	Tahir Qaisrani
Dong Ling Tong	Mirza Hasanuzzaman	Takuya Yamano
Ekrem Kalkan	Mohammad Mehdi Rashidi	Tharek Rahman
Francesco Caruso	Mohammad Taghi Ahmadi	Tony (Panqing) Gao
Giovanni Angrisani	Mohamed A. Sharaf Eldean	Tugba Özacar
Gobithaasan R. U.	Mohammed Al-Abri	Umer Rashid
Godone Danilo	Mohd Afizi Mohd Shukran	Valentina Valentina
Guy L. Plourde	Mohd Hafizi Ahmad	Valter Aragao do Nascimento
Hamidreza Gohari Darabkhani	Monica Caniupán	Veera Gude
Hani Abdualh Alhadrami	Monica Carvalho	Venkatesh Govindarajan
Hui Zhang	Monika Gontarska	Verma Vijay Kumar
Ilki Kim	Muhammad Raza Naqvi	Vijay Karthik
Ioannis Gkigkitzis	Musa Mailah	Vinod Mishra
Jacek Leszczynski	Narayan Ramappa Birasal	Wenzhong Zhou
Jae Woo Lee	Nikolai Perov	Yili Huo
J. Eric Jensen	Övünç Öztürk	Yu Dong
Jiantao Guo	Partha Gangopadhyay	Yuriy Gorbachev
José Ignacio Calvo	Paul William Hyland	

Contents

Effect of Holding Time on the Diffusion Behavior at Interface of Dissimilar Metals Joint between Aluminum and Carbon Steel Joint Using Element Promoter	1
<i>A. J. Saad, Triyono Triyono, A. Supriyanto, N. Muhayat & Z. Yuliadi</i>	
Work Attitude among Malaysian Academicians in the Public Universities: A Social Network Analysis	9
<i>Norhaidah Mohd Asrah, Maman Abdurachman Djauhari & Ebi Shahrin Suleiman</i>	
Efficiency Analysis of Low Power Class-E Power Amplifier	19
<i>Mousa Yousefi, Ziaadin Daie Koozehkanani, Jafar Sobhi, Hamid Jangi & Nasser Nasirezadeh</i>	
BICEP2 and the Gravitino Mass: The Questionable Result	30
<i>AC Tahan</i>	
The Economic Value of Laem Phak Bia Mangrove Ecosystem Services in Phetchaburi Province, Thailand	36
<i>Sitthinan Wiwatthanapornchai, Chuchee Piputsitee & Samakkee Boonyawat</i>	
Reliability Worth Assessment in Composite Power System Planning and Operation	45
<i>Ali Khraiwish Dalabeeh & Eng.Hani Hasan Saleh Al-Hajbi</i>	
Interference in Wireless Networks: Causes, Analyses and Practical Mitigating Techniques	56
<i>A. Hameed & Ali Oudah</i>	
Equality of Google Scholar with Web of Science Citations: Case of Malaysian Engineering Highly Cited Papers	63
<i>Nader Ale Ebrahim, Hadi Salehi, Mohamed Amin Embi, Mahmoud Danaee Marjan Mohammadjafari, Azam Zavvari, Masoud Shakiba & Masoomeh Shahbazi-Moghadam</i>	
Ratio Estimators Using Coefficient of Variation and Coefficient of Correlation	70
<i>Prayad Sangngam</i>	
Impact of Landuse Change on River Floodplain Using Public Domain Hydraulic Model	80
<i>Abolghasem Akbari, Golamali Mozafari, Mohsen Fanodi & Maliheh Sadat Hemmesy</i>	
Detection of Cassava Leaves in Multi-Temporally Acquired Digital Images of a Cassava Field Under Different Brightness Levels by Simultaneous Binarization of the Images Based on Indices of Redness/Greenness	87
<i>Mallika Srisutham, Ryoichi Doi, Anan Polthanee & Masaru Mizoguchi</i>	
Brain Activation and Psychophysiologic Interaction in Association with a Phonological Working Memory Task	97
<i>Ahmad Nazlim Yusoff, Hanani Abdul Manan, Siti Zamratol-Mai Sarah Mukari, Khairiah Abdul Hamid & Elizabeth A. Franz</i>	
Structural Dynamic Study of Roof Waterproofing Materials	115
<i>Alim Feizrakhmanovich Kemalov, Ruslan Alimovich Kemalov, Dinar Zinnurovich Valiev & Ilmira Maratovna Abdrafikova</i>	
Automatic Identification of Ficus deltoidea Jack (Moraceae) Varieties Based on Leaf	121
<i>A Fakhri A Nasir, M Nordin A Rahman, Nashriyah Mat & Rasid Mamat</i>	
The Extent of Sunlight Penetration Performance on Traditional Style's Apartment Façade in Putrajaya, Malaysia	132
<i>Ahmad Sanusi Hassan & Yasser Arab</i>	

Contents

Parameters Estimation of LFM CW Signals Based on Periodic Fractional Fourier Transform <i>Ying Liu, Dianren Chen & Lei Chen</i>	143
Logic of Interval Uncertainty <i>Evgeny A. Kuzmin</i>	152
Economic Growth and Internet Usage Impact on Publication Productivity among ASEAN's and World's Best Universities <i>Hossein Gholizadeh, Hadi Salehi, Mohamed Amin Embi, Mahmoud Danaee, Ali Ordi & Farid Habibi Tanha, Nader Ale Ebrahim & Noor Azuan Abu Osman</i>	169
Justification for Remote Control of Construction and Road-Making Machines <i>Nadegda Savelyevna Sevryugina, Eugene Alexandrovich Volkov & Eugene Pavlovich Litovchenko</i>	179
Design of PID Filter Controller with Genetic Algorithm for MIMO System in Modern Power Generation <i>Anitha Mary. X, L. Sivakumar & J. Jayakumar</i>	186
About the Method of Analysis of Economic Correlations by Differentiation of Spline Models <i>Ruslan Hizrailevich Ilyasov</i>	197
An Auto-Landing Solution for a Drop Test RLV Demonstrator <i>Peng Yong-Tao, Wang Yue-Ping, Wei Wen-Ling & Wang Xiao-Ting</i>	204
Constants of Metal Rubber Material <i>Alexander Mikhailovich Ulanov</i>	216
Vision-Based Row Detection Algorithms Evaluation for Weeding Cultivator Guidance in Lentil <i>Hossein Behfar, HamidReza Ghasemzadeh, Ali Rostami, MirHadi Seyedarabi & Mohammad Moghaddam</i>	224
H.M. The King's Royally Initiated LERD Project on Community Wastewater Treatment through Small Wetlands and Oxidation Pond in Phetchaburi, Thailand <i>Kasem Chunkao, Wit Tarnchalanukit, Paiboon Prabuddham, Onanong Phewnil, Surat Bualert, Kittichai Duangmal, Thanit Pattamapitooon & Chatri Nimpee</i>	233
ADTC in Wind Turbines with Partial State and Wind Speed Estimation <i>Ali Oudah, Izzeldin I. Mohd & A. Hameed</i>	247
Hardware-Software Complex for Laser Scanning in Color Sorting Machines <i>Elnur Megrallyevich Babishov, Gennady Vladimirovich Pakhomov, Vladimir Alekseevich Shulgin, Evgeny Yur'evich Buslov & Dmitry Anatol'evich Minakov</i>	262
Proper Insulated Materials for Temperature Accumulation in Box Technology to Catalyze the Organic Digestion Processing on Community Garbage Disposal <i>Chulabut Chanthasoon & Kasem Chunkao</i>	272
Assessment of Toilet's Indoor Air Quality in Relation to Asthmatic People <i>Ismail Abdul Rahman, Jouvan Chandra Pratama Putra & Ade Asmi</i>	289
Reviewer Acknowledgements for Modern Applied Science, Vol. 8, No. 5 <i>Sunny Lee</i>	296

Effect of Holding Time on the Diffusion Behavior at Interface of Dissimilar Metals Joint between Aluminum and Carbon Steel Joint Using Element Promoter

A. J. Saad¹, Triyono¹, A. Supriyanto², N. Muhayat¹ & Z. Yuliadi³

¹ Mechanical Engineering Department, Sebelas Maret University, Indonesia

² Physics Department, Sebelas Maret University, Indonesia

³ Head of Technology Division, PT. PAL Indonesia (Indonesia Ship Industry)

Correspondence: Triyono, Mechanical Engineering Department, Sebelas Maret University, Surakarta, Indonesia.
Tel: 62-271-632-163. E-mail: triyonomesin@uns.ac.id

Received: June 20, 2014 Accepted: July 15, 2014 Online Published: August 3, 2014

doi:10.5539/mas.v8n5p1

URL: <http://dx.doi.org/10.5539/mas.v8n5p1>

Abstract

Holding time is used for optimizing the bond diffusion between aluminum Al and Carbon steel SS400. The objective of this research was to investigate the effects of holding time on the interface reactions of diffusion welding between aluminum and carbon steel. Holding time variations of 10, 15, 30 and 45 minutes were applied at 950°C using mixture of Cu and Fe powder as elements promoter. Single lap joint configuration was performed in vacuum furnace to join the dissimilar materials which allowed bonding diffusion. Microstructure was examined on the same test piece. It was found that during diffusion process at 950°C, the interfacial zone between aluminum and carbon steel substrate features intermetallic layers. The intermetallic thickness increased with increasing the holding time. Crack or incomplete bonding appeared on the specimens with holding time up to 30 minutes and didn't appear on the specimens with holding time of 45 minutes. Cu rich-element promoter made diffusion penetrated deeper than Fe rich-element promoter in the same holding time. Macrostructure, microstructure and SEM examinations revealed that Al-steel joint had the best result with element promoter content of 60/40 % at 45 minutes holding time. There was no interlayer gap at this specimen. Additionally, from mapping view it can be suggested that in terms of poor interface bonding, Cu molecules were located just around the interface area, on the other hand, in case of strong interface bonding, Cu molecules are scattered throughout the specimen. In fact, the position of Cu molecules can be used as a promising marker for the detection of quality of diffusion joint.

Keywords: aluminum, diffusion bounding, interface, steel, holding time, element promoter

1. Introduction

Aluminum is the most abundant metal and the third most abundant element in the earth's crust, after oxygen and silicon. It makes up about 8% by weight of the earth's solid surface. Aluminum is also chemically reactive to occur naturally as free metal (Kobayashi & Yakou, 2002). There are many applications of Aluminum in our daily life, such as construction machinery, aircraft construction, ship construction, home furnishing and electronics component. For the vehicle industry, Aluminum has established a worldwide position because of its advantages over the other competitive materials.

Because of its lightweight Aluminum provides exceptional unit strengths (strength/density ratio), high corrosion resistance, low maintenance costs, good temperature resistance, ductile which is an intrinsic ability of a material to undergo a certain amount of plastic deformation without the occurrence of macroscopic cracks (Maea et al., 2008). It can be easily joined by all commercial processes such as welding, brazing, or soldering. Aluminum can be easily formed by all common processes, including extrusion (a major advantage) and can be recycled. Beside of the advantages, there are some disadvantages of Aluminum. First, Aluminums is relatively more expensive than steel. Secondly, Aluminum sheets are more difficult to stamp into car body parts. Thirdly, Aluminum is difficult to weld than steel. Lastly, Aluminum doesn't have the strengths levels as steel (Patel et al., 2013).

The effectiveness of Aluminum choice depends on the nature of the aluminum type used. Properties of low work hardening alloy Al 6061–T 6511, a very high work hardening annealed material 1100 Al, and both Al alloys (Al6061-T6511 and annealed 1100 Al) were studied and reported in subsequent researches (Khan et al., 2009; 2010a; 2010b). It was reported that the subsequent expansion and positive cross-effect was observed in the a high work hardening alloy annealed 1100 Al as compared to low work hardening alloy Al 6061–T 6511 where contraction and negative cross-effect with finite deformation yield surfaces was showed. According to Khan et al. (2010a), the subsequent yield surfaces for annealed Al 1100, the rate of kinematic hardening, Young's and shears moduli decreased and isotropic hardening increased with finite plastic deformation, which showed positive cross-effect for linear, bilinear and non-linear uploading.

Steel is an alloy of iron and a small amount of carbon. Carbon is the primary alloy element, which accounts between 0.002% and 2.1% by weight in steel. There is no stiffness advantage in using aluminum over steel. Basically, strength refers to the maximum load that a material can be subjected to without yielding, whereas stiffness refers to how much a material bends when a load is applied (Ajit et al., 2013). Stiffness is quantified by a parameter called Modulus of Elasticity. Steel's modulus is about 3 times stronger than Aluminum but steel's weight is about 3 times heavier than aluminum.

Aluminum welding with other metals is very difficult because there are problems like different melting points, physical natures and other intermetallic differences. Kim et al. (2006) presented that intermetallic diffusion in interface Al and steel is formed. The Intermetallic Compound For motion (IMC) is rapidly developed and grows between the steel and the melt. Only aluminum diffuses into the steel substrate without the dissolution of iron at the interface of the steel-intermetallic compound. This result was supported by Qiu et al. (2009).

According to Kim et al. (2006) the primary diffusing species of the hot dip aluminizing process in their study was aluminum. Al coating on the steel and the short dipping time prevent the iron from dissolving into the aluminum melt. The IMC is confirmed to be Fe_2SiAl_8 with a hexagonal unit cell (space group P63/mmc). Sun and khalel (2007) showed that the intermetallic phases $FeAl_2$ and Fe_2Al_5 were the most dominant phases that could be observed, that they were formed sequentially, in contrast to intermetallic, which formed synchronously in bulk materials. A good diffusion interface cannot be formed if the heating temperature is too low, since extent of diffusion is not sufficient even though the holding time is longer and the pressure is larger. But, if the heating temperature is too high, the grains will grow up seriously and the diffusion transition zone can become wider, which will adversely affect the performance of the diffusion bonding joint (Yajiang et al., 2005).

Furthermore, the usage of filler material such as Cu- Mg enhance the hardness of interface bonding between two dissimilar metal following by perfect holding time (Mahendran et al., 2009). Therefore, this paper aims to determine the effect of holding time on the diffusion characteristics of the joint with the 6061 aluminum and Carbon steel filler using Cu and Fe.

2. Method

2.1 Materials and Diffusion Processes

The used material in this research were aluminum AL 6061 and carbon steel SS400. The chemical composition of the base metals is shown in table 1 and 2.

Table 1. Chemical Composition of Al 6061 according to Dinaharan et al. (2012)

Mg	Si	Fe	Mn	Cu	Cr	Zn	Ni	Ti	Al
0.95	0.54	0.22	0.13	0.17	0.09	0.08	0.02	0.01	Bal.

Table 2. Chemical composition of SS400 as reported by Dinaharan et al. (2012)

Si	Mn	P	S	Ni	Cr	Fe
0.2	0.09	0.53	0.01	0.04	0.03	Bal.

Joined specimens were cut with length of 30 mm and width of 30 mm (Fig. 1). They were joined in single lap joint type and its procedure was based on AWS D8.9. Elements promoter were inserted at interface between aluminum and steel. They were mixture of Al-Fe powder. Two composition were used in this research namely

80%Fe-20%Cu and 60%Fe-40%Cu. Diffusion welding of aluminum to steel used a vacuum furnace at 950°C. The holding time variations were 10, 15, 30 and 45 minutes. Table 3 shows the schedule of diffusion process.

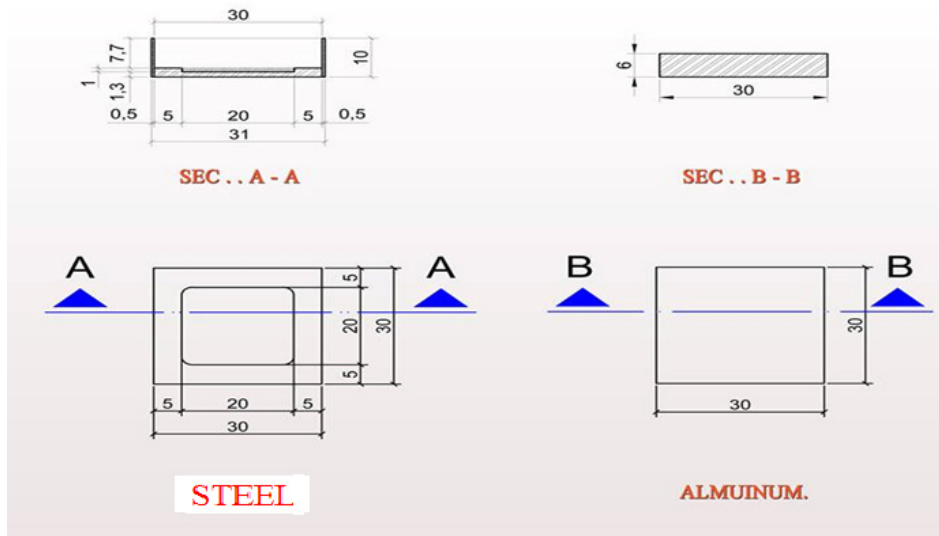


Figure 1. Dimensions of test specimen size in mm

Table 3. Research parameters variations

Run No.	Mixture of Elements Promoter	Holding Time (minutes)
1	60%Fe - 40%Cu	10
2	60%Fe - 40%Cu	15
3	60%Fe - 40%Cu	30
4	60%Fe - 40%Cu	45
5	80%Fe - 20%Cu	10
6	80%Fe - 20%Cu	15
7	80%Fe - 20%Cu	30
8	80%Fe - 20%Cu	45

2.2 Joint Characterizations

The transverse sections of weld passing through the weld interface were prepared by standard metallographic procedure according to ASTM E3-01. The microstructure of mild steel was revealed by using HNO₃ 2.5% solution while that of aluminum was revealed by using HF 5%. Microstructure of interface reaction investigations were carried out using an optical microscopy and SEM.

3. Results

Diffusion aluminium atom to steel and ferro atom to aluminum generally occurred at interface of aluminum-steel during diffusion process. Holding time significantly affect the quality of Al-Fe diffusion. Holding time of 10 minutes, 15 minutes, 30 minutes and 45 minutes gave different result of the diffusion as seen in Figure 2.

Holding time	Fe (80%), Cu (20%)	Fe (60%), Cu (40%)
	10 minutes	
15 minutes		
30 minutes		
45 minutes		

Figure 2. Macrostructure of interface zone

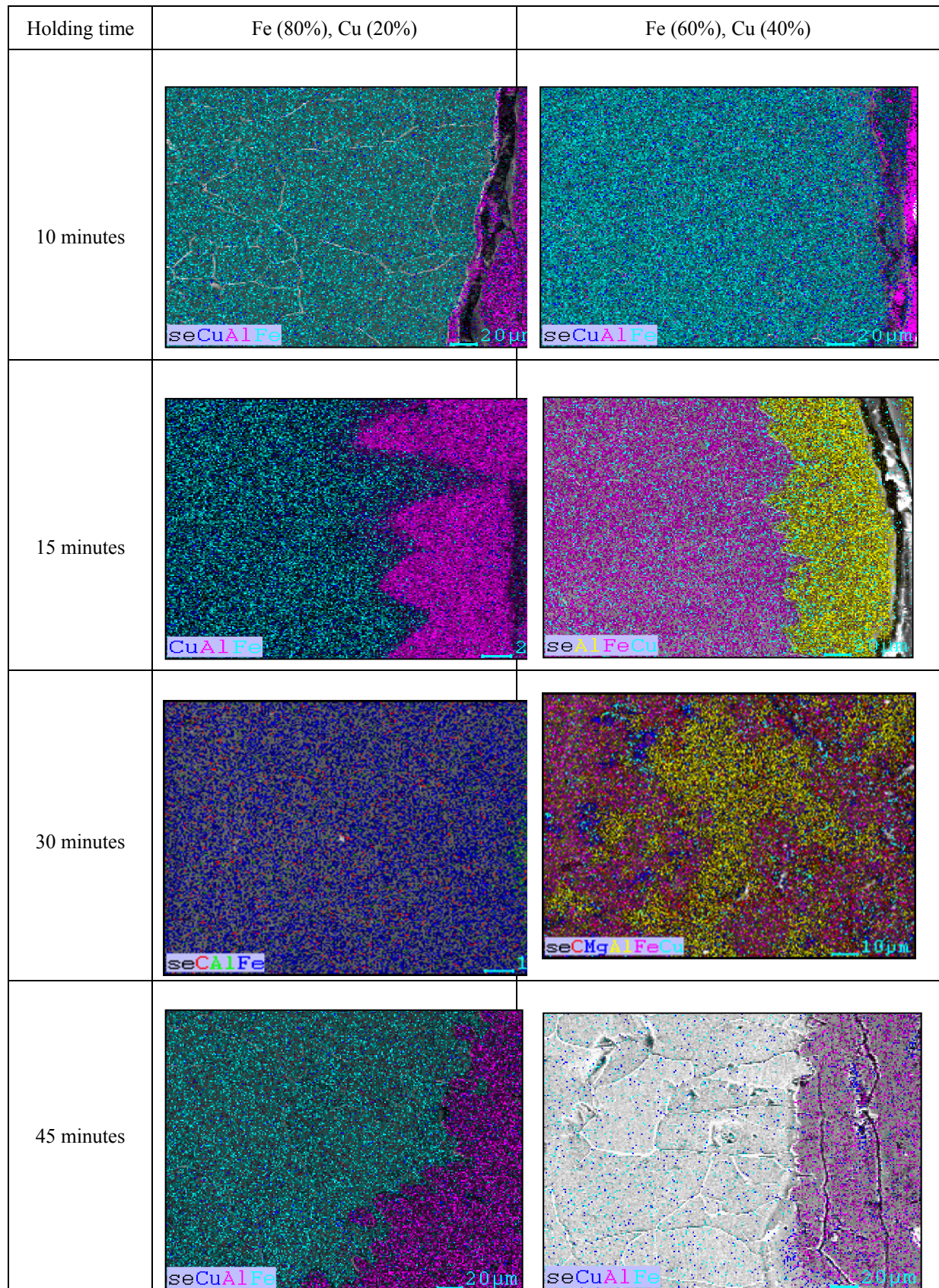
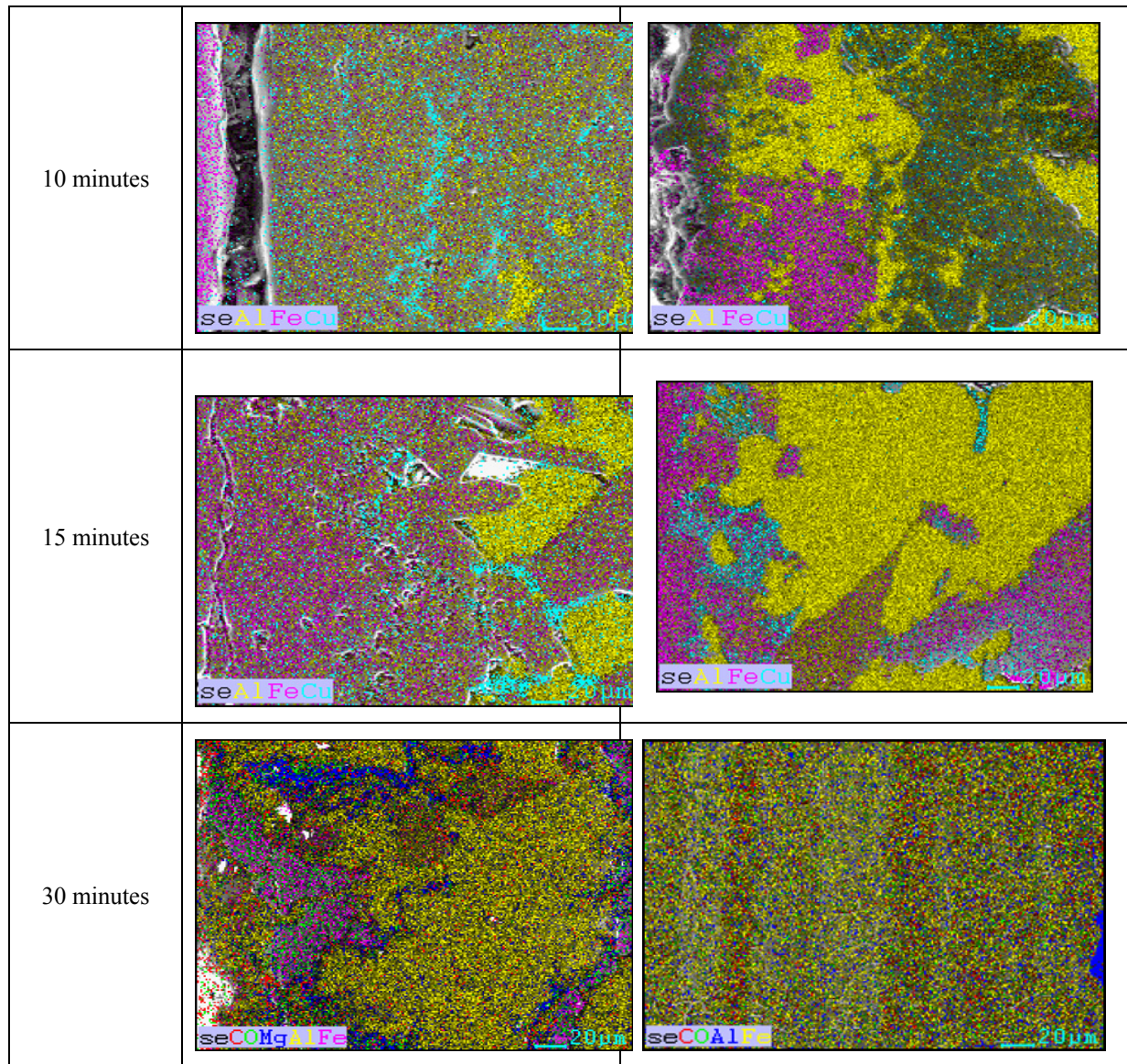


Figure 3. Elements mapping at steel side

At holding time in range of 10-30 minutes, it was found that in the interface there are incomplete bondings and

cracks shown in Figure 2. There were visible cracks on the surface area with small interface layer in specimens with holding time of 10 minutes both the joints with element promoter composition of 60%Fe-40%Cu and 80%Fe-20%Cu. It means that diffusion of aluminum into steel did not happen significantly with 10 minutes holding time. At 15 and 30 minutes holding time, the interface thickness increases with smaller gap and cracks. It was very clear in the figure that the solubility of aluminum increased with increasing of holding time where more aluminum dissolved into steel. The best results were found in specimens with composition of 60%Fe-40%Cu at 45 minutes holding time. In this case, diffused surface area increases because the diffusion of the composition occurs between aluminum and steel which interlock each other resulting in thick interface area and negligible cracking. Additionally, continous intermetallic layer appeared in this specimen, while it didn't appear in other specimens.



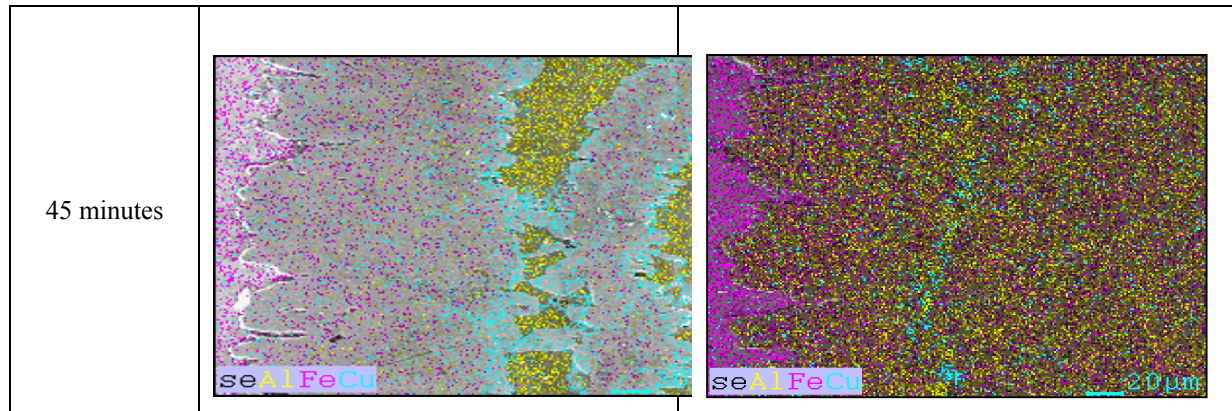


Figure 4. Elements mapping at aluminum side

Generally, thickness increased with increasing holding time. In case holding time up to 30 minutes, small gap and cracks appeared as seen in macrostructure. The thickness of intermetallic layers increases linearly with increasing holding time but until holding time of 45 minutes. As holding time increased, the macro of the intermetallic layer displayed an increasing amount of intercellular primary aluminum solidification. Similar studies have also been carried out for the welding of aluminum and steel friction stir welding by Watanabe et al. (2006) and Rafeng et al. (2009). The main results obtained from the studied on the microscopic structure, the strength of steel with aluminum, welding resistance were as follows. The maximum tensile strength of the joint was about 86% of that of the aluminum alloy base metal. A small amount of intermetallic compounds was formed at the upper part of the steel/aluminum interface, while no intermetallic compounds were observed in the middle and bottom parts of the interface.

Figure 3 and Figure 4 show the influence of holding time on inter-metallic gap between aluminum and carbon steel at two different Fe-Cu combinations. The minimum crack between aluminum and carbon steel was found at combination with holding time 45 minutes whereas significant inter-metallic gap between aluminum and carbon steel was obtained both at 80%Fe-20%Cu and 60%Fe-40%Cu with holding time 10-30 minutes.

Additionally, from the mapping point of view at 20 μm it can be observed that Cu molecules are confined only the interface area and slightly scattered outside the inter-confront area in specimens with holding time up to 30 minutes both at 80%Fe-20%Cu and 60%Fe-40%Cu. On the other hand, Cu molecules were scattered throughout the specimen with holding time of 45 minutes. It means Cu molecules play a vital role of identifying best Al-Fe interface combination i.e. if the Cu molecules are located just around the inter-confront area that can be considered as a bad interface, whereas if the Cu molecules were scattered throughout interface zone that can be considered as a good interface

4. Conclusion

During diffusion process at 950°C, the interfacial zone between aluminum and carbon steel substrate features intermetallic layers. The intermetallic thickness increased with increasing the holding time. Crack or incomplete bonding appeared on the specimens with holding time up to 30 minutes and didn't appear on the specimens with holding time of 45 minutes. Cu rich-element promoter made diffusion penetrated deeper than Fe rich-element promoter in the same holding time. Macrostructure, microstructure and SEM examinations revealed that Al-steel joint had the best result with element promoter content of 60/40 % at 45 minutes holding time. There was no interlayer gap at this specimen. Additionally, from mapping view it can be suggested that in terms of poor interface bonding, Cu molecules were located just around the interface area, on the other hand, in case of strong interface bonding, Cu molecules are scattered throughout the specimen.

Acknowledgments

The authors would like to express their sincere gratitude for Ministry of Education and Culture of Indonesian Republic through MP3EI 2013 grant.

Reference

Kelkar, A. D., Mohan, R., Ghazizadeh, M., & Kimbro, E. (2013). Progressive Failure Analysis of Hybrid Composite Laminates Under Static Tensile Loading. *Journal of International Scientific Publications:*

- Materials, Methods & Technology*, 7(1). Retrieved from <http://www.scientific-publications.net/download/materials-methods-and-technologies-2013-1.pdf>
- Dinakaran, K., Kalaiselvan, Vijay, S. J., & Raja, P. (2012). Effect of Material Location and Tool Rotational Speed on Microstructure and Tensile Strength of Dissimilar Friction Stir Welded Aluminium Alloy. *Archives of Civil and Mechanical Engineering*, 12, 446-456. <http://dx.doi.org/10.1016/j.acme.2012.08.002>
- Khan, A. S., Kazmi, R., Pandey, A., & Stoughton, T. (2009). Evolution of subsequent yield surfaces and elastic constants with finite plastic deformation. Part-I: A very low work hardening aluminum alloy (Al6061-T6511). *International Journal of Plasticity*, 25(9), 1611-1625. <http://dx.doi.org/10.1016/j.ijplas.2008.07.003>
- Khan, A. S., Pandey, A., & Stoughton, T. (2010). Evolution of subsequent yield surfaces and elastic constants with finite plastic deformation. Part II: A very high work hardening aluminum alloy (annealed 1100 Al). *International Journal of Plasticity*, 26(10), 1421-1431. <http://dx.doi.org/10.1016/j.ijplas.2009.07.008>
- Khan, A. S., Pandey, A., & Stoughton, T. (2010). Evolution of subsequent yield surfaces and elastic constants with finite plastic deformation. Part III: Yield surface in tension-tension stress space (Al 6061-T 6511 and annealed 1100 Al). *International Journal of Plasticity*, 26(10), 1432-1441. <http://dx.doi.org/10.1016/j.ijplas.2009.07.007>
- Kim, K. H., van Daele, B., Van Tendeloo, G., & Yoon, J. K. (2006). Observations of intermetallic compound formation of hot dip aluminized steel. In *Materials science forum* (Vol. 519, pp. 1871-1876). Retrieved from <http://www.scientific.net/MSF.519-521>
- Kobayashi, S., & Yakou, T. (2002). Control of intermetallic compound layers at interface between steel and aluminum by diffusion-treatment. *Materials Science and Engineering: A*, 338(1), 44-53. [http://dx.doi.org/10.1016/S0921-5093\(02\)00053-9](http://dx.doi.org/10.1016/S0921-5093(02)00053-9)
- Maea, H., Tengb, X., Baib, Y. & Wierzbickib, T. (2008). Comparison of ductile fracture properties of aluminum castings: Sand mold vs. metal mold. *International Journal of Solids and Structures*, 45(5), 1430-1444. <http://dx.doi.org/10.1016/j.ijsolstr.2007.10.016>
- Mahendran, G., Balasubramanian, V., & Senthilvelan, T. (2010). Influences of diffusion bonding process parameters on bond characteristics of Mg-Cu dissimilar joints. *Transactions of Nonferrous Metals Society of China*, 20(6), 997-1005. [http://dx.doi.org/10.1016/S1003-6326\(09\)60248-X](http://dx.doi.org/10.1016/S1003-6326(09)60248-X)
- Patel, A. S., Neerajd S., & Bhavin, K. (2013). *Asktom.oracle.com* (Volume 1, Issue 11).
- Qiu, R., Satonaka, S., & Iwamoto, C. (2009). Effect of interfacial reaction layer continuity on the tensile strength of resistance spot welded joints between aluminum alloy and steels. *Materials & Design*, 30(9), 3686-3689. <http://dx.doi.org/10.1016/j.matdes.2009.02.012>
- Sun, X., & Khaleel, M. A. (2007). Dynamic strength evaluations for self-piercing rivets and resistance spot welds joining similar and dissimilar metals. *International journal of impact engineering*, 34(10), 1668-1682. <http://dx.doi.org/10.1016/j.ijimpeng.2006.09.092>
- Watanabe, T., Takayama, H., & Yanagisawa, A. (2006). Joining of aluminum alloy to steel by friction stir welding. *Materials Processing Technology*, 178, 342-349.
- Yajiang, L., Wang J., Yin, Y. S., & Ma, H. J. B. (2005). Diffusivity of Al and Fe near the diffusion bonding interface of Fe3Al with low carbon steel. *Mater. Sci.*, 28(1), 69-74. <http://dx.doi.org/10.1016/j.jmatprotec.2006.04.117>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

Work Attitude among Malaysian Academicians in the Public Universities: A Social Network Analysis

Norhaidah Mohd Asrah^{1,2}, Maman Abdurachman Djauhari³ & Ebi Shahrin Suleiman⁴

¹ Department of Mathematics & Statistics, Universiti Tun Hussein Onn Malaysia, Malaysia

² Mathematical Sciences Department, Universiti Teknologi Malaysia, Malaysia

³ Center for Research, Consultation and Training in Statistical Analysis, Universitas Pasundan, Bandung, Indonesia

⁴ Faculty of Management and Human Resource Development, Universiti Teknologi Malaysia, Malaysia

Correspondence: Norhaidah Mohd Asrah, Department of Mathematics & Statistics, Universiti Tun Hussein Onn Malaysia, Malaysia. E-mail: norhaida@uthm.edu.my

Received: April 1, 2014

Accepted: April 21, 2014

Online Published: August 6, 2014

doi:10.5539/mas.v8n5p9

URL: <http://dx.doi.org/10.5539/mas.v8n5p9>

Abstract

This study dealt with a social network analysis approach to comprehend the work attitude amongst academicians in the Malaysian public universities. This work attitude presented the psychological attachment between the employee and the organization. The organizational commitment and workplace spirituality amongst the academicians were highlighted here. A total of 40 factors were found to represent four groups of workplace spirituality and organizational commitment. The similarity amongst the factors was measured with two different kinds of associations. The best measure of association, which was the Tschuprow's measure of association, showed better results than the other measure in measuring the correlation amongst the factors. The connections and relationships amongst the factors were studied by using minimum spanning trees (MST). The interpretation of the MST was conducted by using the overall centrality measure.

Keywords: spearman correlation coefficient, Tschuprow's measure of association, minimum spanning tree (MST), centrality measures

1. Introduction

The study about commitment in an organization has been a subject of interest in the research field for almost four decades. It has an impact on individual performance and also the effectiveness of the organization (Allen & Meyer, 1996). Organizational commitment can best be described as the psychological attachment between an employee and an organization. It is the most important work attitude in the study of organizational and management behaviours. Commitment in the workplace can take in various forms; it also has the potential to influence the organizational effectiveness and employee behaviour (Meyer & Herscovitch, 2001). Many studies have related organizational commitment to other factors like job satisfaction (Darwish, 2002; Therese & Steve, 2006), workplace spirituality (Rego & Cunha, 2008), human resource management (Smeenk, Eisinga, Teelken & Doorewaard, 2006), and research and development (Chang & Choi, 2007).

One of the most important organizations in all nations is the education (Noordin, & Jusoff, 2009). In Malaysia, the education sector has become one of the most essential factors to transform Malaysia into a high-income nation. In 2010, Malaysia has implemented the economic transformation program (ETP). With regards to this, the tertiary education development has been identified as one of the most important strategies to transform Malaysia from a middle-income nation to a high-income nation. Hence, the demands of the tertiary education systems have increased. Due to this fact, Malaysia has made a series of dramatic changes, such as upgrading the university colleges to full public university status, increasing the number of new private universities, and also installing a number of foreign universities in the country. Consequently, the role of academic staff has become more challenging and demanding. The university management needs to take into serious consideration regarding this challenging matter in the academic affairs, as well as in research and administration (Daud, 2012).

This new challenge requires high commitment from the public universities staffs. For example, the organisation of international programmes and research competitions improves the global ranking of these programmes. The

performances of organizational commitment on productivity, efficiency, and quality have become important issues here. It is crucial for the management of the public universities in Malaysia to understand the behaviour and work attitude of academicians. The management also needs to find the solutions to produce high quality graduates with vast knowledge and excellent skills to fulfil the job demands out there.

One of the most important factors related to the organizational commitment is workplace spirituality. Organizations are encouraged to develop this factor because humanistic work environment creates a win-win situation for both employees and the organizations.

In this paper, we used the survey method to collect data related to organizational commitment and workplace spirituality among academicians in the Malaysian public universities. The performances of every factor were studied by using the social network analysis approach to understand the relationships between the factors and to identify the most influential factor(s). Since the data were on an ordinal scale, two kinds of association measures were utilised to measure the similarity among the factors, namely Tschuprow's measure of association and the Spearman correlation coefficient. This was to justify the better measurement between the both. As a result, the Tschuprow's measure of association proved to measure better. Besides, by studying the relationships amongst the factors and identifying the most influential factor(s), the public university management can create a better strategic plan for further development.

This paper is organized as follows. The next section describes the literature review, data preparation, and the methodology of social network. This is followed by the discussion pertaining to the results for both coefficients. The paper ends with a conclusion.

1.1 Literature Review

Workplace spirituality is not about religious beliefs, but it is about being energetically at work for people who perceived themselves as spirited beings. The real spirituality is about people who share and experience the common attraction, attachment, and together with the other members within their work unit and organization as a whole. Hence, spirituality can be considered as the valuing spirits and inner life of the employee (Harrington, Preziosi & Gooden, 2002). The perceptions of employees about workplace spirituality help to explain their level of organizational commitment.

There are three quite distinct forms of psychological linkage between employees and their organizations, namely affective commitment, continuance commitment, and normative commitment (Meyer & Herscovitch, 2001). Affective commitment is more to identify with, involvement in, and emotional attachment to the organization. This kind of employees will remain with the organization because they want to do so. This commitment explains what the employees want in terms of desire to stay in the organization. The core of this commitment is an affective tendency, including their desires, wishes, feelings, and etc (Gonza'lez & Guille'n, 2008). Affective commitment explains the bond to the organization as an affective attachment. This attachment includes the feelings like affection, warmth, belongingness, loyalty, fondness, pleasure, and so on (Jaros, Jermier, Koehler & Sincich, 1993). An employee who perceives higher commitment affectively will likely to remain with their respective organization because they want to (Choong et al., 2011).

While continuance commitment refers to commitment based on the employees' recognition of the costs associated with leaving the organization. This type of employees will stay with the organization because they have to do so. When the employee perceives the cost of leaving is more than the benefits when they continue to remain in the organization, then it is better for the employee to stay with the organization rather than to leave (Choong et al., 2011). Lastly, normative commitment defines the commitment based on a sense of obligation to the organization. These employees will remain because they feel they ought to do so. The decision to stay or leave is based on the employee's belief of the right thing to do (Choong et al., 2011). The employee will stay and remain in the organization if they feel that he/she should be loyal to his/her respective employer.

2. Method

2.1 Data Collection

The data for this research were collected through a survey amongst 204 academic staffs from selected Malaysian public universities. The questionnaires, which were distributed to the respondents, were designed based on the literature study, for example, (Choong's et al., 2011, Yusoff's et al., 2012, and Suleiman's et al., 2012). It consisted of two components, namely workplace spirituality and organizational commitment. The first component consisted of five factors (team's sense of community, alignment between organizational and individual values, sense of contribution to the community, sense of enjoyment at work, and opportunities for the inner life), and three for the second component (affective commitment, continuance commitment, and normative

commitment). The total number of questions was 40.

The responses from the academic staff regarding their beliefs to workplace spirituality and organizational commitment were described by using the Likert-type format. This Likert scale with 5 categories; 1 (strongly disagree), 2 (disagree), 3 (neutral), 4 (agree), and 5 (strongly agree), allowed the academic staff to respond in different degrees to each factor which described the workplace spirituality and organizational commitment in the Malaysian public universities. The advantage of this format was that it allowed the academic staff to express the degree of their response to each factor rather than to a “yes” or “no” answer (Hayes, 2008).

This measurement scale is known as the ordinal scale. This scale is used to identify if the measurements are relevant. The numeric value used in the measurement is a means of arranging the elements being measured in order, from the smallest to the largest. The name ‘ordinal’ refers to ‘order’ of the elements on the basis of their relative size of the measurements (Conover, 1971).

In this study, two measurements were compared; the Tschuprow’s measure of association and the Spearman correlation coefficient. The better measurement was discovered for measuring the relationship between the factors in the Likert scale data. From the previous studies, the Spearman correlation coefficient had always been used to find the correlation between the data in ordinal data (Conover, 1971). However, as for the Likert scale, each scale is represented by a number, but this number does not represent the real number, as it only represents a category. Therefore, the Tschuprow’s measure of association was used to measure the correlation in an ordinal scale too.

2.2 Data Preparation and Analysis

In this section, the steps in social network analysis approach are briefly discussed to understand the relationships amongst the factors and to find the most influential factor(s). The Spearman’s rank coefficient and Pearson’s correlation coefficient were some of the several measures for computing the similarity between a pair of ranking vectors (Tan, Kumar & Srivastava, 2004). In this paper, the data from the survey were transformed into association matrix A of size 40×40 . The elements of the i -th row and j -th column in the association matrix were the measures of association from the Spearman correlation coefficient and the Tschuprow’s measure of association since the data were in the ordinal scale.

These association measures measured the similarity amongst the factors for ordinal data. In many parametric statistical methods, an interval scale of measurement is required. However, in the non-parametric methods, either the nominal or the ordinal scale is appropriate. In each scale of measurement, it has all the properties of the weaker measurement, and therefore, the statistical methods only require a weaker scale to be used against the stronger scales (Conover, 1971).

The Spearman correlation coefficient is one of the oldest and the best methods (Conover, 1971). This measurement has been always used as a test statistic to test the independence between two random variables (Spearman, 1904). Spearman correlation coefficient can best be described as a non-parametric rank statistics to measure the strength of the association between two variables. It uses the monotonic function to describe a relationship between two variables without making any assumption about the distribution of the variables (distribution free) (Hauke & Kossowski, 2011). Furthermore, the Spearman correlation coefficient does not require the variables to be measured on interval scale; it can be used for variables measured at the ordinal level. The measure of correlation by Spearman is expressed as ρ (rho),

$$\rho = 1 - \frac{6T}{n(n^2-1)} \quad (1)$$

where T represents the entire sum in the numerator,

$$T = \sum [R(X_i) - R(Y_i)]^2 \quad (2)$$

T is the ranked difference between two variates, and n is the number of measurements in each of the two variates in the correlation (Jerrold, 1972). $R(X_i)$ is the rank of X_i as compared with the other values of X , for $i = 1, 2, 3, \dots, n$. $R(X_i) = 1$ if X_i is the smallest value in X_1, X_2, \dots, X_n , $R(X_i) = 2$ if X_i is the second smallest and so on.

While $R(Y_i)$ is the rank of Y_i as compared with the other values of Y , for $i = 1, 2, 3, \dots, n$.

The other measurement is the Tschuprow’s measure of association. Tschuprow introduced the measure of association for nominal scale based on the χ^2 (chi square) value from the contingency table (Svante, & Jan, 1978). It is defined as below:

$$R_T = \sqrt{\frac{\chi^2}{N\sqrt{(r-1)(c-1)}}} \quad (3)$$

where

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (4)$$

χ^2 is the value computed from the contingency table, N is the number of units, r is the number of row, and c is the number of columns in the contingency table. The term O_{ij} represents the observed number in cell (i,j) , while E_{ij} represents the expected number of observations in cell (i,j) .

The Tschuprow's measure of association is less well known, but it has some possible theoretical advantages (Bergsma, 2012). The values for Tschuprow's measure of association are between 0 and 1, just like other association measures. The degree of independence between the values of two categorical variables is greater when the values of the association are nearer (Tomizawa, 1994). Tschuprow's measure of association usually uses square tables; row marginal which are identical with the column marginal, and is very seldom used for measuring the association between the two variables. When the contingency table is in square form, it can achieve its maximum value. The usual estimators of these coefficients are simple functions of the Pearson chi-square statistic.

All possible comparisons among the pairs of variables produce a square and a symmetrical association matrix A of size 40×40 . The values in this matrix are the comparison between two variables, a_{ij} is the comparison measure between variables i and variables j . The ecological association matrices are usually symmetric since $a_{ij} = a_{ji}$. The values on the diagonal are compared with the variables. The diagonal value equals to 1 according to the Spearman correlation coefficient and Tschuprow's measure of association.

Besides, the dissimilarity matrix or distance matrix was used to determine the minimal spanning tree (MST). The association matrix cannot be used as the dissimilarity matrix or distance matrix since it does not fulfil the three conditions to define a metric (Mantegna, 1999). The conditions are: (i) $d_{ij} = 0$ if and only if $i = j$, (ii) $d_{ij} = d_{ji}$, and (iii) $d_{ij} \leq d_{ik} + d_{kj}$. The association matrix was then, transformed into dissimilarity matrix, D , by using this formula:

$$d_{ij} = 1 - a_{ij} \quad (5)$$

for all $i, j = 1, 2, \dots, 40$.

From this dissimilarity matrix D , the network amongst the factors was analyzed using the MST. MST is constructed to visualize the important information contained in the network in D . MST is a concept in graph theory that connects weighted graph of n objects. It is a tree with $n-1$ edges that minimizes the sum of the edge distances. MST is built by linking every element in a set of n , together in a graph, characterized by a minimal distance between the nodes. The method used to construct MST by linking a set of n objects is known as Kruskal's algorithm (Mantegna & Stanley, 2000). The MST is also a technique to cluster the nodes in a non-hierarchical clustering by exploring the topological properties of all the factors.

Next, after the network topology of all the factors were constructed, the Pajek software was used to visualize the network. From this network topology, centrality measures were used to understand the importance of each node relative to the others (Borgatti, 2005). There are three different concepts of centrality, as discussed by (Hanneman & Riddle, 2005). The three concepts are degree centrality, closeness centrality, and betweenness centrality, while (Borgatti, 1995) defines four measures of centrality. The latter is eigenvector centrality. Below is the measurement used by Borgatti (2005):

i) Degree centrality is defined as the number of ties that a given node has. The degree of node i is given by: where $d_i = \sum_j a_{ij}$ if the i -th and j -th nodes are linked and 0 otherwise.

ii) Closeness centrality is defined as the total graph-theoretic distance of a given node from all other nodes, $c_i = \sum_j d_{ij}$, where d_{ij} is the number of links in the shortest path from i to j . Larger value indicates less central, while smaller value indicates more central.

iii) Betweenness centrality is the number of the shortest paths that pass through a given node;

$$b_k = \sum_{i,j} \frac{g_{ijk}}{g_{ij}}$$

where g_{ij} is the shortest path from node i to node j , and g_{ijk} is the shortest path from i to j that passes through k .

iv) Eigenvector centrality is also known as a variant of simple degree. An eigenvector of a symmetric square

Degree			Betweenness		Closeness		Eigenvector		Overall		
1	NC6	4	1	WPS8	0.6181	1	WPS8	0.2108	1	WPS8	0.0051
2	WPS2	3	2	AC7	0.5533	2	WPS9	0.2063	2	AC7	0.0048
3	WPS3	3	3	WPS6	0.5385	3	AC7	0.2021	3	WPS2	0.0047
4	WPS6	3	4	WPS9	0.5047	4	WPS10	0.2	4	AC1	0.0047
5	WPS8	3	5	WPS10	0.4966	5	WPS7	0.1921	5	WPS9	0.0047
6	AC1	3	6	WPS7	0.4858	6	WPS6	0.1831	6	WPS6	0.0044
7	AC7	3	7	AC1	0.3887	7	AC1	0.1831	7	AC8	0.0039
8	CC3	3	8	CC4	0.274	8	WPS2	0.1814	8	WPS3	0.0038
9	CC4	3	9	NC4	0.2672	9	AC8	0.1749	9	CC4	0.0036
10	CC6	3	10	NC6	0.2389	10	CC4	0.1632	10	WPS10	0.0033
11	WPS7	2	11	WPS15	0.2294	11	NC4	0.1632	11	NC6	0.0033
12	WPS9	2	12	WPS2	0.193	12	WPS15	0.1618	12	WPS15	0.0033
13	WPS10	2	13	CC3	0.193	13	AC2	0.1579	13	WPS7	0.0029
14	WPS11	2	14	WPS14	0.1889	14	WPS3	0.1566	14	AC2	0.0028
15	WPS12	2	15	AC8	0.1889	15	WPS1	0.1542	15	NC4	0.0027
16	WPS13	2	16	WPS11	0.1457	16	AC6	0.1529	16	CC3	0.0026
17	WPS14	2	17	AC6	0.1457	17	NC6	0.1461	17	WPS1	0.0026
18	WPS15	2	18	WPS3	0.1012	18	CC3	0.145	18	AC6	0.0025
19	WPS16	2	19	CC6	0.1012	19	WPS14	0.1439	19	CC6	0.0025
20	AC2	2	20	WPS12	0.0999	20	CC1	0.1408	20	WPS14	0.0025
21	AC3	2	21	AC2	0.0999	21	AC3	0.1378	21	NC2	0.0024
22	AC5	2	22	AC5	0.0999	22	WPS4	0.1359	22	WPS4	0.0023
23	AC6	2	23	NC2	0.0999	23	WPS5	0.1359	23	WPS5	0.0023
24	AC8	2	24	WPS13	0.0513	24	AC5	0.1349	24	AC3	0.0022
25	NC2	2	25	WPS16	0.0513	25	NC2	0.1296	25	CC1	0.0022
26	NC3	2	26	AC3	0.0513	26	WPS11	0.1287	26	NC5	0.0022
27	NC4	2	27	NC3	0.0513	27	CC6	0.1287	27	NC7	0.0022
28	WPS1	1	28	WPS1	0	28	NC5	0.1279	28	AC5	0.0019
29	WPS4	1	29	WPS4	0	29	NC7	0.1279	29	CC2	0.0019
30	WPS5	1	30	WPS5	0	30	CC2	0.127	30	WPS11	0.0018
31	WPS17	1	31	WPS17	0	31	AC4	0.1215	31	NC3	0.0017
32	AC4	1	32	AC4	0	32	WPS16	0.12	32	CC5	0.0014
33	CC1	1	33	CC1	0	33	WPS12	0.1157	33	CC7	0.0014
34	CC2	1	34	CC2	0	34	NC3	0.1157	34	AC4	0.0013
35	CC5	1	35	CC5	0	35	CC5	0.1144	35	WPS16	0.0011
36	CC7	1	36	CC7	0	36	CC7	0.1144	36	WPS12	0.0011
37	NC1	1	37	NC1	0	37	WPS17	0.1074	37	NC8	0.0011
38	NC5	1	38	NC5	0	38	WPS13	0.1046	38	WPS13	0.0011
39	NC7	1	39	NC7	0	39	NC8	0.104	39	WPS17	0.0011
40	NC8	1	40	NC8	0	40	NC1	0.0949	40	NC1	0.0011

Figure 3. Centrality measures scores of the Tschuprow's measure of association

3.2.4 Eigenvector Centrality

The importance of a node in a network can be measured by using the eigenvector centrality. The relative scores to all the nodes in the network based on the principle that connections with the highest scores of nodes contribute more compared to the lowest scores of nodes. In contrast to degree centrality, eigenvector centrality favours nodes that are connected to nodes that are themselves central within the network (Lohmann, 2010). In the Tschuprow's measure of association, WPS8 (0.3904) had the highest score, and strong relationships with AC7 (0.3809) and WPS2 (0.3048). These nodes had excellent positions and had the highest potential to spread the information throughout the network within a short time.

3.2.5 Overall Centrality

The score for each centrality measure had different roles or functions in every characteristic; thus, the overall centrality measure was needed because the overall centrality measure would identify the most important characteristic. Here, the overall centrality measure was defined as an optimal linear combination of the four centrality measures. The optimality characteristics were based on the PCA of data matrix of size 204 x 4, which represented the 204 respondents and the scores in the four centrality measures. The first principal component

explained 92.8% of the total variations, and the second principal component only explained about 0.062%, which was very small compared to the first principal component. The overall centrality measure was determined since the first principal component was sufficiently adequate. The values of each eigenvector were $e_1 = 0.042046$, $e_2 = 0.002789$, $e_3 = 0.000401$, and $e_4 = 0.000064$. The values were substituted in Equation (6) to find the overall centrality. The most important characteristics were WPS8 (0.0051) and WPS6 (0.0048).

4. Concluding Remarks

The connections among the nodes that were linked to the groups were different from both measurements of the association. The 40 factors were supposed to gather in their own groups. By referring to the MST results, according to the Spearman correlation coefficient, WPS16, WPS17, NC1, NC8, and AC4 were separated from their own groups. WPS16 and WPS17 were directly under the AC group; meanwhile, NC1 and NC8 were directly under the WPS and AC groups relatively. Another node, AC4, was under WPS17, which was directly under the AC group.

In the Tschuprow's measure of association, the connections in the network were better compared to the Spearman correlation coefficient. All nodes gathered in their own groups, except for WPS16, WPS17, and NC1. WPS16 and WPS17 were still directly under the AC group and only NC1 was directly under the WPS group. NC8 was already directly under the NC group.

From these results, the Tschuprow's measure of association showed better performance than the Spearman correlation coefficient. A good measure represents the reality. In this study, all the factors in the same groups were highly correlated to each other. From the MST results, the Tschuprow's measure of association gave better results in the grouping of the 40 factors into their own groups, as compared to the Spearman correlation coefficient. The results from the Tschuprow's measure of association showed that almost all the characteristics were gathered or clustered in their own groups. This means that the Tschuprow's measure correlated better than the Spearman correlation coefficient. It grouped the factors under the right groups, and hence, more effective than the Spearman correlation coefficient.

The results of the overall centrality measure showed that the most important factors were WPS8 (0.0051), WPS6 (0.0048), WPS9 (0.0047), WPS10 (0.0047), and WPS7 (0.0047) for the Tschuprow's measure of association. These 5 factors were the most important and influential factors.

WPS8 = My organization respects my "inner life".

WPS6 = I feel positive about the values prevailing in my organization.

WPS9 = My organization helps me to live in peace/harmony with myself.

WPS10 = The leaders of my organization try to be helpful to the larger social good of the community.

WPS7 = People feel good about their future with the organization.

These 5 factors showed the work attitude among Malaysian academicians in the public universities. The factors were from the same group of WPS. The factors were concerned about the work environment, including the members in the organization. The role of a public university in Malaysia in handling the challenges to increase the tertiary education system is very important and it must include the development of humanistic work environment as well. The high performance of an employee in an organization would definitely help to strengthen and increase the level of productivity, efficiency, and the quality in the organization.

In the future, we suggest that this study is continued with another measure of association, for example, the Pearson correlation coefficient. One can also concentrate on every group instead of studying all the groups. This means that every group has their own MST and centrality measures. By doing so, more information can be obtained based on groups. Perhaps, the results can help the Malaysian government to come up with a better strategic plan for development.

Acknowledgments

The authors are grateful to the Editor and anonymous referees for their comments and suggestions that led to the final presentation of this paper. A special thanks also goes to the Ministry of Higher Education of Malaysia for the sponsorship under FRGS vote number 4F260, Universiti Teknologi Malaysia, Universiti Tun Hussein Onn Malaysia, and Universitas Pasundan, Indonesia, for the research facilities.

References

- Allen, N. J., & Meyer, J. P. (1996). *Affective, continuance, and normative commitment to the organization: An examination of construct validity*. *Journal of Vocational Behavior*, 49, 252–276.

- <http://dx.doi.org/10.1006/jvbe.1996.0043>
- Arme'nio, R., & Miguel, P. C. (2008). Workplace spirituality and organizational commitment: An empirical study. *Journal of Organizational Change Management*, 21(1), 53-75. <http://dx.doi.org/10.1108/09534810810847039>
- Bergsma, W. (2012). A bias-correction for Cramér's V and Tschuprow's T. *Journal of the Korean Statistical Society*.
- Borgatti, S. P. (1995). Centrality and AIDS. *Connections*, 18(1), 112-114.
- Borgatti, S. P. (2005). Centrality and network flow. *Social networks*, 27(1), 55-71. <http://dx.doi.org/10.1016/j.socnet.2004.11.008>
- Chang, J. Y., & Choi, J. N. (2007). The Dynamic Relation Between Organizational and Professional Commitment of Highly Educated Research and Development (R&D) Professionals. *The Journal of Social Psychology*, 147(3), 299-315. <http://dx.doi.org/10.3200/SOCP.147.3.299-315>
- Choong, Y., Lau, T., & Wong, K. (2011). Intrinsic Motivation and Organizational Commitment in the Malaysian Private Higher Education Institutions: An Empirical Study. *Journal of Arts, Science & Commerce*, 2(4), 40-50.
- Conover, W. J. (1971). *Practical Nonparametric Statistics*. John Wiley & Sons, Inc.
- Darwish, A. Y. (2002). Job satisfaction as a mediator of the relationship between role stressors and organizational commitment: A study from an Arabic cultural perspective. *Journal of Managerial Psychology*, 17(4), 250-266. <http://dx.doi.org/10.1108/02683940210428074>
- Daud, N. (2012). The Influence of Quality of Work Life on Organizational Commitment: A study on Academic Staff in Public Institution of Higher Learning in Malaysia. *IEEE Transactions on Innovation Management and Technology Research*, 673-678.
- Djauhari, M. A., Sharif, S., & Djauhari, H. (2012). Network Analysis on Safety Culture and Worker's Behaviour: A Forest of All Minimum Spanning Trees. *International Journal of Basic & Applied Sciences*, 12(6), 29-37.
- Freeman, L. C. (2004). *The Development of Social Network Analysis: A Study in the Sociology of Science*. BookSurge, North Charleston, SC.
- Gan, S. L., & Djauhari, M. A. (2012). An Overall Centrality Measure: The Case of U.S Stock Market. *International Journal of Basic & Applied Sciences*, 12(6), 99-103.
- Gan, S. L., & Djauhari, M. A. (2012). Network Topology of Indonesian Stock Market. *IEEE Transactions on Cloud Computing and Social Networking*, 1-4.
- Gan, S. L., & Djauhari, M. A. (2012). Stock networks analysis in Kuala Lumpur Stock Exchange. *Malaysian Journal of Fundamental and Applied Sciences*, 8(2), 60-66.
- Gonza'lez, T. F., & Guille'n, M. (2008). Organizational Commitment: A Proposal for a Wider Ethical Conceptualization of 'Normative Commitment. *Journal of Business Ethic*, 78, 401-414. <http://dx.doi.org/10.1007/s10551-006-9333-9>
- Hanneman, R. A., & Riddle, M. (2005). *Introduction to social network methods*. Riverside, CA: University of California, Riverside. Retrieved from <http://faculty.ucr.edu/~hanneman/>
- Harrington, W. J., Preziosi, R. C., & Gooden, D. J. (2002). Perceptions of Workplace Spirituality among Professionals and Executives. *Employee Responsibilities and Rights Journal*, 13(3), 155-163. <http://dx.doi.org/10.1023/A:1014966901725>
- Hauke, J., & Kossowski, T. (2011). Comparison of values of Pearson's and Spearman's correlation coefficient on the same sets of data. *Quaestiones Geographicae*, 30(2), 87-93.
- Hayes, B. E. (2008). *Measuring Customer Satisfaction and Loyalty: Survey Design, Use, and Statistical Analysis Methods*. American Society for Quality, Quality Press.
- Jaros, S. J., Jermier, J. M., Koehler, J. W., & Sincich, T. (1993). Effects of Continuance, Affective, and Moral Commitment on the Withdrawal Process: An Evaluation of Eight Structural Equation Models. *Academy of Management Journal*, 36(5), 951-955. <http://dx.doi.org/10.2307/256642>
- Jerrold, H. Z. (1972). Significance Testing of the Spearman Rank Correlation Coefficient. *Journal of the American Statistical Association*, 67(339), 578-580. <http://dx.doi.org/10.1080/01621459.1972.10481251>

- Lohmann, G., Margulies, D. S., Horstmann, A., Pleger, B., & Lepsien, J. et al. (2010). Eigenvector Centrality Mapping for Analyzing Connectivity Patterns in fMRI Data of the Human Brain. *PLoS ONE*, 5(4), 1-8. <http://dx.doi.org/10.1371/journal.pone.0010232>
- Mantegna, R. N., & Stanley, H. E. (2000). *An Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge University Press.
- Mantegna, R. N. (1999). Hierarchical Structure in Financial Markets. *The European Physical Journal B*, 11, 193-197. <http://dx.doi.org/10.1007/s100510050929>
- Meyer, J. P., & Herscovitch, L. (2001). Commitment in the workplace: Toward a general model. *Human Resource Management Review*, 11(3), 299-326. [http://dx.doi.org/10.1016/S1053-4822\(00\)00053-X](http://dx.doi.org/10.1016/S1053-4822(00)00053-X)
- Naylor, M. J., Rose, L. C., & Moyle, B. J. (2007). Topology of Foreign Exchange Markets Using Hierarchical Structure Methods. *Physica A*, 382(1), 199-208. <http://dx.doi.org/10.1016/j.physa.2007.02.019>
- Noordin, F., & Jusoff, K. (2009). Levels of Job Satisfaction amongst Malaysian Academic Staff. *Asian Social Science*, 5(5), 122-128. <http://dx.doi.org/10.5539/ass.v5n5p122>
- Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, 32(3), 245-251. <http://dx.doi.org/10.1016/j.socnet.2010.03.006>
- Roy, R. B., & Sarkar, U. K. (2011). A social network approach to examine the role of influential stocks in shaping interdependence structure in global stock markets. *IEEE Transactions on Social Networks Analysis and Mining*, 567-569.
- Smeenck, S. G. A., Eisinga, R. N., Teelken, J. C., & Doorewaard, J. A. C. M. (2006). The effects of HRM practices and antecedents on organizational commitment among university employees. *International Journal of Human Resource Management*, 17(12), 2035-2054. <http://dx.doi.org/10.1080/09585190600965449>
- Spearman, C. (1904). The Proof and Measurement of Association Between Two Things. *American Journal of Psychology*, 72-101. <http://dx.doi.org/10.2307/1412159>
- Suleiman, E. S., Ismail, W. K. W., Nor, K. M., & Long, C. S. (2012). *Workplace Spirituality And Normative Commitment* (pp. 30-40). Proceeding on 2nd International Conference on Management.
- Svante, J., & Jan, V. (1978). On the Applicability of Truncated Component Analysis Based on Correlation Coefficient for Nominal Scales. *Applied Psychological Measurement*, 2(1), 135-145. <http://dx.doi.org/10.1177/014662167800200113>
- Tan, P. N., Kumar, V., & Srivastava, J. (2004). Selecting the Right Objective Measure for Association Analysis. *Information Systems*, 29, 293-313. [http://dx.doi.org/10.1016/S0306-4379\(03\)00072-3](http://dx.doi.org/10.1016/S0306-4379(03)00072-3)
- Therese, A. J., & Steve, B. (2006). The antecedents of organizational commitment: the case of Australian casual academics. *International Journal of Educational Management*, 20(6), 439-452. <http://dx.doi.org/10.1108/09513540610683694>
- Tomizawa, S. (1994). Two Kinds of Measures of Departures from Symmetry in Square Contingency Tables Having Nominal Categories. *Statistica Sinica*, 325-334.
- Yusoff, N. S., Djauhari, M. A., & Suleiman, E. S. (2012). A social network analysis on organizational commitment in Malaysia. *IEEE Transactions on Cloud Computing and Social Networking*, 1-4.
- Yusoff, N. S., Djauhari, M. A., Sharif, S., & Suleiman, E. S. (2012). Organizational Commitment in Malaysian Public University: An Evidence via Social Network Analysis. *International Journal of Basic & Applied Sciences*, 12(5), 17-21.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

Efficiency Analysis of Low Power Class-E Power Amplifier

Mousa Yousefi¹, Ziaadin Daie Koozehkanani¹, Jafar Sobhi¹, Hamid Jangi¹ & Nasser Nasirezadeh¹

¹ Faculty of Electrical and Computer Engineering, University of Tabriz, Iran

Correspondence: Mousa Yousefi, Faculty of Electrical and Computer Engineering, University of Tabriz, Iran. Tel: 98-411-339-3739. E-mail: m.yousefi@tabrizu.ac.ir

Received: April 12, 2014

Accepted: June 3, 2014

Online Published: August 6, 2014

doi:10.5539/mas.v8n5p19

URL: <http://dx.doi.org/10.5539/mas.v8n5p19>

Abstract

This paper presents an analysis of effect of inductor and switch losses on output power and efficiency of low power class-E power amplifier. This structure is suitable for integrated circuit implementation. Since on chip inductors have large losses than the other elements, the effect of their losses on efficiency has been investigated. Equations for the efficiency have been derived and plotted versus the value of inductors and switch losses. Derived equations are evaluated using MATLAB. Also, Cadence Spectre has been used for schematic simulation. Results show a fair matching between simulated power loss and efficiency and MATLAB evaluations. Considering the analysis, the proposed power amplifier shows about 13 % improvement in power efficiency at 400 MHz and -2 dBm output power. It is simulated in 0.18 μm CMOS technology.

Keywords: low power, class e, power amplifier, efficiency analysis

1. Introduction

High efficiency and low level output power design of power amplifier (PA) is a requirement for optimization of the energy efficiency of the transmitter which is one the key building blocks of sensor nodes in wireless sensor networks. The class-E power amplifier can ideally achieve 100% efficiency. This high efficiency has spurred many research interests on the design and analysis of Class-E Pas (Apostolidou, et al, 2009; Lee, et al 2010; Brama, et al, 2008; Mertens, et al, 2002; Tsai, et al, 1999; Reynaert, 2006). The conventional class-E power amplifier can produce large power levels with good efficiency (Lee, et al 2010; Brama, et al, 2008; Mertens, et al, 2002). Most of the existing Class-E PA designs have been optimized to work at high output power levels, ranging from 23 to 33 dBm (Lee, et al 2010; Brama, et al, 2008; Mertens, et al, 2002; Tsai, et al, 1999; Reynaert, 2006; Mousa, 2013). If these fully integrated PAs are used in applications requiring low level output power such as wireless body sensor networks, the overall efficiency significantly degrades (Tan, et al, 2012). For example Bluetooth and ZigBee standards are short range standards that their output power level are from 0 to 10 dBm (Retz, et al, 2009; Eo, et al, 2007; Bae, et al, 2011) and in wireless body sensor networks it is even under 0 dBm (Cook, et al, 2006; Tan, et al, 2012). Therefore, high efficiency PA with low level output power is critical to short range wireless sensor network.

The efficiency of low power class-E power amplifier (LPCEPA) introduced by Jun Tan (Tan, et al, 2012) is appropriate for use in transmitter block of the sensor node in short range wireless sensor networks. LPCEPA architecture is a proper option for fully integrated PA solutions. Among the elements of a fully integrated design, on chip inductors have large losses than the others and have the most adverse effect on overall efficiency of the transmitter.

In this paper, to investigate the effect of inductor losses on LPCEPA efficiency, the equations of losses of elements and efficiency of PA vs. losses are derived and based on these observations an appropriate PA has been proposed.

The rest of this paper is organized as follows: In Section 2, the circuitry of the LPCEPA and circuit description is presented. Section 3 presents the analytical equations of the losses. In section 4, simulation results of the proposed PA are presented. Section 5 concludes the paper.

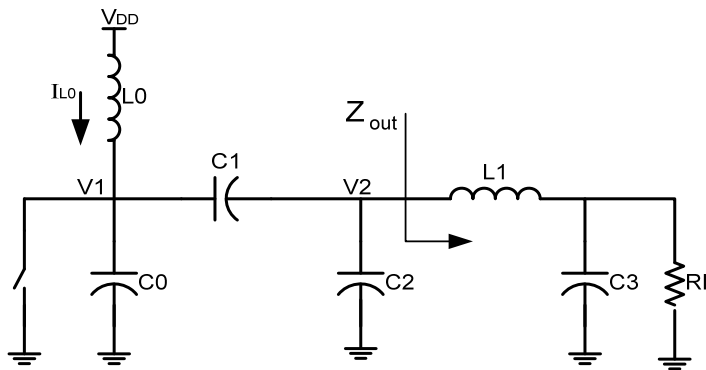


Figure 1. Schematic diagram of the LPCEPA

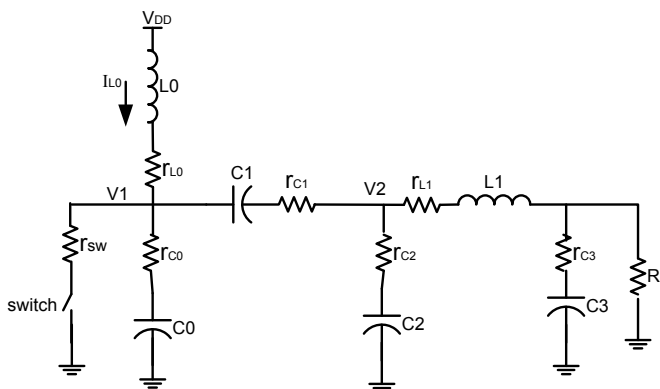


Figure 2. Schematic diagram of the LPCEPA with losses modelled as resistor

2. Circuit Description

With the following assumptions, the circuit model of LPCEPA shown in Figure 1 has been selected for evaluation. In the analysis of PA, except the resistor R_L , all of the elements are supposed to be ideal and the transistor is an ideal switch with zero and infinite resistance when turns on and off, respectively. Two equations define the class-E PA conditions, as below: (Apostolidou, et al, 2009; Lee, et al 2010; Brama, et al, 2008; Mertens, et al, 2002; Tsai, et al, 1999; Reynaert, 2006)

$$V_1(t_1) = 0, \quad \frac{dV_1(t_1)}{dt} = 0 \tag{1}$$

where t_1 is the time that switch turns on.

This equivalent circuit consists of a switch shunt capacitance C_0 , matching network and load resistor R_L . For the analysis, suppose that frequency of input signal, voltage of power supply and duty cycle are known variables. Unknown variables in the circuitry of Figure 1 are six variables; C_0, C_1, C_2, C_3, L_0 and L_1 . To find unknown variables, six independent equations are required. Four new design variables are defined as:

$$C_{eq} = C_0 + \frac{C_1 C_2}{C_1 + C_2} \tag{2}$$

$$\alpha = C_1 / (C_1 + C_2) \tag{3}$$

$$\beta = C_0 / C_{eq} \tag{4}$$

$$q = 1 / (\omega_0 \sqrt{L_0 C_{eq}}) \tag{5}$$

where C_{eq} denotes the total capacitance at node V_1 , α is the ratio of C_1 to C_1+C_2 , β is the ratio of parallel

capacitance C_0 to C_{eq} and q is the normalized frequency. The design challenge is to compute these four new variables. Once calculated, the value of the real elements C_0 ~ C_2 and L_0 can be specified. The two variables I_a and φ are concluded next. When the switch is off, KCL equations at nodes V_1 and V_2 are:

$$C_1 \left(\frac{dV_1(t)}{dt} - \frac{dV_2(t)}{dt} \right) = C_2 \frac{dV_2(t)}{dt} + I_a \cos(\omega_0 t + \varphi) \quad (6)$$

$$i_{L_0}(t) = C_0 \frac{dV_1(t)}{dt} + C_1 \left(\frac{dV_1(t)}{dt} - \frac{dV_2(t)}{dt} \right) \quad (7)$$

and the voltage drops across the inductor L_0 are as follows for the time switch is off and when it is on, respectively.

$$V_{DD} - V_1(t) = L_0 \frac{di_{L_0}(t)}{dt} \quad (8)$$

$$V_{DD} = L_0 \frac{di_{L_0}(t)}{dt} \quad (9)$$

When the switch is ON, node V_2 requires:

$$(C_1 + C_2) \frac{dV_2(t)}{dt} = -I_a \cos(\omega_0 t + \varphi) \quad (10)$$

Solving differential equations, the waveform of V_1 , V_2 and I_L when switch is off are:

$$V_1(t) = A_1 V_{DD} \cos(\omega_1 t) + A_2 V_{DD} \sin(\omega_1 t) + \kappa V_{DD} \frac{1}{q^2 - 1} \sin(\omega_0 t + \varphi) + V_{DD} \quad (11)$$

$$V_2(t) = \frac{C_1}{C_1 + C_2} V_1(t) - \frac{I_a}{(C_1 + C_2) \omega_0} \sin(\omega_0 t + \varphi) + A_3 V_{DD} \quad (12)$$

$$i_{L_0}(t) = C_{eq} \omega_1 [A_2 V_{DD} \cos(\omega_1 t) - A_1 V_{DD} \sin(\omega_1 t)] + \kappa V_{DD} C_{eq} \omega_0 \frac{q^2}{q^2 - 1} \sin(\omega_0 t + \varphi) + V_{DD} \quad (13)$$

The variable κ is defined as:

$$\kappa = \frac{\alpha I_a}{C_{eq} \omega_0 V_{DD}} \quad (14)$$

For determining A_1 , A_2 and κ , boundary conditions of voltage waveform V_1 can be used. The initial condition of V_1 is $V_1(0) = 0$. Solving the linear algebraic equations, the variables of A_1 , A_2 and κ are:

$$A_1 = \frac{q(\sin(a_2) - \sin(a_1) + \sin(a_3) - \sin(a_4)) + \sin(a_1) + \sin(a_2)}{q \sin(a_1) - q \sin(a_2) - 2q \sin(\varphi) - \sin(a_1) - \sin(a_2)} \quad (15)$$

$$A_2 = \frac{\sin(a_5) - \sin(\varphi)}{\sin(\varphi) \sin(a_4 + \varphi)} - A_1 \frac{\sin(\varphi) \cos(a_4 + \varphi) - \sin(a_5)}{\sin(\varphi) \sin(a_4 + \varphi)} \quad (16)$$

$$\kappa = \frac{-2q(q^2 \cos(2q\pi(1-D)) - \cos(2q\pi(1-D)) - q^2 + 1)}{q \sin(a_1) - q \sin(a_2) - 2q \sin(\varphi) - \sin(a_1) - \sin(a_2)} \quad (17)$$

and for the time interval that switch is on, V_1 is zero. Also, the current of inductor L_0 is:

$$i_{L_0}(t) = \frac{V_{DD}}{L_0} t + i_L(t_1) \quad (18)$$

By solving differential equation (10), V_2 can be calculated as:

$$V_2(t) = -\frac{I_a}{(C_1 + C_2)\omega_0} \sin(\omega_0 t + \varphi) + A_3 V_{DD} \quad (19)$$

to satisfy that all waveforms are periodic with the period of T , φ is:

$$\varphi = \tan^{-1}\left(\frac{k_2 + 2q\pi Dk_3 - k_1}{-g_2 - 2q\pi Dg_3 + g_1}\right) \quad (20)$$

The variables $a_1, a_2, a_3, a_4, a_5, k_1, k_2, k_3, g_1, g_2$ and g_3 used in equation above are given in the Appendix A. The output impedance Z_{out} at the first frequency can be computed by:

$$Z_{out} = (R_L \parallel \frac{1}{jC_3\omega_0}) + jL_1\omega_0 \quad (21)$$

Because V_2 is a periodic waveform, it can be expanded into its Fourier series. At the first frequency ω_0 , Z_{out} is:

$$Z_{out} = \frac{V_{2_1}}{I_a} \exp[(\varphi_1 - \varphi)j] = Z_{out_real} + Z_{out_imag} \quad (22)$$

V_{2_1} and φ_1 are amplitude and phase of the fundamental harmonic of V_2 , respectively. From (21) and (22) equations, the C_3 and L_1 can be calculated.

With computing average current of L_0 inductor, consumption power can be calculated as below:

$$P_{DC} = V_{DD}^2 C_{eq} h \quad (23)$$

The variable h is given in the Appendix A

3. Analysis of Power Losses and Efficiency

In analysis of losses of LPCEPA, losses of elements are shown by a resistor in series with elements. Figure 2 shows the equivalent circuit of the power amplifier. The r_{L0} and r_{L1} are losses of the inductors L_0 and L_1 , respectively. To simplify derivation of loss equations for inductors, root means square (RMS) current of L_0 inductor and maximum current of L_1 inductor are determined with suppose that currents of elements stay no changed when parasitic resistance are not zero.

The RMS current of inductor L_0 is:

$$I_{rms,L_0} = \sqrt{\frac{1}{T} \int_0^T i_{L_0}^2(t) dt} \quad (24)$$

and power loss in L_0 inductor is determined by:

$$P_{loss,L_0} = I_{rms,L_0}^2 r_{L_0} \quad (25)$$

For the time in which the switch is OFF ($0 < t < t_1$) i_{L0} is presented by equation 13 and for ON time of the switch it is given by equation 18. By integrating current from 0 to t_1 and from t_1 to T , I_{rms,L_0} can be calculated as:

$$I_{rms,L_0} = C_{eq} \sqrt{\frac{1}{T} (E_1 + E_2 + E_3 + E_4 + E_5 + E_6 + E_7)} \quad (26)$$

where variables $E_1 \sim E_7$ are given in the Appendix B.

Also, power loss in inductor L_1 is determined by:

$$P_{loss,L_1} = 0.5 I_a^2 r_{L_1} \quad (27)$$

I_a is:

$$I_a = \frac{\kappa C_{eq} \omega_0 V_{DD}}{\alpha} \quad (28)$$

The current that flows through capacitor C0 is:

$$i_{C_0} = C_0 \frac{dV_1}{dt} \quad (29)$$

The RMS current of capacitor C₀ is given by equation 30.

$$I_{rms,C_0} = C_0 \sqrt{\frac{1}{T} (E_1 + E_2 + E_3 / q^4 + E_4 + E_5 / q^2 + E_6 / q^2)} \quad (30)$$

The power loss in shunt capacitor C₀ is shown in equation 31.

$$P_{loss,C_0} = I_{rms,C_0}^2 \cdot r_{C_0} \quad (31)$$

Same procedure for capacitors C₁ and C₂ is done and equation for their current, RMS current and power loss are given by equations 32 ~ 37.

$$i_{C_1} = C_1 \left(\frac{dV_1}{dt} - \frac{dV_2}{dt} \right) \quad (32)$$

$$I_{rms,C_1} = C_{eq} \sqrt{\frac{1}{T} (E_1 + E_2 + E_3 + E_4 + E_5 + E_6 + E_8)} \quad (33)$$

$$P_{loss,C_1} = I_{rms,C_1}^2 \cdot r_{C_1} \quad (34)$$

$$i_{C_2} = C_2 \frac{dV_2}{dt} \quad (35)$$

$$I_{rms,C_2} = \frac{C_1 C_2}{C_1 + C_2} \sqrt{\frac{1}{T} (E_1 + E_2 + E_3 / q^4 + E_4 + E_5 / q^2 + E_6 / q^2)} \quad (36)$$

$$P_{loss,C_2} = I_{rms,C_2}^2 \cdot r_{C_2} \quad (37)$$

Considering Figure 2 the current flowing through C₃ can be calculated as:

$$I_{m,C_3} = I_a \frac{RC\omega_0}{\sqrt{(RC\omega_0)^2 + 1}} \quad (38)$$

And its power loss is:

$$P_{loss,C_3} = 0.5 I_{m,C_3}^2 \cdot r_{C_3} \quad (39)$$

To calculate losses of switch resistance, its current must be known. When switch is open, its current is zero and when it's closed, its current is calculated using KCL in V₁ node as follows:

$$i_{sw} = i_{L0} - i_{C1} \quad (40)$$

and its RMS value is calculated by:

$$I_{rms,sw} = \sqrt{E_9 + E_{10} + Ceq^2 E_7 / T - C_1^2 C_2^2 E_6 / (q^2 (C_1 + C_2)^2 T)} \quad (41)$$

then, the losses of switch is:

$$P_{loss,sw} = I_{rms,sw}^2 \cdot r_{sw} \quad (42)$$

Total power loss in the LPCEPA is the sum of power loss in inductors L_0 and L_1 and power loss in capacitances $C_0 \sim C_3$.

$$P_{Loss} = P_{loss,L0} + P_{loss,L1} + P_{loss,C0} + P_{loss,C1} + P_{loss,C2} + P_{loss,C3} + P_{loss,sw} \quad (43)$$

Total power consumption is given by equation 23 and efficiency of the PA is:

$$\eta = \frac{P_{DC} - P_{loss}}{P_{DC}} \quad (44)$$

4. Simulation Results

The purpose of design of the LPCEPA is to be used in the transmitter of a short range wireless sensor network structure. Major advantages of LPCEPA are low level output power, high efficiency and on chip implementation of all elements. Since on chip inductors have large losses than the other elements, their effect on overall efficiency of PA has been discussed. To investigate the losses of on chip inductors in the PA, a schematic of LPCEPA (Figure 2) has been simulated with element values listed in Table 1 using Cadence Spectre. Furthermore, equations governing this structure have been evaluated in MATLAB. In these simulations, only the losses of inductors and switch have been considered. Waveforms of the voltage of the drain terminal and the current of the inductor L_0 are shown in the Figure 3. Figure 4 shows the efficiency versus the losses of L_0 , L_1 and switch when considered separately and together. Figure 5 shows the same for MATLAB evaluation.

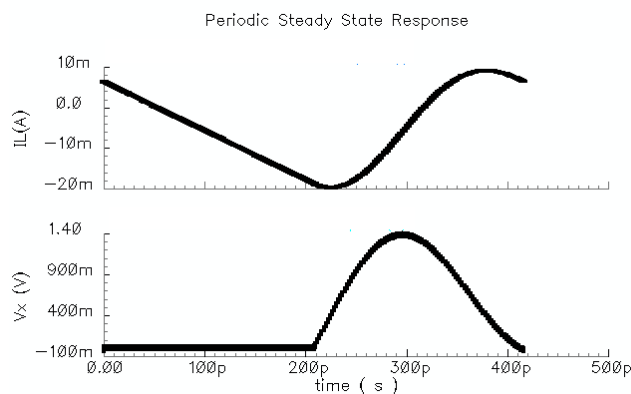


Figure 3. Waveforms of the drain node voltage and L_0 inductor current

Table 1. Simulation parameters and element values for simulated

<i>Parameters & Elements</i>	<i>Value</i>	<i>Unit</i>
L_0	3.3	nH
L_1	3.3	nH
C_0	240	fF
C_1	1	pF
C_2	1.3	pF
C_3	1.6	pF
R_L	50	ohm
Frequency	2.4	GHz
Duty cycle (D)	0.5	--
Power Supply	0.5	Volt

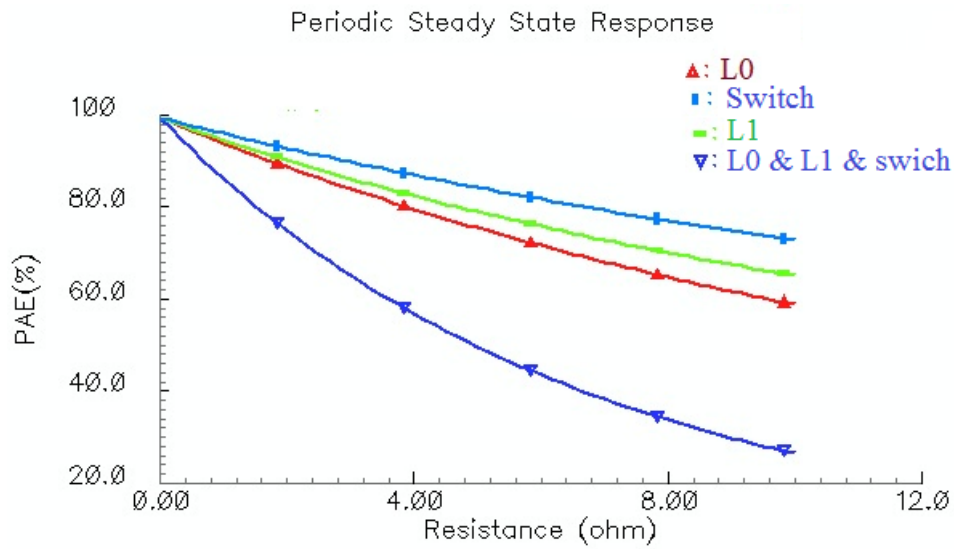


Figure 4. Efficiency of PA versus losses of inductors and switch (Cadence Spectre)

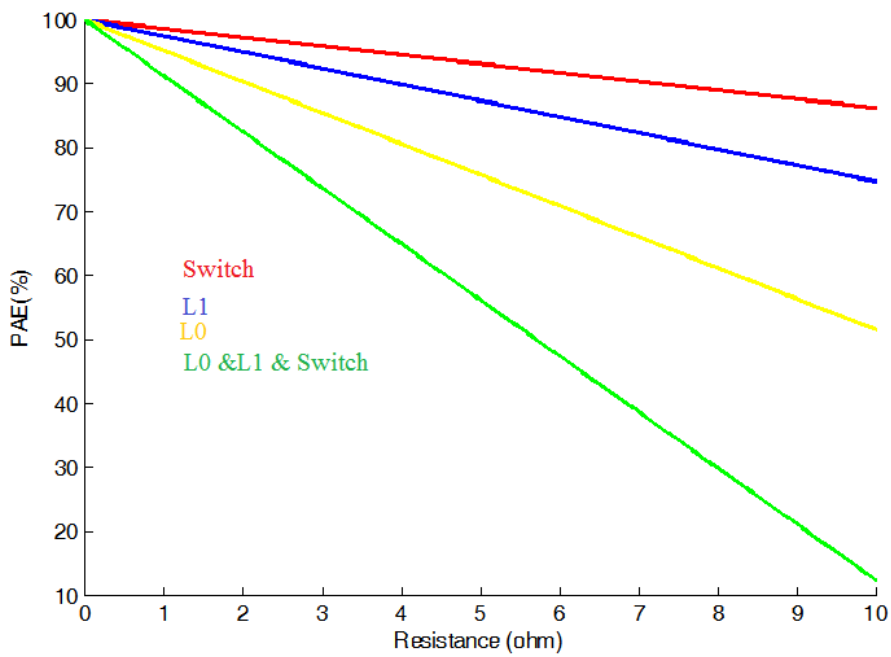


Figure 5. Efficiency of PA versus losses of inductors and switch (MATLAB)

Considering the efficiency analysis of the amplifier it can be concluded that to decrease the losses, output network can be changed such that the first harmonic RMS current does not pass through the inductor. To do this, we propose the following circuit. Simulation results for two structures show about 13 % improvement in efficiency.

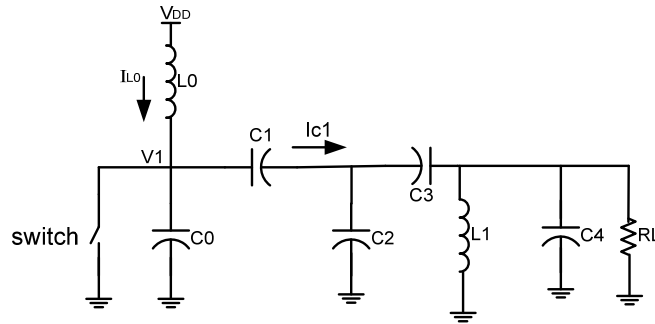


Figure 6. Proposed PA

To investigate the results, structures of Figure 1 and Figure 6 have been simulated using the component values listed in Table 2. PAE of both circuits versus series loss of inductors have been shown in Figure 8. For a 4-ohm series resistance, the difference in PAE of two circuits is 19%. Also, the analysis of both structures employing on chip inductors for -2 dBm output power at 400 MHz frequency shows that PAE of circuit LPCEPA (Figure 1) is 20% and that of circuit proposed PA (Figure 6) is 33%. Results are summarized in Table 3.

The proposed fully integrated amplifier for 400 MHz and -2 dBm has been shown in Figure 6. Component values are listed in Table 2. As supply voltage varies from 0.35 V to 0.6 V the output power varies from -3.1 dBm to 1.1 dBm (Figure 9).

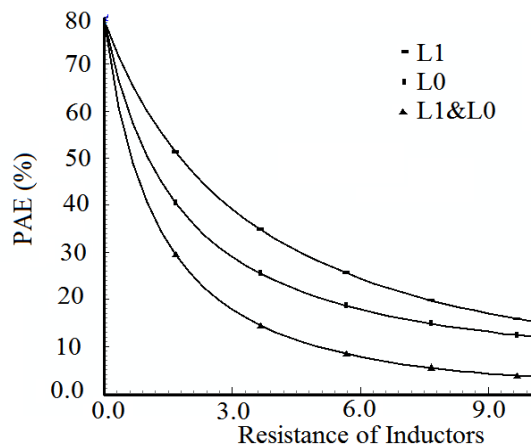


Figure 7. Efficiency of LPOCEPA versus losses of inductors (Cadence Spectre) with 2.4 GHz

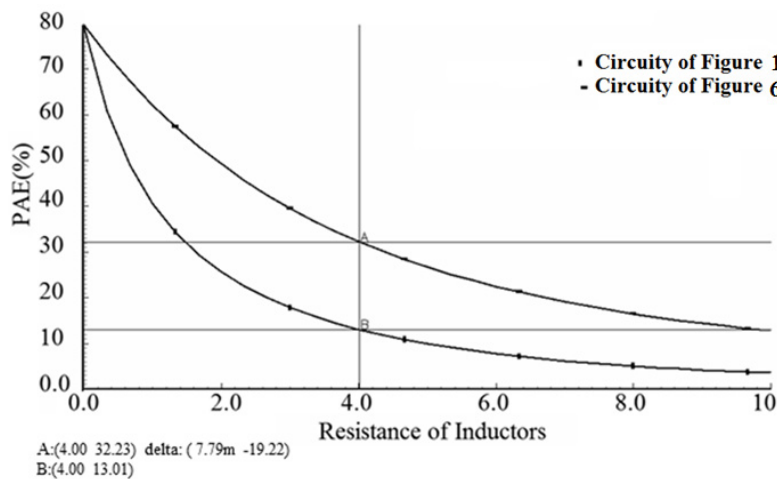


Figure 8. PAE of proposed PA Figure 5 and Figure 4 of circuit with losses of inductors

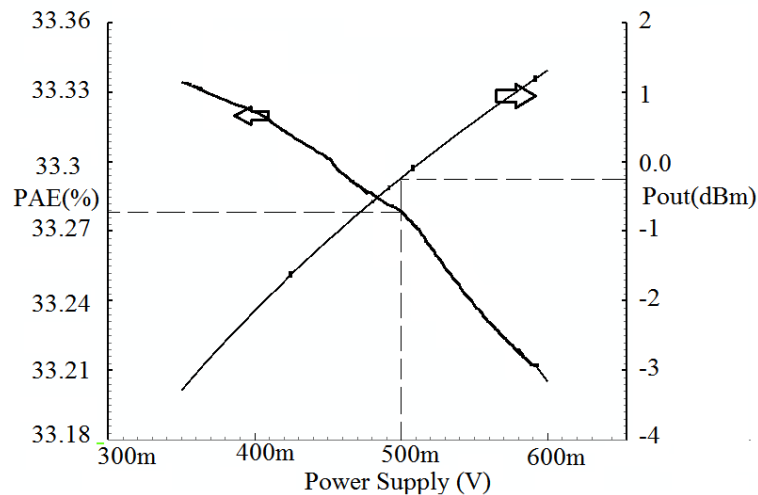


Figure 9. PAE and Pout of the proposed PA versus power supply

Table 2. Component value list for circuits of Figure 1 and Figure 6

Parameter	Circuitry of Figure 1		Circuitry of Figure 6	
Technology (μm)	0.18		0.18	
Power Supply	0.5 Volt		0.5 Volt	
Frequency	400 MHz		400 MHz	
Duty cycle	0.4		0.4	
Sub.	Inductors	capacitors	Inductors	capacitors
0	10.5 nH	1.65 pF	10.5 nH	4.2 pF
1	9.2 nH	5.52 pF	9.2 nH	9 pF
2	---	12.9 pF	---	12.9 pF
3	---	32.8 pF	---	40p
4	---	---	---	4p

Table 3. Results of the amplifiers of Fig. 1 and Fig. 6 for the component values listed in Table 1

	Frequency	Pout	PAE
Circuitry of Fig. 1	400(MHz)	-2 dBm	20%
Circuitry of Fig. 6	400(MHz)	-2 dBm	33%

5. Conclusion

In this paper, the effect of element losses on PA efficiency has been investigated. Since LPCEPA is suitable for fully integrated implementation and among the integrated elements on chip inductors have larger losses, equations for the efficiency have been derived and plotted versus the value of inductors losses. Results show that one of the inductors has larger contribution to the overall drop in efficiency. Derived equations are evaluated using MATLAB. Cadence Spectre has been used for schematic simulation. A fair matching between simulated power loss and efficiency and MATLAB evaluations can be seen from the plots. Considering the analysis, the proposed power amplifier shows about 13 % improvement in power efficiency at 400 MHz and -2 dBm output power level. The proposed PA is simulated in 0.18 μm CMOS technology.

References

Apostolidou, M. P. V., Heijden, D. M., Leenaerts, W., Sonsky, J., Heringa, A., & Volokhine, I. (2009). A 65 nm CMOS 30 dBm class-E RF power amplifier with 60% PAE and 40% PAE at 16 dB back-off. *IEEE Journal Solid-State Circuits*, 44, 1372–1379. <http://dx.doi.org/10.1109/JSSC.2009.2020680>

- Bae, L., Yan, H., & Yoo. (2011). A Low Energy Injection-Locked FSK Transceiver With Frequency-to-Amplitude Conversion for Body Sensor Application. *IEEE Journal Solid-State Circuits*, 46, 928-937. <http://dx.doi.org/10.1109/JSSC.2011.2109450>
- Brama, L., Larcher, A., Mazzanti, & Svelto. F. (2008). A 30.5 dBm 48% PAE CMOS class-E PA with integrated balun for RF applications. *IEEE Journal Solid-State Circuits*, 43, 1755-1762. <http://dx.doi.org/10.1109/JSSC.2008.925605>
- Cook, A., Berny, A., Molnar, S., Lanzisera, K., & Pister. (2006). Low-power 2.4 GHz transceiver with passive RX front-end and 400 mV supply. *IEEE Journal Solid-State Circuits*, 41, 2757-2766. <http://dx.doi.org/10.1109/JSSC.2006.884801>
- EO, H. J., Yu, S. S., Song, Y. Y., Ko, & Kim, J. Y. (2007). *A fully integrated 2.4 GHz low IF CMOS transceiver for 802.15.4 ZigBee applications* (pp. 164-167). Solid-State Circuits Conference ASSCCIEEE Asian. Jeju. <http://dx.doi.org/10.1109/ASSCC.2007.4425756>
- Lee. K. H., An, H., & Kim. (2010). Analysis and design of fully integrated high-power parallel-circuit class-E CMOS power amplifiers. *IEEE Transaction Circuits Systems*, 57, 725-734. <http://dx.doi.org/10.1109/TCSI.2009.2023944>
- Mertens, M. S. J., & Steyaert. (2002). A 700-MHz 1-W fully differential CMOS class-E power amplifier. *IEEE Journal Solid-State Circuits*, 37, 137-141. <http://dx.doi.org/10.1109/4.982419>
- Retz, H., Shanan, K., Mulvaney, S., O'Mahony, M., Chanca, P., Crowley, C., ... Quinlan. (2009). *A highly integrated low-power 2.4 GHz transceiver using a direct-conversion diversity receiver in 0.18 m CMOS for IEEE802.15.4* (pp. 414-415). WPAN. Solid-State Circuits Conference. San Francisco, CA. <http://dx.doi.org/10.1109/ISSCC.2009.4977484>
- Reynaert, M., & Steyaert. (2006). RF Power Amplifiers For Mobile Communications. New York, Springer.
- Tan, C. H., Heng, Y., & Lian. (2012). Design of Efficient Class-E Power Amplifiers for Short-Distance Communications. *IEEE Transaction ON circuit and systems*, 59, 2210-2220. <http://dx.doi.org/10.1109/TCSI.2012.2188951>
- Tsai, P. R., & Gray. (1999). A 1.9-GHz 1-W CMOS class-E power amplifier for wireless communications. *IEEE Journal Solid-State Circuits*, 34, 962-970. <http://dx.doi.org/10.1109/4.772411>

Appendix A

$$k_1 = -2q \sin(n_1) \sin(n_2) - 2 \cos(n_2) \cos(n_1) - 2q^2 (\cos(n_2) - 1)(1 - \cos(n_1)) + 2 \cos(n_1) \quad (42)$$

$$k_2 = -2 \cos(n_1) + 2 \cos(n_2) \cos(n_1) \quad (43)$$

$$k_3 = 2q \sin(n_1) \cos(n_2) - 2 \sin(n_2) \cos(n_1) \quad (44)$$

$$a_1 = 2\pi(1-D)(q+1) + \varphi \quad (45)$$

$$a_2 = 2\pi(1-D)(q-1) + \varphi \quad (46)$$

$$a_3 = 2\pi q(1-D) + \varphi \quad (47)$$

$$a_4 = 2\pi q(1-D) - \varphi \quad (48)$$

$$a_1 = 2\pi(1-D) + \varphi \quad (49)$$

$$g_1 = 2q \sin(n_1)(1 - \cos(n_1)) - 2q^2 (\cos(n_2) - 1) \sin(n_1) + \sin(n_1) \cos(n_2) \quad (50)$$

$$g_2 = 2 \sin(n_1) - 2 \sin(n_1) \cos(n_2) \quad (51)$$

$$g_3 = 2q \cos(n_2) \cos(n_1) - 2q + 2 \sin(n_2) \sin(n_1) \quad (52)$$

$$n_1 = 2\pi(1-D) \quad (53)$$

$$n_2 = 2\pi q(1-D) \quad (54)$$

Appendix B

$$J_1 = q\omega_0 A_2 V_{DD} \quad (56)$$

$$J_2 = -q\omega_0 A_1 V_{DD} \quad (57)$$

$$J_3 = \kappa\omega_0 V_{DD} \frac{q^2}{q^2 - 1} \quad (58)$$

$$J_4 = \frac{1}{2q\omega_0} \sin(4\pi qD) \quad (59)$$

$$J_5 = \frac{1}{2q\omega_0} \sin(4\pi qD + 2\varphi) - \frac{\sin(2\varphi)}{\omega_0} \quad (60)$$

$$J_6 = \frac{1}{(q+1)\omega_0} \sin((q+1)2\pi D + \varphi) + \frac{1}{(q-1)\omega_0} \sin((q-1)2\pi D - \varphi) - \frac{2q}{(q^2-1)\omega_0} \sin(\varphi) \quad (61)$$

$$J_7 = -\frac{1}{(q+1)\omega_0} \cos((q+1)2\pi D + \varphi) - \frac{1}{(q-1)\omega_0} \cos((q-1)2\pi D - \varphi) + \frac{2q}{(q+1)\omega_0} \cos(\varphi) \quad (62)$$

$$E_1 = 0.5J_1^2(TD + J_4) \quad (63)$$

$$E_2 = 0.5J_2^2(TD - J_4) \quad (64)$$

$$E_3 = 0.5J_3^2(TD + J_5) \quad (65)$$

$$E_4 = \frac{J_1^2 J_2^2}{q\omega_0} \sin^2(2\pi qD) \quad (66)$$

$$E_5 = J_1^2 J_3^2 J_6 \quad (67)$$

$$E_6 = J_2^2 J_3^2 J_7 \quad (68)$$

$$E_7 = \frac{V_{DD}}{3L_0^2} (1-D)^3 T^3 + i_{L_0}^2(t_1)(1-D)T + \frac{V_{DD}}{L_0} i_{L_0}(t_1)(1-D)^2 T^2 \quad (69)$$

$$E_8 = \frac{T}{2}(1-D) + \frac{1}{4\omega_0} (\sin(\varphi) - \sin(4\pi D + 2\varphi)) \quad (70)$$

$$E_9 = kC_{eq} V_{DD} i_{L_0}(t_1) [\sin(\varphi) - \sin(2\pi D + \varphi)] / \pi / \alpha \quad (71)$$

$$E_{10} = -C_{eq}^2 V_{DD}^2 k\omega_0 [2\sin(\varphi) - 2D\sin(2\pi D + \varphi) - \cos(\varphi) / \pi + \cos(2\pi D + \varphi) / \pi] \quad (72)$$

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

BICEP2 and the Gravitino Mass: The Questionable Result

AC Tahan¹

¹ETHERMED™, Cambridge, MA, USA

Correspondence: AC Tahan, ETHERMED, PO Box 391987, Cambridge, MA 02139, USA. E-mail: actahan@ethermed.com

Received: April 30, 2014

Accepted: June 23, 2014

Online Published: August 8, 2014

doi:10.5539/mas.v8n5p30

URL: <http://dx.doi.org/10.5539/mas.v8n5p30>

Abstract

The gravitino has been explained to exist with a low mass (Tahan 2013). BICEP2 results support a mass close to 10^{14} GeV. Though the BICEP2 conclusions have been celebrated by numerous physicists with fewer scientists indicating problems with the data, $r = 0.20_{-0.05}^{+0.07}$ from BICEP2 (BICEP2 Collaboration 2014) can be understood improbable by simply focusing on the mass of the gravitino. This manuscript presents a method involving increasing string vibrations to measure the gravitino mass in the laboratory.

Keywords: gravitino, BICEP2, dark matter, string experimentation

1. Background

In attempts to understand how gravity fits in the Standard Model the gravitino has been an unavoidable subject since it is a superpartner of the graviton, specifically regarding supergravity analogous to electroweak theory W and Z bosons. Accordingly, the gravitino has been well-studied including for troubling Universe topics as dark matter, which has forced the issue of the superparticle mass if wishing to study gravity in the evolution of the Universe. Inflation emerged as the most popular scenario for the development of the Universe and consequently the gravitino has been tied to different inflation theories. The mass of the superpartner has usually been a primary point in the ideas related to how the gravitino can exist with inflation, even having led to disfavoring gravitino dark matter because of discussions regarding how the mass could have prevented the existence of the Universe, e.g. the gravitino problem (Weinberg, 1982). Therefore, the gravitino became a less popular topic of study, particularly considering that evidence for supersymmetry or string theory has been nonexistent.

A recent report from the BICEP2 collaboration has reintroduced the possibility for the existence of the gravitino with a large mass, which has not been a primary topic in papers referencing the findings. The mass may not have been well-discussed so far since perhaps it has been considered, like other particles, an unavoidable result of the overall events suggested by the BICEP2 data. Still, the BICEP2 work discussing quantum gravity requires clarification for the existence of a high mass gravitino, particularly considering past studies. By considering the mass in relation to work performed with an instrument that permitted graviton appearance in the lab (Tahan 2011) while remembering theoretical studies of the gravitino, the high mass alone would suggest that the BICEP2 findings are questionable.

1.1 Introduction

Though the BICEP2 team has not presented all data and the $r = 0.20_{-0.05}^{+0.07}$ result is inconsistent with Wilkinson Microwave Anisotropy Probe (WMAP) satellite (Hinshaw et al., 2013) and Planck satellite information (Planck Collaboration, 2014) thus far, the BICEP2 report has been received predominantly with great fanfare. The BICEP2 conclusions support quantum gravity, which again would suggest the existence of the gravitino. However, according to the BICEP2 findings the gravitino mass would have to be about 10^{14} GeV, which was less acceptable before the BICEP2 report due to reasons including a large mass and decay products possibly influencing overclosure of the Universe (Jeong & Takahashi, 2012) or Big Bang Nucleosynthesis (BBN) (Kawasaki, Kohri, Moroi, & Yotsuyanagi, 2008).

Apart from scholars using the BICEP2 results to support their work, the majority of subsequent papers have discussed how the visible Universe could exist involving inflation in consideration of the data, though researchers have written that *B*-mode polarization is not necessarily evidence for inflation (Brandenberger, 2011). Few publications have addressed the limitations of the BICEP2 instrumentation, e.g. compared to the Planck satellite regarding operating frequencies. Rarely have scientists scrutinized the data, including specific points on

graphs that may not fit explanations. Little criticism has been presented regarding unusual methods the BICEP2 team may have used to analyze or compare data particularly to information from other projects, and minimal discussion has existed regarding what the BICEP2 team may have overlooked, e.g. the Wollenbein New Loop (Wollenbein, 2007; Liu, Mertsch, & Sarkar, 2014). Papers indicating that BICEP2 data diverge from theory, for instance not allowing string theory and inflation to coexist (Sakellariadou 2014), have not expressed the results to be worrisome but instead have explained the discrepancies to mean that BICEP2 has shown that events in the progression of the Universe need to be modified to allow for BICEP2 findings. And the gravitino seldom has been mentioned; manuscripts suggesting that the high mass makes the BICEP2 results questionable are uncommon. Instead scientists have proposed situations to allow for a large gravitino mass (Fan, Jain, & Ozsoy, 2014).

The inability of synchrotron teams to find supersymmetry suggests a broken symmetry, permitting more massive superparticles than related Standard Model bodies. Prior to the BICEP2 conclusions certain symmetry breaking theories permitting near apart from significantly above 1TeV gravitinos were disfavored (Jeong, Park, Stewart, & Kadota, 2004, Endo, Hamaguchi, & Takahashi, 2006), additional references could have been presented. Manuscripts have not properly clarified what type of supersymmetry breaking would have allowed for events including the massive gravitino, i.e. particular scenarios conflict discounted proposals or ideas in different papers are contradictory or incompatible.

2. Tool for Explorations beyond the Standard Model

A tool for explorations beyond the Standard Model exists that has allowed for the visualization of an open string and D-brane (Tahan, 2011). Results supported a sub-keV gravitino mass, being the lightest superparticle (LSP) involving minor violation of R-parity and potentially dark matter that includes fifth dimension primordial black holes (Tahan, 2013). With inclusion of a slight violation of R-parity the gravitino LSP has a duration beyond the lifetime of the Universe (Takayama & Yamaguchi, 2000) and primordial black holes and the small massed gravitinos would be cold dark matter in consideration of entropy (Lemoine, Moulta, & Jedamzik, 2007). Additionally, work with the Figure 1 set-up that showed the prevalence of branes supported the existence for a division of an underlying superparticle and graviton area and the Standard Model visible sector.

Conclusions due to use of the Figure 1 set-up force the BICEP2 results in consideration of the gravitino mass to be considered highly improbable and suggest that a method for determining the validity of the BICEP2 work would be to use the Figure 1 technique to understand the mass of the gravitino. The symmetry breaking due to the set-up was periodic. By that, multiple Hydrogen branes were influenced during approximately hour long trials so that by surrounding the set-up with proper detectors, several opportunities for gravitino mass measurements would arise per trial.

The gravitino weakly couples to matter thereby making direct detection doubtful at colliders. Researchers preferring late decay in the Cosmos especially involving the LSP gravitino have proposed discovery through next to lightest superparticle (NLSP) decay. NLSP decay to gravitinos would be unlikely if the gravitino is the LSP with slight violation of R-parity. The NLSP would be ephemeral and decay straight to R-parity even particles (Takayama & Yamaguchi, 2000), lowering the possibility for gravitinos to be found at synchrotrons thereby suggesting that solely the Figure 1 procedure could permit gravitino detections including perhaps other superpartners if incorporating the right detectors with the set-up. Also, this work should not be considered to disfavor inflation. Describing small massed gravitinos as the LSP and cold dark matter support a relationship between supersymmetry and inflation (Tahan, 2013). This manuscript is intended only to suggest that disagreement regarding the accuracy of BICEP2 findings, particularly if upcoming data as from the Planck satellite do not resolve questions related to the results, possibly can be settled with a direct measure of the gravitino mass using the Figure 1 set-up. Though regardless of future collaboration reports, only the Figure 1 instrument allows for specific studies and appearances, e.g. the direct observations of strings (Tahan, 2011).

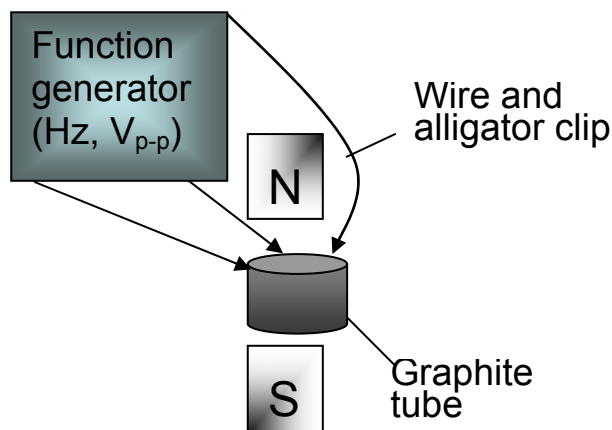


Figure 1. Basic apparatus

Laser light (Quartet Standard Laser Pointer), $\approx .017\text{m}$ from the middle and about $.01-.02\text{m}$ from the bottom of a graphite tube creating an approximately $.004-.005\text{m}$ line on the side of the tube closest to the S pole of a gap magnet, for the D-brane trials was removed for Figure 1, which is the basic set-up. Incorporation of the laser was to examine the possibility of the bending of spacetime resulting from gravitons emerging due to symmetry breaking (Tahan, 2012) and interacting with the tube, imparting mass-energy thereby causing the tube to bend spacetime sufficiently to alter the direction of the light. Sulfuric acid (H_2SO_4) (typically $20.\text{mL}$, 96% concd. Mallinckrodt Analytical Reagent, ACS) was in the tube for each experiment. Approximately 99.7kPa and 98-100% were the barometric pressure and relative humidity (Boston, MA, CW1378) during the largest laser light D-brane trial (Tahan, 2011). The tube was clamped in the static magnetic field, N and S being the poles of the gap magnet, measuring about 2000Gs. Small distance variations toward the S pole were immaterial: $\approx .004-.012\text{m}$ should allow for observations. Wires from a breadboard to which the function generator was attached to a LED supplied $\approx 2\text{Hz}$ ($2.000-2.012\text{Hz}$) and $V_{p-p}=4.312-4.437$, predominately 4.375, to the graphite container. The low frequency was not intended to alter Hydrogen alignment in the magnetic field. The set-up was ineffective when the magnet was removed. One wire with an alligator clip is required; three wires were connected triangularly to the top of the tube to be sure that the frequency and amplitude were being supplied--in view of potential inoperability in an experiment as because of clip corrosion from acid exposure. The graphite tube, clamp that held the tube, and the stand to which the clamp was attached that included the $\approx .153\text{kg}$ clamp holder were $\approx .098-.099\text{kg}$, near $.230\text{kg}$, and about 1.458kg respectively. The set-up requires the frequency and amplitude, tube (Crucible, Saed/Manfredi G40, $1.5''\text{OD} \times 1.25''\text{ID} \times 3.75''\text{DP}$), constant magnetic field strength ($\approx 2000\text{Gs}$), and concentrated sulfuric acid--the only components always needed. Manners for supplying the parts can be tested, for example Hydrogen from a different source than H_2SO_4 . But modifications to the components may be needed depending on the change: e.g. the amplitude may require adjustment. When utilizing only one wire, if the set-up is inoperable and the clip is well-attached and not requiring replacement, especially if a researcher had attached the three wires dissimilarly modification of the amplitude may be necessary, imagining the three wires to have supplied quanta differently for suitable string vibrations. The graphite tube and acid supply should be changed occasionally to avoid variables, for instance impurities in the acid or eventual deformation of the tube due to the acid that may inhibit regular frequency conduction. The acid should be replaced in the tube for each trial unless the examiner wishes to test a particular aspect of the set-up involving the acid. The Figure 1 apparatus permits periodic appearances of superpartners and gravitons that can be presented similarly to a reaction: gluon, gluon, gluon, gluino, gluino, gluino, photon, proton, electron, W^- boson, Higgs boson, Z boson, axion, dilaton \rightarrow graviton, gravitino, neutron, electron neutrino, graviscalar, graviphoton, photon, photon, photon, sneutrino, selectron, saxion (Tahan, 2012).

Use of the set-up allowed for the writing of the reaction in the Figure 1 description that was utilized for a calculation 125.81GeV of the Higgs Boson mass (Tahan, 2012). After learning that the calculated mass was not dissimilar from collider discoveries (Incandela, 2012; Gianotti, 2012), in view of the reaction the gravitino was

thought should not be greater than 1keV, also while considering theoretical work instructing that gravitinos with a larger mass would have overclosed the Universe (Cho & Uehara, 2004)--appreciating the superparticle as the LSP therefore avoiding issues with BBN. The gravitino being a light superpartner, particularly the LSP, would support the body as dark matter (Giudice & Rattazzi, 1999; Takayama & Yamaguchi, 2000). The reaction can be utilized to predict bodies and rewritten to contain or remove particles if understood required primarily after using suitable detectors with the set-up.

3. String Experimentation

Prior to work with the Figure 1 set-up, strings were unknown to exist and scientists believed that the predicted, required high energies to observe strings made the possibility of related work unlikely. The BICEP2 report being supportive of quantum gravity, i.e. the graviton, would be suggestive of the existence of strings since quantum gravity is well-accepted among theorists as a string theory (Scherk & Schwarz, Phys. Letters, 1974). The Figure 1 method is the only tool that permits the direct visualization of strings, including allowing for the guiding of gravitons to locations that results in the gravitons imparting mass-energy that occurred in the lab for the Figure 1 graphite tube, causing the curving of spacetime and resultantly the bending of the laser light around the tube (Tahan, 2011).

For experiments, the initial hypothesis for use with the Figure 1 apparatus that led to gravitons in the laboratory concerned principles of string theory regarding Hydrogen. The method involved 2Hz with a particular amplitude to create vibrations of the Hydrogen brane that resulted in increasing energy and eventually symmetry breaking that was periodic in the concentrated H₂SO₄ supply for Hydrogen, i.e. symmetry breaking for various Hydrogen happening at distinct times in the usual roughly one hour experiments. Specifically, the foundation of Hydrogen (protons) is the string, principally gluon field strings, with tension that can be altered with a magnetic field thereby permitting increased vibrations due to electromagnetic energy of the proper frequency and amplitude in consideration of the manipulated tension.

The escalating vibrations of a string are identical to an increasing energy (Scherk & Schwarz, Nucl. Phys. B, 1974). The 2Hz was useful mainly due to influencing gluon field strings: a strong force string underlying a uniform field of quanta (Yang & Mills, 1954). The field strength (coupling constant) between quarks is consistent (Mills, 1994); as energy increased the coupling constant decreased (Aharony, Gubser, Maldacena, Ooguri, & Oz, 2000). Hydrogen maintaining alignment in the magnetic field caused the tension of the Hydrogen brane and associated strings to stay continuous thus allowing for growing vibrations. Consequently ultimately the proton, involving electron capture (Tahan, 2012), fell apart thereby permitting the glueball and symmetry breaking. In other words, if appreciating Yang-Mills theory to be a quantum gauge theory with the gauge group SU(N) being equivalent to a string theory with 1/N as the string coupling constant ('t Hooft, 1974, Maldacena 1998) the strong force coupling constant steadily reached the limit 1/3, a large N analogous to a vibration limit (Aharony, Gubser, Maldacena, Ooguri, & Oz, 2000), because of the incessant 2Hz quanta that resulted in deconfinement. Namely, the coupling constant can be considered inversely proportional to an energy that corresponds to the gluon field. Increasing string vibrations result in a gluon field advancing toward a limit infinity ('t Hooft, 1974). At high energy due to approaching the limit, the coupling constant became zero that permitted liberation of gluon strings. The glueball occurring from the coalescing of the strings before the massive superpartners of the graviton (Tahan, 2012) suggested that confinement and the Higgs mechanism are attributes of the strong force string (Fradkin & Shenker, 1979).

Using 2Hz with a particular amplitude could allow the Figure 1 method to be thought a low energy technique. However, what should be remembered is that a string is energy. Specifically, a problem for scholars has been to define energy, which is resolved by equating energy to a string that becomes a greater quantity of energy the more it vibrates. The 2Hz was a resonant frequency for the altered strings of the Hydrogen in the magnetic field, creating steadily growing string vibrations thereby increasing energy that influenced the structure of the Hydrogen. The required point at which the symmetry breaking process occurred was reached as the strings vibrated that can be visualized with the involved particles of the Figure 1 reaction, which again can be edited with future studies using the instrument. By that, the needed high energy for deconfinement was provided by the Hydrogen, videlicet its strings.

The ease of set-up permits testing predictions inexpensively and in a typical laboratory setting. Thus, possible due to the Figure 1 apparatus string experimentation is the practical use of string resonance. Through the manipulation of strings the Figure 1 method can lead to globally needed applications including using the tool for research, for example to manufacture gravitons and superpartners as the gravitino.

4. Conclusions

This manuscript emerged due to concern regarding how quickly the BICEP2 report was applauded and accepted in the scientific community (Cowen 2014), though again a few scientists--particularly experimentalists--have expressed displeasure regarding BICEP2 data. The acknowledgment is not immensely surprising if considering Large Hadron Collider (LHC) results.

Physicists are bothered that LHC teams have not presented more discoveries especially beyond the Standard Model (Grant, 2013), to the extent of a panic that would explain swift approval of the BICEP2 information that was unforeseen. Specifically, a theorist makes predictions consistent with findings from experiments. Unexpected data provide more opportunities for explanations. And though the Higgs boson result has been celebrated disappointment seems to be the main product for the first years of the LHC, from which the conclusions were anticipated.

With the emergence of string experimentation due to the Figure 1 technique frustration or worries regarding the future of theoretical physics or physics studies generally should not exist. The method is a low-cost, facile set-up and permits theoretical work and experiments on a lab bench (Tahan, 2012). Collider teams not having detected supersymmetry should require the instrument to be accepted at least alternatively for string and superpartner examinations, as to measure the gravitino mass, since it is the only tool that manifests superparticles apart from strings: for instance the open string and D-brane of the laser light when part of the Figure 1 set-up (Tahan, 2011). For improved opportunities to observe D-branes with open strings pristine graphite tubes should be used since more graphite will leak to the acid, where gravitons emerge and will interact with the atoms thereby creating larger high gravitational areas that will greater influence the light (Tahan, 2011; Tahan, 2013). Therefore, instead hope should subsist if the apparatus is appreciated as the viable tool for studies outside the Standard Model.

References

- Aharony, O., Gubser, S. S., Maldacena, J., Ooguri, H., & Oz, Y. (2000). Large N field theories, string theory and gravity. *Phys. Rep.*, 323, 183-386. [http://dx.doi.org/10.1016/S0370-1573\(99\)00083-6](http://dx.doi.org/10.1016/S0370-1573(99)00083-6)
- BICEP2 Collaboration, P. A. R. Ade et al. (2014). BICEP2 I: Detection of B-mode polarization at degree angular scales, arXiv:1403.3985 [astro-ph.CO].
- Brandenberger, R. H. (2011). Is the spectrum of gravitational waves the "Holy Grail" of inflation? arXiv:1104.3581v1 [astro-ph.CO].
- Cho, G. C., & Uehara, Y. (2004). Cosmological gravitino problem confronts electroweak physics. *Phys. Rev. D*, 69, 075003. <http://dx.doi.org/10.1103/PhysRevD.69.075003>
- Cowen, R. (2014). Telescope captures view of gravitational waves. *Nature*, 507, 281-283. <http://dx.doi.org/10.1038/507281a>
- Endo, M., Hamaguchi, K., & Takahashi, F. (2006). Moduli-induced gravitino problem. *Phys. Rev. Lett.* 96, 211301. <http://dx.doi.org/10.1103/PhysRevLett.96.211301>
- Fan, J. J., Jain, B., & Ozsoy, O. (2014). Heavy gravitino and split SUSY in the light of BICEP2, arXiv:1404.1914v1.
- Fradkin, E., & Shenker, S. H. (1979). Phase diagrams of lattice gauge theories with Higgs fields. *Phys. Rev.*, 19D, 3682-3697. <http://dx.doi.org/10.1103/PhysRevD.19.3682>
- Gianotti, F. (2012). Update on the Standard Model Higgs searches in ATLAS, CERN Seminar.
- Giudice, G. F., & Rattazzi, R. (1999). Theories with gauge-mediated supersymmetry breaking. *Phys. Rept.*, 322, 419-499. [http://dx.doi.org/10.1016/S0370-1573\(99\)00042-3](http://dx.doi.org/10.1016/S0370-1573(99)00042-3)
- Grant, A. (2013). Hard times for theorists in a post-Higgs world. *Science News*, 183(13), 22-25. <http://dx.doi.org/10.1002/scin.5591831323>
- Hinshaw, G. et al. (2013). Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Cosmological parameter results, arXiv:1212.5226v3 [astro-ph.CO].
- Incandela, J. (2012). Update on the Standard Model Higgs searches in CMS, CERN Seminar.
- Jeong, D., Park, W. I., Stewart, E. D., & Kadota, K. (2004). Modular cosmology, thermal inflation, baryogenesis and a prediction for particle accelerators. *JHEP*, 11, 046. <http://dx.doi.org/10.1088/1126-6708/2004/11/046>
- Jeong, K. S., & Takahashi, F. (2012). A Gravitino-rich Universe, TU-921, arXiv:1210.4077v1 [hep-ph].
- Kawasaki, M., Kohri, K., Moroi, T., & Yotsuyanagi, A. (2008). Big-Bang nucleosynthesis and gravitino. *Phys.*

- Rev.*, 78D, 065011. <http://dx.doi.org/10.1103/PhysRevD.78.065011>
- Lemoine, M., Moultaqa, G., & Jedamzik, K. (2007). Natural gravitino dark matter in SO(10) gauge mediated supersymmetry breaking. *Phys. Lett.*, 645B, 222-227. <http://dx.doi.org/10.1016/j.physletb.2006.12.025>
- Liu, H., Mertsch, P., & Sarkar, S. (2014). Fingerprints of Galactic Loop I on the Cosmic Microwave Background, arXiv:1404.1899v1 [astro-ph.CO].
- Maldacena, J. (1998). The large N limit of superconformal field theories and supergravity. *Adv. Theor. Math. Phys.*, 2, 231-252.
- Mills, R. L. (1994). *Space, time, and quanta: An introduction to contemporary physics, Part III*. New York: W.H. Freeman.
- Planck Collaboration, P. A. R. Ade et al. (2014). Planck 2013 results. XVI. Cosmological parameters, arXiv:1303.5076v3 [astro-ph.CO].
- Sakellariadou, M. (2014). How well do we understand the thermal history of the Universe? Implications of the recent BICEP2 data, KCL-PH-TH/2014-9, arXiv: 1403.6688v1 [astro-ph.CO].
- Scherk, J., & Schwarz, J. H. (1974). Dual models and the geometry of space-time. *Phys. Letters.*, 52B, 347-350. [http://dx.doi.org/10.1016/0370-2693\(74\)90059-8](http://dx.doi.org/10.1016/0370-2693(74)90059-8)
- Scherk, J., & Schwarz, J. H. (1974). Dual models for non-hadrons. *Nucl. Phys.*, 81B, 118-144. [http://dx.doi.org/10.1016/0550-3213\(74\)90010-8](http://dx.doi.org/10.1016/0550-3213(74)90010-8)
- 't Hooft, G. (1974). A planar diagram theory for strong interactions, *Nucl. Phys.*, 72B, 461-473. [http://dx.doi.org/10.1016/0550-3213\(74\)90154-0](http://dx.doi.org/10.1016/0550-3213(74)90154-0)
- Tahan, A. C. (2011). Exposing strings in the laboratory with a novel technique. *Appl. Phys. Res.* 3(2), 39-51. <http://dx.doi.org/10.5539/apr.v3n2p39>
- Tahan, A. C. (2012). Diagrammatic presentation for the production of gravitons and supersymmetry. *Mod. Appl. Sci.*, 6(9), 76-83. <http://dx.doi.org/10.5539/mas.v6n9p76>
- Tahan, A. C. (2013). Low mass gravitino: Re-introducing the superpartner as dark matter with consideration to inflation due to experimentation. *Mod. Appl. Sci.*, 7(12), 43-55. <http://dx.doi.org/10.5539/mas.v7n12p43>
- Takayama, F., & Yamaguchi, M. (2000). Gravitino dark matter without r-parity. *Phys. Lett.*, 485B, 388-392. [http://dx.doi.org/10.1016/S0370-2693\(00\)00726-7](http://dx.doi.org/10.1016/S0370-2693(00)00726-7)
- Weinberg, S. (1982). Cosmological constraints on the scale of supersymmetry breaking. *Phys. Rev. Lett.*, 48, 1303-1306. <http://dx.doi.org/10.1103/PhysRevLett.48.1303>
- Wolleben, M. (2007). A new model for the Loop-I (The North Polar Spur) Region. *Astrophys. J.*, 664, 349-356. <http://dx.doi.org/10.1086/518711>
- Yang, C. N., & Mills, R. L. (1954). Conservation of isotopic spin and isotopic gauge invariance. *Phys. Rev.*, 96, 191-195. <http://dx.doi.org/10.1103/PhysRev.96.191>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

The Economic Value of Laem Phak Bia Mangrove Ecosystem Services in Phetchaburi Province, Thailand

Sitthinan Wiwatthanapornchai¹, Chucheep Piputsitee² & Samakkee Boonyawat¹

¹ Faculty of Environment, Kasetsart University, Bangkok 10900, Thailand

² Faculty of Economics, Kasetsart University, Bangkok 10900, Thailand

Correspondence: Sitthinan Wiwatthanapornchai, Faculty of Environment, Kasetsart University, Bangkok 10900, Thailand. Tel: 66-825-634-888. E-mail: wiwatt_s@hotmail.com

Received: May 6, 2014

Accepted: June 9, 2014

Online Published: August 6, 2014

doi:10.5539/mas.v8n5p36

URL: <http://dx.doi.org/10.5539/mas.v8n5p36>

Abstract

Thailand is a one of the nation in Southeast Asia, covered by numerous mangrove areas approximately 244,000 hectares. Phetchaburi province is the one of the province in Thailand where the mangrove area has been increasing continually since King's Royally Initiated Laem Phak Bia Environmental Research and Development Project has been set up. The mangrove ecosystems functions are vital to the livelihood of the surrounding community. Laem Phak Bia community is one that has been served from mangrove ecosystem services. This study assessed the economic value of Laem Phak Bia mangrove ecosystem services in an area approximately 237.44 hectares using Participatory Economic Valuation (PEV) by the villagers ranking and rating the importance of mangrove ecosystem services with a valuable thing for living that is the rice value. The results showed that this mangrove area was worth a total economic value about 100 million Baht per year or 424 thousand Baht per hectares per year. It could be divided into the value of regulation functions, production functions, habitat functions and information functions, which were about 38, 8, 25, and 29 million Baht per year, respectively.

Keywords: ecosystem services, participatory economic valuation, mangrove, Thailand

1. Introduction

Nowaday, There is approximately 244,000 hectares mangrove area in Thailand. Almost area is in the south. The value of the mangrove from ecosystem services are ecosystem functions that are directly beneficial to humans. Moreover, it is also important for the survival of aquatic plants, animals and humans. Value of the mangrove ecosystem can be classified as the direct use value, which is as a timber and non-timber usage or site for recreation, and as the indirect use value that is the role of nature such as major habitats and food source for fish or shrimp, and the value of coastal protection from wind or storm. Not only use value, mangrove forest has non-use value, which reflects the value of the satisfaction of the society for the existence of mangrove area, or heritage for descendants.

In Thailand, economic valuation of mangrove area has many case studies such as in Phang Nga Bay, in Pak Phanang, in Southern of Thailand, or in Krabi River Estuary. Almost studies assessed economic value in term of direct use, indirect use and non-use value and evaluated value by using market value or willingness to pay. The economic value of mangrove forest was estimated as a total value that could not be classified on the mangrove ecosystem functions and services (Sathirathai, 1998; Sathirathai & Edward, 2001; Edward et al., 2002; Seenprachawong, 2002; Islam & Ikejima, 2010; Janekarnkij, 2010).

Phetchaburi Province is the one of Thailand that the mangrove area has dramatically increased since His Majesty the King's Initiative Laem Pak Bia environmental research and development project (LERD) has been launched in 1990 and located on Laem Phak Bia Sub-district, Ban Laem District. The mangrove area has increased steadily up to 64 hectare or has an average of growth rate up to 3.71 hectares per year. (Chunkao & Nimpee, 2012) There are four villages located nearby Laem Phak Bia mangrove area and they get the services from mangrove ecosystems.

The objectives of this study are to (1) demonstrate the role of mangrove ecosystem functions and services and (2) estimate mangrove ecosystems economic values. This value is the reflect of the importance of the mangrove

ecosystem services for the communities. The estimated value can be used as a guideline for policy makers to make an alternative decision-making use and preserve the mangrove natural resource in term of sustainable development for the future.

2. Method

2.1 Study Area

LERD has been established since 1990 at Laem Phak Bia sub-district, Ban Laem district, Phetchaburi province with the coverage area 3.24 square kilometers. It is placed on between the latitude of 13° 02' 40" to 13° 03' 20" N and longitude of 100° 05' 10" to 100° 06' 05" E (UTM 1442240 to 1443480 N and 0619271 to 0619271 E). Laem Phak Bia Mangrove has area approximately 237.44 hectares as shown in Figure 1.

From secondary data collected by Department of Provincial Administration, Thailand, in Laem Phak Bia sub-district, there are four villages which are Moo 1 Ban Paneun, Moo 2 Ban Donnai, Moo 3 Ban Donklang, and Moo 4 Ban Donkadee. In 2013, the local population is 2,378 with 968 Households. Livelihood of people living in coastal villages relies mainly on fishery and laboring in the fishery and tourism business (in the short term per year) and these area villagers do not utilize neither timber nor non-timber product of mangrove forest (Table 1).

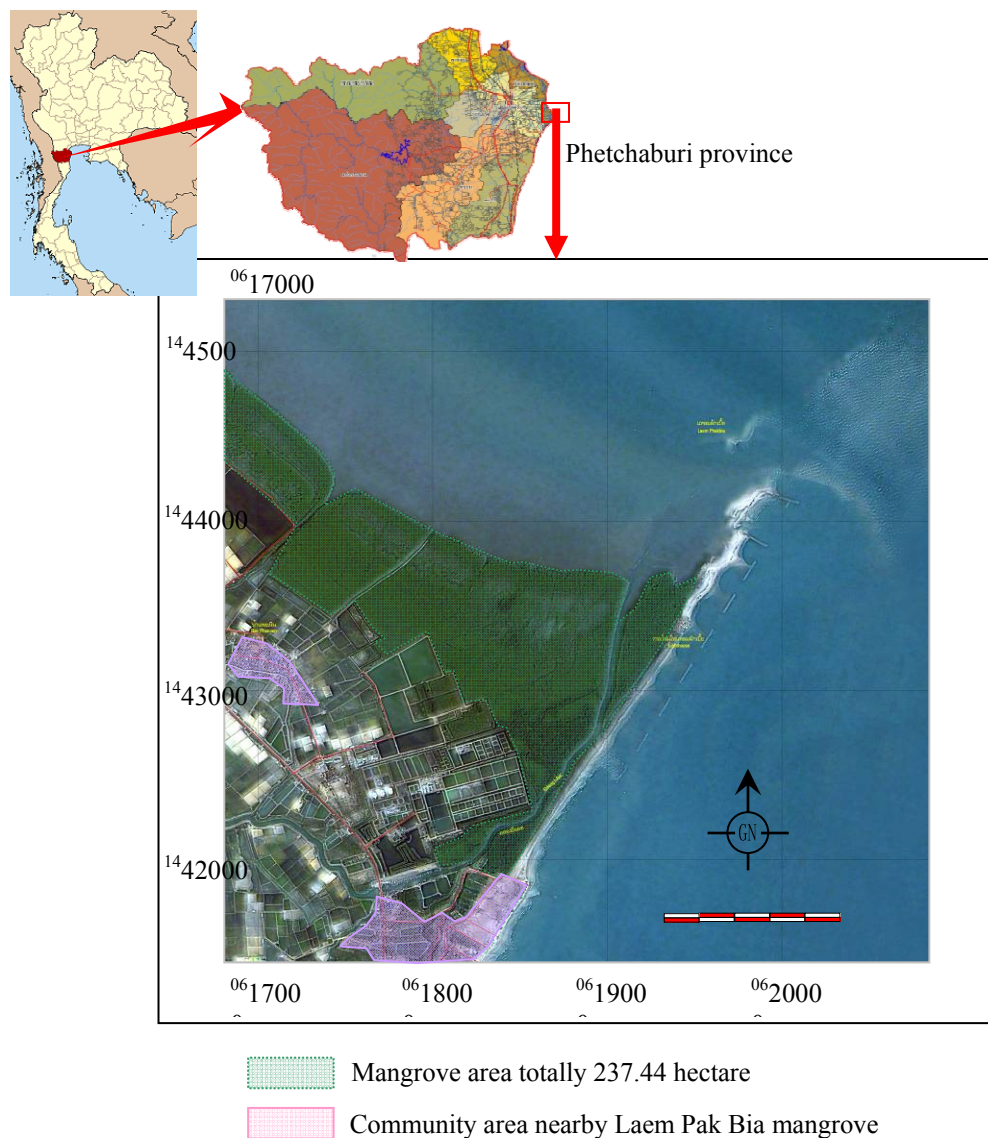


Figure 1. Map of Laem Pak Bia Mangrove and community around mangrove area (community target)
 Source: Jitthaisong et al. (2012).

Table 1. Population and number of household nearby Laem Phak Bia mangrove and resource utilization from mangrove forest in 2013

Village	Population	No. of Household	Major types of resource utilization			
			Timber / Non-timber	Fishery	Labor in Fishery	Tourism
Moo 1 Ban Paneun	889	227	X	√√√	√√	√
Moo 2 Ban Donnai	346	108	X	√√√	√√	√
Moo 3 Ban Donklang	479	422	X	√√√	√√	√
Moo 4 Ban Donkadee	664	211	X	√√√	√√	√
Total	2,378	968				

Source: Department of Provincial Administration, Thailand (2013).

2.2 Ecosystem Functions and Ecosystem Services

Mangrove ecosystem functions are the capacity of mangrove natural processes and components to provided goods and services that satisfy human or social needs directly and indirectly (Costanza and Folke, 1997; De Groot, 1992, 1994; De Groot et al., 2002). Ecosystem functions can be grouped into four primary categories, including (1) regulation functions that relate to the capacity of natural and semi-natural ecosystem to regulate essential ecological processes and life-support systems such as gas regulation, climate regulation, waste treatment, carbon storage or nutrient regulation (2) production functions that provide many ecosystem goods and services for human consumption such as wood, food, raw material or energy resource (3) habitat functions that provide refuge and reproduction habitat to wild plants and animals and thereby contribute to the biodiversity or genetic diversity conservation and (4) information functions that provide the opportunity to maintenance human health and spirit by the scenery and landscape for human enjoyment or recreation and aesthetic experience.

Ecosystem services are the benefits provided by ecosystem to humans that contribute to making human life both possible and worth living. Leam Pak Bia mangrove has many ecosystem services include (1) provisioning services like timber and non-timber product on forest, fisheries or fresh air; (2) supporting services that are a natural process such as primary production or nutrient for plants or aquatic animals; (3) regulating services such as air and climate regulation, wastewater treatment or coastal erosion control from wind and (4) cultural services that the community can use for recreation or education.

This study evaluates mangrove ecosystem functions and ecosystem services by assessment from the important level of mangrove ecosystem functions and ecosystem services that provide to the communities. There are 5 important levels, starting from 1 which is the least important to 5 which is the most important. Then, the average of the important level was analyzed and the result was interpreted by comparing the average value with the important rank, including (1) score 5.00 – 4.21 is the most importance; (2) score 4.20 – 3.41 is the importance; (3) score 3.40 – 2.61 is the moderate; (4) score 2.60 – 1.81 is the less importance; and (5) score 1.80 – 1.00 is the least importance.

2.3 Mangrove Values

Laem Phak Bia mangroves have many tangible and intangible values. In term of economic value, mangrove values are assessed from an importance of goods and services that provided to the communities. Total economic values of mangroves include use values and non use values (Figure 2). Use values or values in use is the utility of consuming a goods and services from mangroves that divided into three types which are direct use values, indirect use values and option values. Direct use values are all goods and benefits of using mangrove resources directly to produce timber and non-timber products and non commodity benefits such as recreation, education, or research in mangrove forest. Indirect use values are the environment services that value to the community such as nursery and habitat, coastal protection, wastewater treatment, or carbon storage. Option values are the current use of mangrove forest land saved for future use. Non use values or passive values are bequest values, which are the value of keeping mangrove resource intact for future generations and existence values, which are the intrinsic value of the mangrove resource (Costanza & Folke, 1997; De Groot, 1992, 1994; De Groot et al., 2002; Pearce et al., 1992).

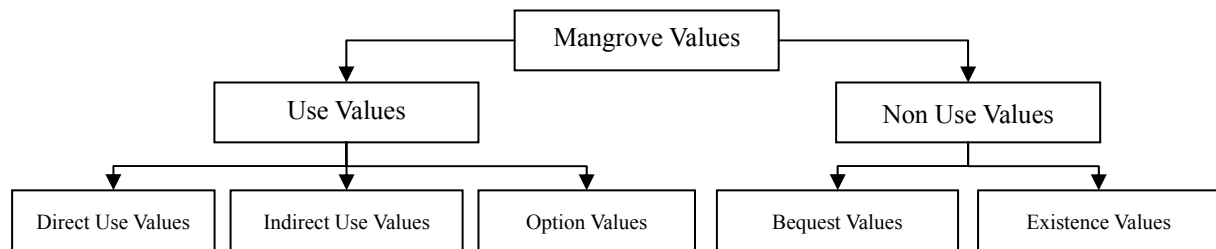


Figure 2. Classification of mangrove value

2.4 Participatory Economic Valuation (PEV)

Economic valuation method is a tool to allocate quantitative values to the goods and services provided from natural resources or mangrove ecosystem services (Costanza & Folke, 1997; Daily, 1997; Daily et al., 2000; Quoc et al., 2012). Economic valuation of mangrove ecosystem services can be useful in indicating the opportunity cost of other land-use practices, and it can help in land-use decision-making. In economic valuation of natural resources, there are several methods such as market value approach, travel cost method, hedonic pricing, contingent ranking, or contingent valuation method. The best way to evaluate natural resource is the contingent valuation method that can assesses both use value and non use value via willingness to pay or willingness to accept. However, the weakness of this method is that it can not distribute value by ecosystem functions. The limitations of conventional valuation suggest that it might be a good thing to encourage active involvement of villagers in the economic valuation of their mangrove forests. However, participatory approach is a new technique in land use development in developing countries.

Participatory Environmental Valuation (PEV) is the one of the economic valuation method that uses hypothetical market. PEV is a technique to compare the value of the natural resources relies on the value of items that are valued in the market and important to survival in society (Emerton & Mogaka, 1996; Kuchelmeister, 2003; Rosale et al., 2003; IUCN, 2005, 2006, 2007).

In the overall, Participatory Environmental Valuation (PEV) has advantages over other conventional valuation approaches which are: (1) PEV fully utilizes the data and knowledge from local decision-makers and villagers; (2) PEV maximizes the involvement and feedback of these groups; (3) PEV raises the awareness about the importance of mangrove valuation and knowledge of how to use and conserve it properly; (4) PEV can identify total value of mangrove and distribute in each ecosystem services of mangrove; (5) PEV eliminates subjective bias that make overestimate value because PEV use value of item in the communities to compare with the value of mangrove resources.

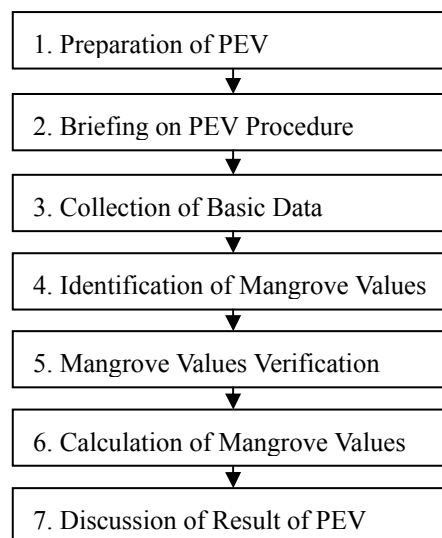


Figure 3. Participatory Economic Valuation (PEV) procedures

Participatory Economic Valuation (PEV) has 7 steps to assess the value of mangrove ecosystem services (Figure 3). First step is to prepare of background data by collecting the secondary data in term of yield and market data in the communities. Second step is to brief the villager about the purpose, scope and the expected results of PEV and to collect the primary village data including mangrove resources utilization. Then, proceed to the next steps of PEV by collecting the primary village data and identifying mangrove values by interview key stakeholders in the communities. This step will get the first overview of the mangrove values the villagers perceive, the major direct benefits, mangrove ecosystem services, optional values, and non-use values in the village and beyond. To reconfirm the identified mangrove forest values, the sited are visited, and the physical impact area is collected in step 5. Then, step 6 is calculating the direct benefits and costs of predominant forest models in the village, subsequently also ecosystem services, optional values and non-use values. Finally in the last step, the results of the PEV exercise are discussed at a village meeting and recommendations for further actions are made by the villagers.

The calculation of the mangrove ecosystem service value performed in step 6 is the mean of proportion of importance of mangrove ecosystem service and importance of value item in the community. In this study, rice is referred to the value item that all villagers will consume it for their living, then plus with the value of rice as following equation;

$$\text{Value of MES} = \sum_i (\text{Value of MES}_{x_1}, \text{Value of MES}_{x_2}, \dots, \text{Value of MES}_{x_n})$$

$$\text{Value of MES}_{x_n} = \text{Average} \left(\frac{\text{IMES}_{x_{n1}}}{\text{IR}_1} \times \text{RV}_1, \frac{\text{IMES}_{x_{n2}}}{\text{IR}_2} \times \text{RV}_2, \dots, \frac{\text{IMES}_{x_{nm}}}{\text{IR}_m} \times \text{RV}_m \right)$$

Whereas	Value of MES	: Value of Mangrove Ecosystem Services
	Value of MES _{x_i}	: Value of Mangrove Ecosystem Services in type x _i
	IMES _{x_ij}	: Important Score of Mangrove Ecosystem Services in type x _i of person j
	IR _j	: Important Score of Rice for living of person j
	RV _j	: Value of Rice for living (Baht/Year)
	i	: 1,2, ..., n
	j	: 1,2, ..., m

3. Results and Discussion

3.1 Important rank of Leam Pak Bia Mangrove Ecosystem Services

This study surveyed from interviewing the head of household totally 225 samples in Leam Pak Bia communities who received ecosystem services from a mangrove area. Mangroves had many ecosystem functions as shown in Table 2. The first priority was the regulation functions that were the natural role to act as a regulator in the environment. For example, mangroves could purify air pollution, as a wastewater filter from the community, and as a protective wall of wind or storm. Moreover, mangrove forest could be carbon storage in the tree and nutrient storage in the mud land. The second was the information functions; the communities received this aesthetic scene from recreation, bird watching, or ecotourism in a mangrove area. For the third function; habitat functions; communities were aware of the importance of the benthos habitat and nursery source for fish. This was due to the benthos harvest and the fishery that was the main occupation of the communities. The last was the production functions that were the direct utilization of forest resources in the mangrove area such as timber or non-timber products, the communities did not aware of the importance in these functions because mangrove area was a reserved forest area, and the communities recognized the value that they would receive ecosystem services from other ecosystem functions.

Table 2. Important rank of Laem Phak Bia mangrove ecosystem functions and ecosystem services

Ecosystem Services	Moo 1	Moo 2	Moo 3	Moo 4	Total Average
Regulation functions					
Purify air pollution	4.28	4.09	3.66	4.03	4.04
Wastewater treatment	4.06	4.14	3.82	4.06	4.02
Coastal protection	4.43	4.17	3.74	4.00	4.11
Nutrient Storage	4.26	4.20	3.74	4.06	4.08
Carbon Storage	4.43	4.17	3.68	3.99	4.09
Production functions					
Timber and fuel wood collection	2.54	3.00	2.92	3.18	2.89
Plant and herb collection	2.31	3.29	3.12	3.24	2.92
Habitat functions					
Nursery and food habitat of aquatic animal	4.46	4.09	3.66	4.00	4.08
Benthos habitat	4.64	3.71	3.58	3.65	3.96
Genetic diversity and biodiversity source	4.38	3.94	3.58	3.71	3.93
Information functions					
Natural education	4.63	3.91	3.64	3.96	4.09
Recreation	4.54	3.97	3.70	3.85	4.06
Bird watching	4.44	3.89	3.72	3.87	4.02
Nation heritage	4.81	4.11	3.72	4.19	4.27

Note. 1) score 5.00 – 4.21 is the most importance.

2) score 4.20 – 3.41 is the more importance.

3) score 3.40 – 2.61 is the moderate importance.

4) score 2.60 – 1.81 is the less importance.

5) score 1.80 – 1.00 is the least importance.

3.2 Economic Value of Laem Pak Bia Mangrove Ecosystem Services

In this study, PEV exercise was performed in December 2013 by interviewing totally 225 samples. All villagers consumed rice as a valuable product for surviving. The average of consumed rice in the household was equal to 23.50 kilograms per month in the average price of 25 Baht per kilograms, and; therefore, the average rice value in this community was equal to 7,050 Baht per years. In the PEV exercise, the villagers were a concern on ecosystem functions and ecosystem services from the interviewer briefing. Ecosystem functions and ecosystem services were included of regulation functions, production functions, habitat functions, and information functions. Next step, the villagers, ranked the ecosystem functions and the ecosystem services from the importance that effected to himself and the communities and compared with the valuable thing, rice value, and ranked with them all. Then, all of ecosystem functions and ecosystem services and rice value were rated by villagers. This important value score would be different for each of the villagers.

Table 3 showed the average value of Laem Phak Bia mangrove ecosystem functions and ecosystem services. In overall of Laem Phak Bia communities, they focused on the occupations and the housing that shown from the important rank of ecosystem services. There was more important and the enormous value in the coastal erosion control from wind in 11.05 percentages and the benthos habitat for harvesting in 10.43 percentages. Moreover, they gave a high value in mangrove forest for the next generation in 10.88 percentages. On the other hand, the communities gave less important on production functions that shown the low percentage of timber and fuel wood, and plant and herb collection in 3.41 and 4.52 percentages, respectively.

As shown in Table 3, the PEV calculations indicated Laem Phak Bia mangrove economic value in each ecosystem functions and ecosystem services. The total value was about 10 thousand Baht per household per year or equal 100 million Baht per year and this value was based on rice value. Total economic value of mangrove could distribute value in term of ecosystem functions including;

Regulation functions; the villagers in Laem Phak Bia communities were strongly aware of the regulation functions of mangrove forest in term of coastal protection from wind and sea wave that was the first priority; the value of this function was 11,490.39 Baht per household per year. Other regulation functions were wastewater treatment, nutrient storage, purify air pollution, and carbon storage in the average value of 8,819.89, 7,510.48,

6,093.90, and 5,773.76 Baht per household per year, respectively.

Table 3. Economic value of Laem Phak Bia mangrove ecosystem functions and ecosystem services

Ecosystem services	Average Value (Baht/Household/Year)					Total Value (Million Baht/Year)	Total in %
	Moo 1	Moo 2	Moo 3	Moo 4	Total Average		
Regulation functions							
Purify air pollution	7,745.39	8,686.79	12,868.47	5,049.63	6,093.90	5.90	5.86
Wastewater treatment	13,259.30	12,568.43	4,373.93	5,459.02	8,819.89	8.54	8.48
Coastal protection	13,883.44	19,183.47	6,576.92	8,609.75	11,490.39	11.12	11.05
Nutrient storage	8,662.12	11,640.00	5,202.18	5,862.88	7,510.48	7.27	7.22
Carbon storage	5,297.03	14,130.54	3,346.49	3,761.99	5,773.76	5.59	5.55
Total Regulation functions	48,847.28	66,209.23	32,367.99	28,743.27	39,688.42	38.42	38.16
Production functions							
Timber and fuel wood collection	3,543.63	7,315.40	1,472.07	3,129.16	3,544.74	3.43	3.41
Plant and herb collection	6,323.05	8,204.94	9,676.67	3,466.64	4,704.62	4.55	4.52
Total Production functions	9,866.68	15,520.34	11,148.74	6,595.80	8,249.36	7.99	7.93
Habitat functions							
Nursery and food habitat of aquatic animal	9,397.02	11,449.71	13,869.81	6,170.23	7,294.07	7.06	7.01
Benthos habitat	16,550.31	10,599.41	7,228.11	7,587.11	10,844.13	10.50	10.43
Genetic diversity and biodiversity source	8,116.65	12,444.79	4,974.40	6,288.76	7,539.21	7.30	7.25
Total Habitat functions	34,063.98	34,493.91	26,072.32	20,046.10	25,677.41	24.86	24.69
Information functions							
Natural education	9,686.65	17,822.82	5,163.71	6,033.60	8,843.15	8.56	8.50
Recreation	3,926.62	14,181.77	1,640.05	2,267.73	4,512.39	4.37	4.34
Bird watching	5,373.92	15,014.74	3,196.04	3,138.75	5,714.11	5.53	5.50
Nation heritage	16,104.15	16,206.51	6,567.43	7,227.04	11,317.94	10.96	10.88
Total Information functions	35,091.34	63,225.84	16,567.23	18,667.12	30,387.59	29.42	29.22
Total Economic Value (Baht/Household/Year)	127,869.28	179,449.32	86,156.28	74,052.29	104,002.78		100.00
No. of Household	227	108	422	211	968		
Total Economic Value (Million Baht/Year)	29.03	19.38	36.36	15.63	100.67		

Note. Average current rate on 2013 is 1 US dollar equal 30.73 Baht.

Production functions; this ecosystem function and ecosystem service for Laem Phak Bia communities had less value than the others because the villagers did not access to use timber and non-timber product in the mangrove forest. The value of the timber and fuel wood collection service and the plant and herb collection service of Laem Phak Bia mangrove were equal to 3,544.74 and 4,704.62 Baht per household per year, respectively.

Habitat functions; almost villagers in Laem Phak Bia communities did an offshore fishery as a main occupation, so they would give importance to habitat functions of Laem Phak Bia mangrove forest as a housing and nursery of aquatic animal. The highest value of habitat functions for this community was a benthos habitat on mud land nearly mangrove and communities that valuing 10,844.13 Baht per household per year. Habitat functions as a genetic diversity and biodiversity source and nursery and food habitat of aquatic animal had the value of 7,539.21 and 7,294.07 Baht per household per year, respectively.

Information functions; these ecosystem functions built and maintained a healthy body and mind of the human who had experience in mangroves such as a recreational or learning in natural resource or the value store for future use or next generation. The villagers in Laem Phak Bia communities had high awareness of mangrove ecosystem service for their next generation and then they gave high value for conserving Laem Phak Bia mangrove forest as a nation heritage, which valuing 11,317.94 Baht per household per year. Next, natural education from the mangrove area had the value of 8,843.15 Baht per household per year. The value of

recreation and bird watching in Laem Phak Bia mangrove were less than precious services, which were 4,512.39 and 5,714.11 Baht per household per year, respectively, because almost villagers did not yield from tourism in this mangrove area.

From the total economic value of Laem Phak Bia mangrove ecosystem services, when compared with the other study in Thailand, the total economic value was equal 100 million Baht per year in an area approximately 237.44 hectares or about 424 thousand Baht per hectares per year. This value was higher than the mangrove value assessed from market value approach because the value from PEV study was a sum of ecosystem services; however, and market value approach could only estimate the value of timber, non-timber, fishery, or recreation. Moreover, the value from the study using CVM or willingness to pay for assessing use and non-use value evaluated from the hypothetical market was less than the value from PEV study because the value of the CVM study depended on emotion and scenario whereas the PEV value depended on the proxy goods like the rice value. Finally, the PEV exercise could assess the total economic value in all dimensions of mangrove ecosystem functions and services.

4. Conclusion

It could be clearly seen that Laem Phak Bia mangrove forest had the enormous economic value in all direct use value, indirect use value and non-use value. From the PEV exercise, the results showed that a total economic value of Laem Phak Bia mangrove forest was about 100 million Baht per year. Total economic value of mangrove could be distributed in term of ecosystem functions including regulation functions, production functions, habitat functions, and information functions that were 38, 8, 25, and 29 million Baht per year, respectively. This value reflected the importance of mangrove ecosystem services for the community in term of natural protection from wind or coastal protection as a regulation function from mangrove ecosystem service. These mangroves were the housing and the nursery of benthos and aquatic animals as a habitat function. In summary, these mangroves had the high value in the indirect value. Moreover, this mangrove area also had high non-use value as a national heritage for the next generation. On the other hand, the direct use value from Laem Phak Bia mangrove forest was very low because the communities did not utilize neither timber nor non-timber product, then did not try to use as a recreation or learning, and almost villagers gave an importance for the existing mangrove forest for the future.

From the result of this study, the communities were aware of the importance of the mangrove ecosystem functions especially regulation functions, habitat functions, and information functions that are the key to the success of the operations of government agencies trying to reduce deforestation, to raise awareness and to promote reforestation of mangroves in Thailand. The policy makers should use PEV valuation method to assess the value of ecosystem functions and ecosystem services that are different in each region or locality, in order to determine the appropriate policy or guideline on the utilization and the preservation of natural resource.

Acknowledgments

This research was funded by the King's Royally Initiated Laem Phak Bia Environmental Research and Development Project, The Chaipattana Foundation.

References

- Chunkao, K., & Nimpee, C. (2012). *The survey research of the rise of the coastal area nearly the King's Royally Initiated Laem Phak Bia Environmental Research and Development Project.* (in Thai). The King's Royally Initiated Laem Phak Bia Environmental Research and Development Project Annual Report, The Chaipattana Foundation., Thailand.
- Costanza, R., & Folke, C. (1997). *Valuing Ecosystem Services with Efficiency, Fairness and Sustainability as Goals.* Washington D.C.: Island Press.
- Daily, G. C. (1997). *Valuing and Safeguarding Earth's Life Support Systems.* Washington D.C.: Island Press.
- Daily, G. C., Soderqvist, T., Aniyar, S., Arrow, K., Dasgupta, P., Ehrlich, P. R., ... Walker, B. (2000). The Value of Nature and the Nature of Value. *Science*, 289, 395–396. <http://dx.doi.org/10.1126/science.289.5478.395>
- De Groot, R. S. (1992). *Functions of Nature: Evaluation of Nature in Environmental Planning, Management and Decision Making.* Wolters-Noordhoff, Groningen.
- De Groot, R. S. (1994). Environmental functions and the economic value of natural ecosystems. *Investing in Natural Capital: The Ecological Economics Approach to Sustainability.* Washington D.C.: Island Press, 151–168.
- De Groot, R. S., Wilson, M. A., & Boumans, R. M. J. (2002). A Typology for the Classification, Description and

- Valuation of Ecosystem Functions, Goods and Services. *Ecological Economic*, 41, 393–408. [http://dx.doi.org/10.1016/S0921-8009\(02\)00089-7](http://dx.doi.org/10.1016/S0921-8009(02)00089-7)
- Department of Provincial Administration. (2013). *Population and number of household in Laem Phak Bia sub-district, Ban Laem district, Phetchaburi province*. Department of Provincial Administration, Thailand.
- Edward, B. B., Strand, I., & Sathirathai, S. (2002). Do Open Access Conditions Affect the Valuation of an Externality? Estimating the Welfare Effects of Mangrove-Fishery Linkages. *Environmental and Resource Economics*, 21(4), 343-367. <http://dx.doi.org/10.1023/A:1015129502284>
- Emerton, L., & Mogaka, H. (1996). Participatory Environmental Valuation of Forest Resources in the Aberdares, Kenya. *PLA Notes*. London, 6–10.
- Islam, M. S., & Ikejima, K. (2010). Gear type, species composition and economic value of fisheries in the mangroves of Pak Phanang, Thailand. *Wetlands Ecology and Management*, 18(1), 27-36. <http://dx.doi.org/10.1007/s11273-009-9145-7>
- IUCN. (2005). *Balancing the Returns to Catchment Management: The Economic Value of Conserving Natural Forests in Sekong, Lao PDR*. A Report to the World Conservation Union, Centre for Social and Economic Research on the Global Environment.
- IUCN. (2006). *Valuation, Rehabilitation and Conservation of Mangroves in Tsunami Affected Areas of Hambantota, Sri Lanka*. A Report to the World Conservation Union, Centre for Social and Economic Research on the Global Environment.
- IUCN. (2007). *Environmental and Socio Economic Value of Mangroves in Tsunami Affected Areas: Rapid Mangrove Valuation Study, Panama Village in South Eastern Coast of Sri Lanka*. A Report to the World Conservation Union, Centre for Social and Economic Research on the Global Environment.
- Janekarnkij, P. (2010). *Assessing the Value of Krabi River Estuary Ramsar Site Conservation and Development*. ARE Working Paper No. 2553/4 (December 2010).
- Jitthaisong, O., Dhanmanonda, P., Chunkao, K., & Teejuntuk, S. (2012). Water quality from Mangrove forest: The King's royally initiated Laem Phak Bia Environmental Research and Development Project: Phetchaburi Province, Thailand. *Modern Applied Science*, 6(8), 1-8. <http://dx.doi.org/10.5539/mas.v6n8p1>
- Kuchelmeister, G. (2003). *Participatory Economic Evaluation – Experience in Forest Valuation with Villagers in Vietnam*. Paper Present for the Frontiers 2 Conference: European Applications in Ecological Economics 12-15 February 2003, Tenerife, Canary Islands, Spain.
- Pearce, D. W., Moran, D., & Fripp, E. (1992). *The Economic Value of Biological and Cultural Diversity*. A Report to the World Conservation Union, Centre for Social and Economic Research on the Global Environment.
- Quoc, T. V., Kuenzer, C., Quang, M. V., Moder, F., & Oppelt, N. (2012). Review of valuation methods for mangrove ecosystem services. *Ecological Indicators*, 23, 431–446. <http://dx.doi.org/10.1016/j.ecolind.2012.04.022>
- Rosales, R., Kallesoe, M., Gerrard, P., Muangchanh, P., Phomtavong, S., & Khamsoiphou, S. (2003). *The Economic Returns from Conserving Natural Forests in Sekong, Lao PDR*. IUCN / The World Conservation Union Asia Regional Environmental Economics Programme and WWF Lao Country Office, Vientiane.
- Sathirathai, S. (1998). *Economic Valuation of Mangroves and the Roles of Local Communities in the Conservation of Natural Resources: Case Study of Surat Thani, South of Thailand*. Economy and Environment Program for Southeast Asia. (EEPSEA), Research Report.
- Sathirathai, S., & Edward, B. B. (2001). Valuing Mangrove Conservation in Southern Thailand. *Contemporary Economic Policy*, 19(2), 109-122. <http://dx.doi.org/10.1111/j.1465-7287.2001.tb00054.x>
- Seenprachawong, U. (2002). *An Economic Valuation of Coastal Ecosystems in Phang Nga Bay, Thailand*. EEPSEA Research Report Series No. 2002-RR5. Singapore: IDRC.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

Reliability Worth Assessment in Composite Power System Planning and Operation

Ali Khraiwish Dalabeeh¹ & Eng.Hani Hasan Saleh Al-Hajbi²

¹ Department of Electrical Engineering, Faculty of Engineering Technology, Al-Balqa' Applied University, Jordan

² National Electric Power Company, Amman, Jordan

Correspondence: Ali Khraiwish Dalabeeh, Department of Electrical Engineering, Faculty of Engineering Technology, Al-Balqa' Applied University, P.O.BOX 15008, Amman 11134, Jordan. E-mail: alidalabeeh@yahoo.com

Received: May 20, 2014

Accepted: June 3, 2014

Online Published: August 6, 2014

doi:10.5539/mas.v8n5p45

URL: <http://dx.doi.org/10.5539/mas.v8n5p45>

Abstract

This study presents a comprehensive procedure for evaluating the reliability indices incorporating reduction in interruption cost, and to relate reliability investments with customer's benefits. An analytical algorithm is used to obtain the final optimal operational solution by determining unit commitment for each generating unit to reach the maximum profit (minimizing cost) subjected to operational constraints. The paper describes a method based on the complementing short-term planning with long-term planning take into account the cost benefit approach. The proposed method has been demonstrated on the RBTS to show the feasibility and the economic effect of the proposed method, and was applied to the Jordanian Electric Power System (JEPS) to show its application capability.

Keywords: reliability indices, power system planning, power system operation, interruption cost

1. Introduction

The primary function of a modern power system is to supply its customers with electrical energy as economically as possible with an acceptable degree of reliability. Load curtailment takes place in the system as a result of the inadequacy of the generation unit's capacity to meet the demand or the deficiencies in transmission lines. This is a result of poor planning and failure to take the appropriate steps to use available resources to meet deal with the increase in future loads economically with acceptable reliability and quality. Evaluation of restructured power systems provide both economic and reliability information for generation companies and transmission companies to enhance generation capacity and to expand transmission lines, so a decision should be made to choose among the various transmission system expansion or the new generating units addition alternatives the most reliable and beneficial one not only from a power utility' standpoint but also from customers' viewpoint. There is a need to know customer and user outage costs and to be included and considered in system planning and operational decisions. Power suppliers aimed to enhance system reliability within the available amount of investment while the customer needs fewer service interruptions, and less destruction caused by these interruptions. The main objective of power system restructuring and deregulation is to introduce competition in the power industry and to allow customers to select their suppliers based on price and reliability (Wang & Billinton, 2003). In recent years and at the same time as the restructuring in the power industry, various studies have been carried out into the Transmission Expansion Planning (TEP) in electricity markets and its adjustment to market conditions (Shayeghi, et al., 2008; Buygi, et al., 2003; Wu, et al., 2006). In Sirjani, et al., (2008) and Fu, et al. (2008), there are some criteria and models for creating competition in the electricity market. In (De la Torre, et al., 2008), by considering that users' benefit increase is one of the TEP goals a model is presented that can be utilized by all participants. Also, in Shrestha and Fonseka (2004), optimization the users' benefit and investment cost is evaluated as a single-objective model. In Wangdee (2005), Eliassi, et al., (2009), a probabilistic criterion, known as Expected Customer Interruption Cost (ECOST) due to transmission constraint is presented to evaluate the value of reliability. In (Bresesti, et al, 2009), a method is presented for evaluating the users' benefit and reliability in TEP, which is based on Monte Carlo simulation. TEP should serve its users, so the benefit of both participants in the market and investment cost are considered as economic criteria for the

electricity market, and the congestion cost of the network as a factor for encouraging market competition, also a probabilistic criterion was considered to evaluate the value of reliability (Mohammad & Mojtaba, 2011). TEP model need to include the most relevant short-term transmission related details of electricity markets into the network expansion planning (Fonseka & Shrestha, 2009), Markets in different deregulated power systems are operated by different organizations. These markets operate on either day-ahead or hour-ahead market policy. The most common form of market at present is Power Pool. In a pool system both generating utilities and customers bid for selling and buying electrical power. A generating utility would be out of the competitive market if its price is too high and on the other hand a customer would have no power if its offer were too low. Thus the pool fixes a single price for every hour which is determined by basic supply demand relationship of economics (Shahidepour, et al., 2002; Weron, 2006). The price of electricity is the most important signal to all market participants and the most basic pricing concept is market-clearing price. The intersection of demand and the stepped bid curve determines the market clearing price. The market operators are responsible to maintain a balance between the supply and demand of power. All successful bidders will get paid the market clearing price irrespective of their bidding prices. The fear of getting out of business encourages the supplier to bid the most competitive prices in order to compete for dispatch in the wholesale marketplace (Shahidepour, et al., 2002; Weron, 2006). In a deregulated power system generating utilities and customers may sign contracts for selling and buying of power in addition to the existence of a power pool.

2. Proposed Method

The proposed method is addressed in this section and an illustrative RBTS is employed to clarify the issue. For the sake of effectiveness and simplicity the method is based on the two state models of the generating units and transmission lines, a Markov state of the studied system is defined where every component is in a given operating state. The up state represents the unit available with full capacity; the down state represents the unit out of service. All the possible states of the studied system make up the state space. A generating and transmission failure state array as a sampling tool is constructed using a Generic Algorithm (GA) (Samaan, 2004). Demand is modelled as annualized or annual load. The amount of curtailed load, if needed, is determined for the whole system at each load bus for each sampled state by using a load flow program.) As the sampling process stops, the annualized (or annual) indices are calculated for the whole system and at each load bus. The probability of any two state generation unit to be down is equal to its forced outage rate and the probability failure of a two state transmission line (PTi) is calculated using its failure rate (λ_i) and repair rate (μ_i) as follows (Samaan, 2004).

$$PT_i = \frac{\lambda_i}{(\lambda_i + \mu_i)} \quad (1)$$

The total number of states (N_{states}) for all possible combinations of generating units and transmission lines installed is:

$$N_{states} = 2^{ng+nt} \quad (2)$$

Where (ng) is the total number of generation units and (nt) is the total number of transmission lines in the system. Each possible combination is considered as a raw which represents a system state. A search for failure system state is done and save such states in a state array. Each raw consists of a binary number genes. Each number represents the component's state in the state sample (contingency). The first (ng) in the raw represent generation units while the remaining (nt) represent transmission lines. If any component takes a zero value this means that the component is in the down state while a one value means the component is in the up state. To illustrate the raw construction, consider the small RBTS test system shown in Figure 2. It consists of 2 generator (PV) buses, 4 load (PQ) buses, 9 transmission lines and 11 generating units.

Consider the state that all system components are up, the raw (system state) representing this state is:

$$[1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1]$$

Where the first four elements represent the four generation units (40MW, 40MW, 20MW and 10MW, respectively) at busbar number one, the second seven elements represent the seven generation units (40MW, 20MW, 20MW, 20MW, 20MW, 5MW and 5MW respectively) at busbar number two, the remaining nine elements represent the nine transmission lines in the system. The state probability calculated for each system state, the system states with an associated state probability greater than a threshold value is considered while the remaining system states removed from the evaluation process to reduce the processing time with a negligible error. Each considered system state is evaluated through an evaluation function. The evaluation function returns zero if it is a success state and a state probability if it is a failure state. The failure states that include outages up to a given order only considered also to reduce the processing time. For the failure considered states, the

evaluation function calls a linear programming optimization load flow model that to determine the amount of load curtailment for the load buses in the system. If no load curtailment is needed, then the failure considered state is a success state, otherwise it represents a failure state. The annualized (or annual) reliability indices for each load bus and for the whole system is calculated [Appendix A & B]. Then the annual expected interruption costs are evaluated. These outage costs can then be used as input data to worth assessments of system planning and operational decisions. To enhance the system reliability by adding new transmission lines or new generating units various alternatives should be considered. To select the beneficial alternative the annual expected interruption costs (EIC) re-evaluated and the one which has the much lower EIC has to be chosen. The price of electricity is important signal to all market participants and the most basic pricing concept is market-clearing price which is as states previously is the intersection of demand and the stepped bid curve. The market operators are responsible to maintain a balance between the supply and demand of power. All successful bidders will get paid the market clearing price irrespective of their bidding prices. In the environment of the power market to maintain a balance between the supply and demand of power based on the cost benefit approach, an optimisation models should be established to find the optimal units operations. The expected duration, within a certain period of time over which the load demand exceeds the total generated power, is calculated based on the units that are cleared in the market (the units cleared in market depend on the demand and the reserve) and the expected energy not supplied for every hour also calculated to find the optimal units operations (reach the maximum profit) to supply market demand for each hour. The proposed method illustrated in flow chart as shown in Figure 1.

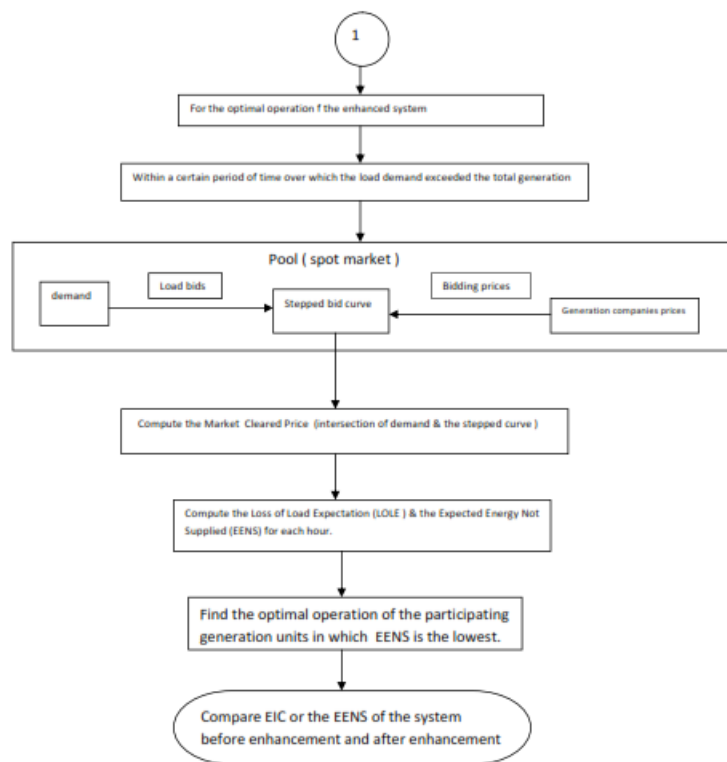


Figure 1. Flow-Chart of the System Short-Term Operation Stage

3. Case Study

3.1 Reliability worth Assessment in Composite System Planning of RBTS

RBTS (Figure 2) is used to evaluate the annual expected interruption costs of three alternatives:

- (A1): the addition of transmission line between Buses 3 and 5 (50Km).
- (A2): the addition of a transmission line between Buses 5 and 6 (50Km).
- (A3): the addition of a generating unit at Bus 6 (20Km).

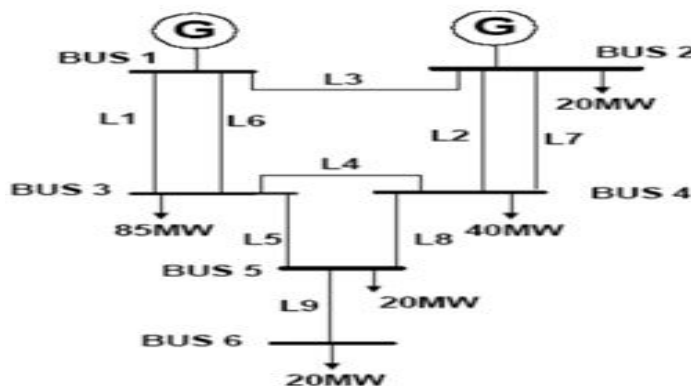


Figure 2. 6-bus RBTS

Table 1, Shows the annual expected energy not supplied (EENS), expected interruption costs (EIC), and interruption energy assessment rate (IEAR) in the RBTS at different Load Levels (0.0% - 10.0%) MW. The annual expected interruption costs (EIC) of the base system and the three alternatives for load increment 0%, 2%, 4%, 6%, 8%, and 10% are shown in Table 2. It can be seen that alternative 1 (A1) has basically the same EIC indices as the base case. This means that the addition of the line between Buses 3 and 5 does not improve the reliability of the system. A2 is the addition of the shorter line between Buses 5 and 6 which leads to much lower EIC than A1. This indicates that the different line addition location have completely different impacts on composite system reliability. It is interesting to note that A2 have higher EIC values than A3 for load increases from to 0% to 10%. A 20MW generating unit requires much higher investment cost than one 50Km line.

Table 1. Annual EENS, EIC, and IEAR in the RBTS at different load levels

Load increment %	Peak Load MW	EENS MWh/yr	EIC (k\$/yr)	IEAR(\$/kWh)
0.0	185	147.84	629.32	4.21
2.0	188.7	166.48	703.34	4.22
4.0	192.4	180.28	764.26	4.24
6.0	196.1	193.00	823.67	4.27
8.0	199.8	212.34	907.12	4.27
10.0	203.5	237.40	1022.57	4.31

Table 2. EIC Indices of the base system and Three alternatives for RBTS(k\$/yr)

Load Increment%	Base System	Alternatives		
		A1	A2	A3
0.0	629.32	622.73	88.41	28.12
2.0	703.34	697.75	110.18	31.89
4.0	764.26	754.38	168.93	39.12
6.0	823.67	816.91	221.47	55.31
8.0	907.12	901.96	297.46	74.57
10.0	1022.57	1014.21	384.74	101.44

Table 3. The annual investment for all alternatives in RBTS

No.	Alternatives description	Annual investment
A1	Add line (50 Km) @ Bus 3-5	Rejected
A2	Add line (50 Km) @ Bus 5-6	281.11 k\$/yr
A3	Add generator unit (20 MW) @ Bus 6	1803.35 k\$/yr

Further analysis can be conducted to evaluate the total cost, which is the sum of annual expected interruption cost and annual investment for the alternatives. The economic life of the power system facilities was assumed to be 30 years and the discount rate 10%. The unit capital cost of a 230KV line is 53\$/Km and the capital cost of a 20MW generating unit 17M\$. The annual investment for all alternatives can be shown in Table 3. At the present load level (0.0%), Alternative 2 can reduce the expected interruption cost by 540.91 k\$/yr (629.32-88.41). This reduction is much larger than its annual investment of 281.11 k\$/yr and therefore Alternative 2 is a beneficial option even the present load level. On the other hand, the reduction of the EIC due to Alternative 3 at 10.0% load increment is 921.13 k\$/yr (1022.57-101.44). This reduction is still smaller than the annual investment of A3 (1803.35 k\$/yr). This indicates that the addition of the 20 MW generating unit is not a cost effective option even when the load has 10.0% growth.

3.2 Reliability worth Assessment in Composite System Operation of RBTS

Table 4. Generating Unit data for the RBTS

Economical Loading order	Company	Unit Size	Unit No.	Forced Outage Rate
1	Genco 1	40	7	0.020
2	Genco 1	20	3	0.015
3	Genco 1	20	4	0.015
4	Genco 2	40	10	0.030
5	Genco 2	40	11	0.030
6	Genco 2	20	9	0.025
7	Genco 2	10	8	0.025
8	Genco 1	20	5	0.015
9	Genco 1	20	6	0.015
10	Genco 1	5	1	0.010
11	Genco 1	5	2	0.010

Also RBTS is used to study the reliability indices in a spot market. It is assumed that two generating companies own the generators at bus 1 and bus 2. Genco 1 owns the seven hydro units at bus 2 with a total capacity of 130 MW. The four thermal units with a total capacity of 110 MW at bus 1 are considered as Genco 2. The total installed capacity of the test system is 240 MW. The two generating companies bid into the pool to meet the demand at buses 2,3,4,5, and 6. It is assumed that the generating companies bid according to the economical loading order of the generators. The economical loading order of the generators, their forced outage rates and rating of the generating units are shown in Table 4.

Table 5. Reliability indices for two typical hours in the spot market

Typical Hour	Demand	Demand Reserve	Units committed (Genco 1)	Units committed (Genco 2)	Committed Capacity(MW)	LOLE (h/h)	EENS (MWh/h)
4354	105.13	115.643	7,3,4	10	120.0	0.082	1.94
2622	83.472	91.819	7,3,4	10	120.0	0.052	0.46

Table 6. The effect on reliability indices due to Gencos' choice in participating

Case	Units participating	Demand	Committed Capacity	Committed Units of Genco 1	Committed Units of Genco 2	LOLE (h/h)	EENS (MWh/h)
1	4,10,11,9,8,6,5	105.13	120.0	4	10,11,9	0.0951	1.4952
2	7,3,10,9,8,5,6	105.13	120.0	7,3	19,9	0.0896	1.3608
3	4,10,9,5,6,1,2	105.13	120.0	4,5,6	10,9	0.0999	0.9936

Table 7. The effect on reliability indices due to capacity and participate Gencos

Units participating	Demand	Committed Capacity	Committed units of Genco 1	Committed Units of Genco 2	LOLE (h/h)	EENS (MWh/h)
7,3,4,10,11,9,8,5,6,1,2	105.13	120.0	7,3,4	10	0.082	1.94
7,4,10,11,9,8,5,6	105.13	140.0	7,4	10,11	0.078	0.91

The annual curve of the test system has 8760 points that is dependent on the annual system peak load, weakly peak load in percentage of annual peak, daily peak load in percent of weekly peak, and hourly peak load in percentage of the daily peak load, the annual system peak load is taken as 185 MW. The reliability of the system can be expressed by many reliability indices. Two reliability indices namely loss of load expectation (LOLE) and expected energy not supplied (EENS), which can provide consistent measures of the reliability of the system are determined in this paper. The reliability indices are determined based on the units cleared in the market. The units cleared in the market depend on the demand and the reserve. The operating reserve is taken as ten percent of the demand predicted. Reliability indices are calculated for every hour in the spot market based on the demand for that hour. Reliability indices for the 4354th hour and the 2622nd hour of the year are shown in Table.5. It is assumed that all the generators bid at their marginal costs for these hours. The total demand is different in both the cases but the committed capacity is the same. Reserve margin is 14.1% of the load for the 4354th hour. 43.8% reserve is provided for the 2622nd hour. The reliability indices in the first case are higher compared to those of the latter. Table 6 shows the effect on the reliability indices due to the choice of the generating companies to participate, on a typical hour (4354th hour) of the spot market for the same demand and committed capacity. For the first two cases shown in the table, generators 10 and 9 are committed. In the case generator 4 (of Genco 1) is participating to meet the load as against generator 3 (of Genco 1) in the second case. This difference in choosing generating units may not affect the reliability indices as both these generating units have the same rating, failure and repair rates. Generator 11 (of Genco 2) in the first case has a FOR of 0,03 but generator 7(of Genco 1) in the second case has a FOR of 0.02, although both have the same capacity. Since the less reliable generator is participating in the first case, the LOLE and EENS in this case are much more than the corresponding values in the second case. The change in the reliability indices in the first and the third case is due to the number of generators participating in the market for the same committed capacity and demand. In both cases generators 4, 10, and 9 are participating. In the first case generator 11 of 40 MW is participating whereas in the third case two generators 4 and 6 of capacity 20 MW each are participating to meet the load. For this demand and the choice of units, the participation of the bigger unit is showing better reliability than two smaller units of same capacity. Another example in Table 7 shows the reliability indices for two cases of different choice of Gencos to participate in the market and different capacities.

4. An Application Case of the Proposed Method

4.1 Application of Reliability worth Assessment in JEPS Planning

JEPS (single line diagram is shown in Appendix A.3) is used to evaluate the annual expected interruption costs of four alternatives:

- (A1): the addition of a 132 KV double circuit OHL between Buses 17 and 19 (70 Km).
- (A2): the addition of a 400 KV double circuit OHL between Buses 1 and 4 (360 Km).
- (A3); the addition of a generating unit at Bus 20 (20 MW).
- (A4): the addition of a generating unit at Bus 4 (120 MW).

Table 8. System EENS, EIC, and IEAR in the JEPS at different load levels

Load increment %	Peak Load (MW)	EENS (MWh/yr)	EIC (K\$/yr)	IEAR (\$/kWh)
0.0	2230.00	2219.55	22982.39	10.35
2.0	2274.60	3845.30	41225.69	10.36
4.0	2319.20	7760.53	80572.70	10.38
6.0	2363.80	14900.42	154918.22	10.40
8.0	2408.40	27977.45	292524.49	10.46
10.0	2453.00	48744.72	506006.02	10.38

Table 9. EIC Indices of the Base System and Four Alternatives for JEPS (K\$/yr)

Load Increment %	Base System	Alternatives			
		A1	A2	A3	A4
0.0	22982.39	20859.02	21907.5	16862.99	8329.92
2.0	41225.69	38582.92	39282.92	33804.63	19563.11
4.0	80572.70	77717.7	78117.7	62705.06	41502.38
6.0	154918.22	151813.3	153857.9	118620.4	79556.49
8.0	292524.49	286715.72	289051.5	239444.4	145612
10.0	506006.02	499881.64	502524.7	418796.1	294731.1

Similarly to the RBTS the annual Expected EENS, EIC, and EIAR are calculated and shown in Table 8, Also the annual Expected Interruption Costs (EIC) of the base system and the four alternatives for load increment 0%, 2%, 4%, 6%, 8%, and 10% are shown in Table 9.

Table 10. The unit capital cost and the annual investment for all alternatives in JEPS

No.	Alternative Description	Unit Capital Cost	Annual Investment
A1	Add line 132 KV (70 Km) @ Bus 17-19	135 K\$/Km	1002.40 K\$/yr
A2	Add line 400 KV (360 Km) @ Bus 1-4	400 K\$/Km	Rejected
A3	Add GT gen. (20 MW) @ Bus 20	17 M\$	1803.33 K\$/yr
A4	Add ST gen (120 MW) @ Bus 4	120 M\$	12730.00 K\$/yr

Table 11. Alternative Annual net gain at different load level

Load Increment %	Base System	Alternative Annual net gain (K\$/yr)			
		A1	A2	A3	A4
0.0	22982.39	1120.97	-14203.51	4316.07	1922.47
2.0	41225.69	1640.37	-13335.63	5617.73	8932.58
4.0	80572.70	1852.59	-12823.41	16064.30	26340.31
6.0	154918.22	2102.52	-14218.12	34494.45	62631.73
8.0	292524.49	3613.04	-12998.77	50083.47	132989.21
10.0	506006.02	5121.98	-11797.12	85406.55	198544.89

Further analysis can be conducted to evaluate the total cost, which is the sum of annual expected interruption cost and the annual investment for the alternatives. The unit capital cost and the annual investment for all alternatives can be shown in Table 10. At the present load level (0.0% increment), Alternative 1 can reduce the expected interruption cost by 2123.37 K\$/yr (22982.39-20859.02). This reduction is much larger than its annual investment of 1002.40 K\$/yr and therefore Alternative 1 is a beneficial option even the present load level. Table 11 shows the alternatives annual net gain at different load levels. This indicates that the addition of the 120 MW units is the best option even when the load has 10.0% growth

4.2 Application of Reliability worth Assessment in JEPS Operation

The JEPS is used to evaluate the reliability indices in a hybrid market. It is assumed that the JEPS become restructured and the followed market pricing criteria is bidding into the pool which is managed by an independent system operator (ISO).

In Jordan Samra Electric Power Generation Company (SEPGCO) was established in 2002 as a new Generation company in order to subject it under privatization process. In 2007, 60% of Central Electricity Generation Company (CEGCO) was privatized. The first Independent Power Producer (IPP) was established in 2007 as Amman East Station (AES). A second IPP power station was established by the end of 2010.

Table 12. The profit of each main generating unit in CEGCO

	HTPS99 G1	HTPS264 G2	REHAB200 G3	REHAB50 G4	ATPS650 G5
Profit (\$/MWh)	-09.31	-04.61	33.92	55.54	45.19

Table 13. The optimal output with the maximum profit

Load (MW)	HTPS99 (MW)	HTPS264 (MW)	REHAP200 (MW)	REHAP50 (MW)	ATPS650 (MW)	Profit (k\$/h)
800	0	0	100	50	650	35.5425

Table 14. Reliability indices for the 2204th hour in the hybrid market

Demand	Demand +Reserve	Units committed (CEGCO)	Units committed (SEGCO)	Units committed (AES)	Committed Capacity (MW)	LOLE (h/h)	EENS (MWh/h)
2050	2255	1,2,11-16,19,31-35,37-40	24,25,26	28,29	2240	0.0003	2.53

CEGCO produces electricity (in MWh) from four main power plants in Jordan; ATPS, HTPS, REHAB, and Risha. SEGCO have a 300 MW combined cycle plant (two GT's and one ST) and two GT's with 100 MW for each. While the AES have only two GT's, 120 MW for each. CEGCO has a long term bilateral contract to run Risha generating units as much as possible as the high priority to run because these units consume a national natural gas. The maximum capacity for this power plant reaches 150 MW. CEGCO has another bilateral contract with a customer to supply a demand of 800 MW, the average tariff of the energy sold is 70.54 \$/MWh. This profit of each generating unit is detailed as in Table 12. A MATLAB program file, using Revised Simplex Method, has been used to find the optimal units operations to reach the maximum profit for CEGCO to supply a demand of 800 MW with a profit 35.5425 k\$/h as shown in Table 13. It is noticed that HTPS did not participate to the contract because HTPS generating units can't compete against the other company due to high running cost. CEGCO prefer to run it to reserve generators. Reliability indices for the 2204th hour (the daily peak in the 1st of April) are shown in Table 14. It is assumed that all the generators bid at their marginal costs for this hour.

5. Conclusion

The different line addition location has completely different impacts on composite system reliability. The different choice of the generating companies to participate in the market for the same committed capacity and demand affect the reliability indices. A Matlab program file, using Revised Simplex Method, is used to find the optimal solution, by determining unit commitment for each generating unit, to reach the maximum profit (or minimize cost) subjected to operational constraints. The needs for power system planners to develop operational strategies by complementing short-term planning with long-term planning based on the cost benefit approach.

References

- Brestitusi, P., Calisti, R., Cazzo, M. V., Gatti, A., Provenzano, D., Vaiani, A., & Vailati, R. (2008). The benefits of transmission expansions in the competitive electricity markets. *International Journal of Energy*, 34, 274–280. <http://dx.doi.org/10.1016/j.energy.09.008>
- Buygi, M. O., Shanechi, H. M., Balzer, G., & Shahidehpour, M. (2003). *Transmission Planning Approaches in Restructured Power Systems*. 2003 IEEE Power Tech Conference Proceedings, vol. 2, June, Bologna, Italy. Digital Object Identifier: 10.1109/PTC.2003.1304666
- De la Torre, S., Conejo, A. J., & Contreras, J. (2007). Transmission Expansion Planning in Electricity Markets. *IEEE Transactions on Power Systems*, 23(1), 238-248. <http://dx.doi.org/10.1109/TPWRS.913717>
- Eliassi, M., Seifi, H., & Haghifam, M. R. (2009). Multi-Objective Value-Based Reliability Transmission Planning Using Expected Interruption Cost Due to Transmission Constraint. *Electric Power and Energy Conversion Systems ~EPECS*, 9, 1-8.
- Fonseka, P. A. J., & Shrestha, G. B. (2009). *Economic Transmission Planning Model Suitable for a Regulated Network Service Provider*, 64 / *JOURNAL OF ENERGY ENGINEERING* © ASCE / SEPTEMBER 2009. http://dx.doi.org/10.1061/_ASCE_0733-9402-135:3-64

- Fu, R., Wei, P., Sun, Y., & Tang, G. (2008). *A new congestion monitoring index constrained multistage transmission expansion planning under market environment* (pp. 978-983). 3rd International Conference on Electric Utility Deregulation and Restructuring and Power Technologies ~DRPT, Nanjing. <http://dx.doi.org/10.1109/DRPT.2008.4523548>
- Mohammad, T. A., & Mojtaba, S. (2011). A Multi-Objective Framework for Transmission Expansion Planning in Electricity Market. *International Review of Electrical Engineering (I.R.E.E.)*, 6(4).
- National Electrical Power Company (NEPCO). *Electricity in Jordan, Development of Electricity*. Retrieved from http://www.nepco.com.jo/english_improve.html
- Samaan, N. (2004). *Reliability Assessment of Electric Power Systems using Genetic Algorithms*. Texas A&M University. Retrieved from <http://hdl.handle.net/1969.1/1054>
- Shahidepour, M., & Yamin, H. (2002). *Market Operations in Electric Power System Forecasting, Scheduling, and Risk Management*. New York: John Wiley & Sons. <http://dx.doi.org/10.1002/047122412X.ch1>
- Shayeghi, H., Jalilzadeh, S., Mahdavi, M., & Hadadian, H. (2008). Studying the influence of two effective parameters on network losses in transmission expansion planning using DCGA. *International Journal of Energy Conversion and Management*, 49, 3017–3024. <http://dx.doi.org/10.1016/j.uncommon.06.013>
- Shrestha, G. B., & Fonseka, P. A. J. (2004). Congestion-Driven Transmission Expansion in Competitive Power Markets. *IEEE Transactions on Power Systems*, 19(3), 1658-1665.
- Sirjani, R., Khaki, B., Neghab, A. P., Mehrtash, A., & Parastar, A. (2008). *A new method for increasing competition in transmission expansion planning* (pp.717–721). 2008 IEEE International Conference on Sustainable Energy Technologies ~ICSET 2008~, November, Singapore.
- Wang, P., & Billinton, R. (2003). IEE Proc.-Gener. Transm. Distrib., 150(5). <http://dx.doi.org/10.1049/ip-gtd:20030723>
- Wangdee, W. (2005). *Bulk Electric System Reliability Simulation and Application*. Ph.D. dissertation, Dept. Elec. Eng., University of Saskatchewan, December. Retrieved from <http://library2.usask.ca/theses/available/etd-12152005-133551/unrestricted/WijamangdeePhDthesis.pdf>
- Weron, R. (2006). *Modelling and Forecasting Electricity Loads and Prices: a statistical approach*. John Wiley & Sons Ltd. HD9685.A2W47 2007 333.793213015195—dc22.
- Wu, F. F., Zheng, F. L., & Wen, F. S. (2006). Transmission investment and expansion planning in a restructured electricity market. *International Journal of Energy*, 31, 954-966. <http://dx.doi.org/10.1016/j.energy.2005.03.001>

Appendix A

Reliability Indices

Three different classes of reliability indices can be computed: (a) probability, (b) frequency and (c) duration indices.

a. Probability index:

The state probability SP_j for each chromosome j is calculated as follows:

$$SP_j = \prod_{i=1}^{ng} G_i \cdot \prod_{i=1}^{nt} T_i$$

Where $G_i = 1 - FOR_i$ if its generator's state = 1 (up state) or $G_i = FOR_i$ if its generator's state = 0 (down state), and $T_i = 1 - PT_i$ if its transmission line's state = 1 or $T_i = PT_i$ if its transmission line's state = 0 .

A threshold probability value is set depending on the required accuracy. If the state probability calculated for a state is less than the threshold value this state is ignored.

The Expected Power Not Supplied (EPNS) for the new state is calculated and the result is saved in the state array.

$$EPNS_j = LC_j \cdot SP_j$$

Where LC_j is the amount of load curtailment for the whole system calculated by the optimal power flow.

b. Frequency Indices:

The failure frequency for the j state is calculated using the conditional probability:

$$FS_j = SP_j \cdot \sum_{i=1}^{ng+nt} [(1 - b_i) \cdot \mu_i - b_i \cdot \lambda_i]$$

Where FS_j is the failure frequency for state j , and b_i is the binary value of component's state number (i) representing a generator unit or transmission line.

c. Duration index:

The failure duration for the j state is calculated using the following equation:

$$T_{sj} = \frac{SP_j}{FS_j}$$

Where T_{sj} is the failure duration for state j .

Appendix B

Assessment of Composite System Adequacy Indices

Annualized adequacy indices for the whole system and for each load bus are calculated using the data saved in the state array. These indices are, Loss of Load Probability (LOLP), Loss of Load Expectation (LOLE), Expected Power Not Supplied (EPNS), Expected Energy Not Supplied (EENS), Loss of Load Frequency (LOLF), and Loss of Load Duration (LOLD). These indices are calculated considering only saved failure states and ignoring success ones.

Let the total number of saved failure states to be (nf), Then the adequacy indices for the whole system are calculated as follows:

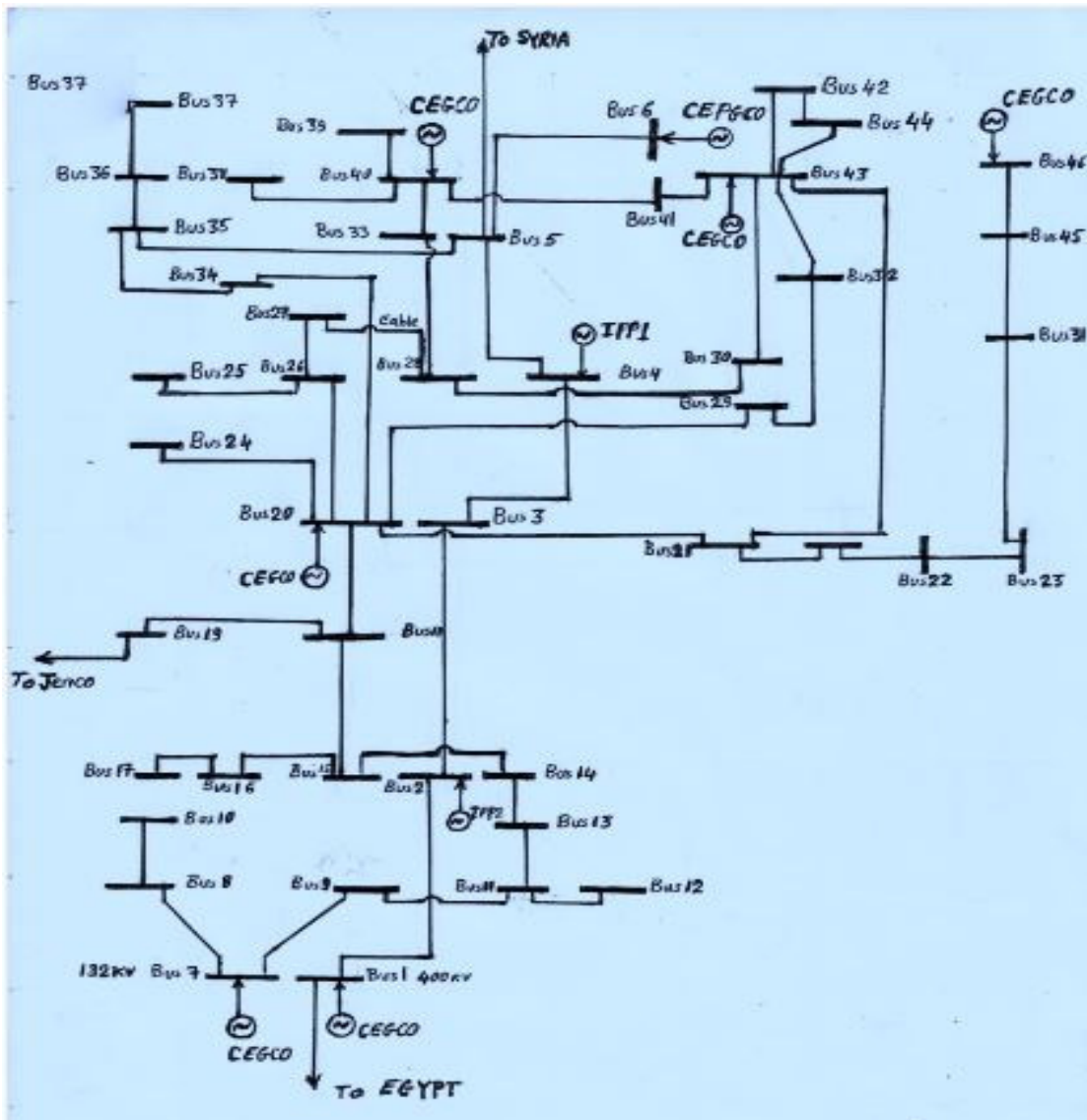
$$LOLP = \sum_{j=1}^{nf} SP_j, LOLF = \sum_{j=1}^{nf} FS_j, EPNS = \sum_{j=1}^{nf} EPNS_j, LOLE = LOLP \cdot 8760, LOLD = \frac{LOLE}{LOLF}, EENS = EPNS \cdot 8760$$

The same set of indices can be calculated for each load bus considering only failure states resulting in load curtailment at this bus and ignoring all other states.

Appendix C

The Jordanian Electrical Power System (JEPS)

The transmission network of the JEPS consists of 46 bus locations connected by 92 lines and transformers, as in the figure below. The transmission lines are at two voltages, 400KV and 132KV. The JEPS has 40 generating units rating from 10MW to 200MW. The basic annual peak load for the test system is 2230MW and the total installed generating capacity is 2525MW.



Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

Interference in Wireless Networks: Causes, Analyses and Practical Mitigating Techniques

A. Hameed¹ & Ali Oudah¹

¹ Faculty of Manufacturing Engineering, Universiti Malaysia Pahang (UMP), 26600 Pekan Pahang, Malaysia

Correspondence: A. Hameed, Faculty of Manufacturing Engineering, Universiti Malaysia Pahang (UMP), 26600 Pekan Pahang, Malaysia. E-mail: multicore.processor@yahoo.com

Received: May 25, 2014 Accepted: June 4, 2014 Online Published: August 6, 2014

doi:10.5539/mas.v8n5p56

URL: <http://dx.doi.org/10.5539/mas.v8n5p56>

Abstract

This paper addresses the key characteristics of interference in wireless networks. Moreover, it considers the impact of interference from Code Division Multiple Access base station to Wideband-Code Division Multiple Access one. A detailed discussion of major interference mitigation techniques in wireless networks, in general, and CDMA and W-CDMA systems, in particular is also imparted on. The techniques and findings in this paper can be easily used in the analyses of other wireless technologies.

Keywords: interference, spurious emissions, inter-modulations, mitigation techniques, guard band, isolation

1. Introduction

Coexistence of wireless systems in various frequency bands has become a commonplace (Shamsan et al. 2012). In a multi-operator situation and in an area where several network operators are striving to furnish mobile services, placing a cell site without interference is often a hurdle faced by operators (Roke Manor Research Ltd 2008). It is vital to realize the risk of interference between systems when planning in such an environment (Nguyen & Zaghoul, 2007).

In circumstances where several mobile networks of different air interfaces operate in a specific province in a combination of frequency bands (i.e. 900, 1800, 1900 and 2100 MHz), the probability of developing interference is inevitable (Oudah, 2013). This end-result is blended in territories where there is no site or districting standard and when there are many operators for each of the reserved frequency bands (Li & Tatesh, 2009). In some regions, identical spectrum portions have been licensed to various operators in contiguous geographical domains. Indeed, it is a great obstacle to fight interference when the two systems operate in the same frequency band in neighboring service regions with non-collocated sites. In nations where siting constraints are widespread, particularly in urban and suburban areas, an operator is further restricted as controls are also put on the position of cell sites together with the corresponding antenna masts (Ball et al. 2008).

In occasions where the same frequency band is employed by several operators, interference may still be reduced while respecting siting stipulations, by sharing towers, antennas, etc.; however, additional obstacle is to create a tower that can fit a number of antennas while retaining the overall interference to a lowest (Kazemitabar, 2010). The installation of antennas at distinctive heights (vertical separation) for different frequency bands in disparate clutter types is an additional key consideration that can mitigate interference levels (ITU-R M.2039-2 2010).

The objective of this paper is to describe key features of interference in emerging cellular systems. Furthermore, the interference between Code Division Multiple Access (CDMA) and Wide-CDMA (WCDMA) technologies is analyzed and potential mitigation techniques are also discussed. To serve its purpose, this paper is organized in the following order: section two is an in-depth discussion of major interference causes and measures. Section three embarks on the analysis of interference between two coexisted systems, i.e. CDMA and W-CDMA, where potential interference issues are analyzed and discussed along with possible mitigation techniques. Section four concludes the findings.

2. Interference Related Factors

Separation between networks is interpreted as the attenuation between the transmitter reference point in the interfering unit, namely Base Station (BS) or Mobile Station (MS) and the receiver reference point (MS or BS). As shown in Fig. 1, it is likely that one BS could potentially disrupt another BS. The separation between the

reference points in this case is given below:

$$\begin{aligned} \text{Available Isolation} = & \text{Tx Feeder loss} - \text{Tx Antenna gain} + \text{Pr opagation loss} - \text{Rx Antenna gain} \\ & + \text{Rx Feeder loss} + \text{Extra attenuation provided by special filters} \end{aligned} \quad (1)$$

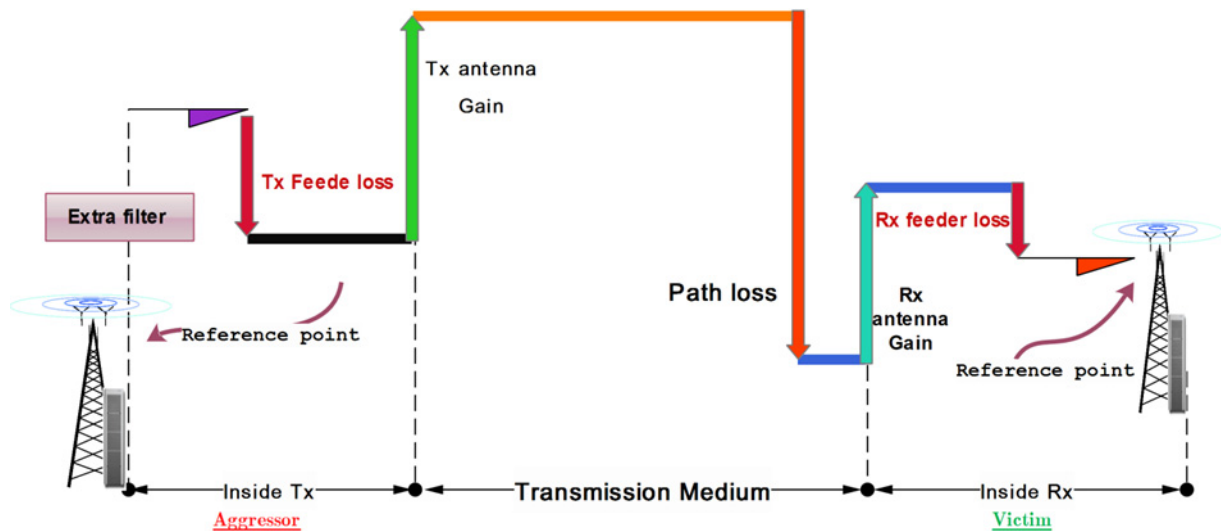


Figure 1. Transmission metrics between two adjacent systems

For gaps more than 10 m, the path loss can be bordered to a free space pathloss model based on the ambience. For short antenna splits, the amount of the separation, i.e., (propagation loss – the combined transmitted and receiver antenna gains) may be approximated by evaluating coupling loss of the typical antennas. For extended separation, particular band pass filters are required.

Employing extra filters in transmit as well as receive paths, one can alleviate or at least lessen the impact of the interference that emerges owing to the coexistence of several wireless networks. The needed filter operation prerequisite can differ hugely relying on the operation frequency, bandwidth, attenuation condition and expense.

Other important elements in interference analysis are the so-called spurious emissions. These mostly give rise to all undesired transmissions from a transmitter. The emissions may incorporate inter-modulation (IM) products stimulated by various frequencies used internally in the Base Station transmitter, the inter-modulation induced by extrinsic carrier frequencies, the harmonics (peaks at double the carrier frequency) and the noise. All of these can result in a co-channel or adjacent channel interference to a neighboring victim receiver.

Interestingly, spurious emissions are sporadic transmissions outside the carrier frequency, but transmitter noise is the weakest level of successive wide-band emission. Transmitter noise cannot be reduced with radio frequency (RF) planning solely, as it is interrelated to the Noise Figure (NF) of the transmit link. This wide-band RF noise is also thought of as sideband noise. The majority of this noise is formed in the exciter section and magnified in following stages and ended at the transmitter output phases. Some transmitters are worse than others in forming this noise.

Furthermore, Inter-modulation (IM) products are established both in transmit and receive links, as two or more frequencies are combined and amplified in non-linear equipment. IM products of order 'n' are the sums and differences in 'n' terms of the primary frequencies. Typically, the greater the order of the IM product, the lesser is its strength. Furthermore, if one of the terms of the product is lower than the rest, it follows that the consequent IM product energy will also decline significantly. The imperfections of the equipment give rise to those unwanted frequencies that will result in a co-channel interference at the victim receiver. Normally, there are two major types of IM products: transmitted and receiver-generated IM(s). On one hand, the transmitter IM products are generated in the BS by mixing of carriers in the same power amplifier, connectors, duplex filter, combiner or the antenna. Transmitter IM can cause co-channel interference, which needs high separation between cells or additional filters to be used to in the intruding cell. Appending filters turns out to be challenging if the intruder cell belongs to a different operator. On the other hand, two or more electromagnetic waves can give rise to inter-modulation in the victim receiver featuring some imperfections. These inter-modulation waves generate

interference if they are in the range of the receiver's bandwidth as they are received and amplified in a way similar to that of the wanted transmission. Accordingly, inter-modulation measures are essential in order to diagnose those surrounding transmitters that may impair the operation of the receiver owing to inter-modulation impact. Receiver (Rx) IM is created in the receiver amplifiers. The primary disparity between Tx IM and Rx IM with respect to co-existence is that the amount of co-channel interference from Tx IM is more conceivable than interference from Rx IM.

Remarkably, the intensity of Rx IM3 and IM5 rises 3 and 5 times as fast (in dB) as the received carriers creating this IM. Given that the powers of Rx IM products step up quickly with the power of the strong received (undesired) carriers, Rx IM is often affiliated with a Rx IM rejection level. Carriers higher than the IM rejection level can lead to serious interference while carriers lower than the level are tolerable or harmless. Both IM3 and IM5 ought to be thought about as receiver IM interference. Usually, the IM behavior of a receiver is reported in terms of inter-modulation rejection ratio and is associated with the 3rd Inter-mod intercept point (IIP3) of the receiver interface. For instance, the Pin IM (intermodulation power in dBm) equivalent to $3 \times P_i - 2 \times \text{IIP3}$ where P_i is the equal power in the two overlapping signals in dBm and IIP3 is the 3rd order Inter-mod intercept point of the receiver in dBm. An additional filter in front of the influenced receiver stage can solve the dilemma.

Normally, receivers are devised to acknowledge particular forms of electromagnetic waves within the bounds of a fixed frequency band. Nevertheless, receivers additionally respond to unwanted signals featuring different modulation and frequency peculiarities. Occasionally, the interfering signals fall into one of the following fundamental types:

- Co-Channel Interference (CCI): This is due to emissions featuring frequencies that present within the bounds of the narrowest pass-band of the receiver.
- Adjacent Channel Interference (ACI): This is owing to undesired signals possessing frequency parts that exist within or nearby the receiver pass-band. Generally, adjacent channel interference is the interference between two wireless networks whose frequency carrier assignments are adjacent or contiguous, particularly in cases where their cell towers are not collocated.
- Out-of-band Interference (OOBI): This occurs when signals having frequency parts, that are outside of the receiver pass-band. If this interference overtakes the noise floor of the victim receiver, it weakens the noise figure of the receiver. In order to combat this type of impairment, either the out-of-band signals must be to be decreased (by inserting filters to the interfering transmitter) or the desired signal level at the victim receiver be improved.

The receiver selectivity serves a significant purpose in attenuating adjacent channel interference. Furthermore, the intermediate frequency (IF) selectivity is the ability of a receiver to filter out any adjacent channels interference. The combination of the transmitter emission mask, the RF and IF selectivities together decide the frequency separation requirements.

Usually, receivers are subject to enormous out-of-band signals that give rise to a spurious response in the receiver. A spurious response can be initiated if the frequency of an intruding signal is such that the signal or one of its harmonics can merge with a local oscillator or one of its harmonics to output an IF in the receiver IF pass-band. The main crucial frequency in this respect is the image frequency of the receiver. In order to hold positive spurious response rejection, the sensitivity threshold of the image frequency should be considered. However, this causes detailed image frequency/spurious response qualities of the receiver if interference issues owing to image frequency are found.

A high interference level inside the RF bandwidth of a receiver can lead to interference even if the emission is outside the pass-band of the IF amplifiers resulting in a decrease of gain for the wanted signal owing to imperfections in the receiver front end. Such end-result gives rise to decreased Signal-to-Noise ratio (S/N) of the receiver, if a specific saturation reference power threshold is exceeded. This occurrence is ordinarily named blocking or desensitization. For this goal, the interference level at the front end of the receiver requires to be calculated over the whole RF bandwidth of the receiver. Namely, the receiver blocking or desensitization level is the highest level of a non-cochannel or adjacent channel interference that may be experienced as a low input signal (about three dB above the sensitivity level) without lowering the receiver operation.

Typically, noise, spurious emissions and transmitter IM products outside the allocated frequency block are inherent in any transmitter. These emissions give rise to noise and can stimulate interference to a neighboring receiver that is set to a weak signal from a transmitter (BS or MS) which is far away. As a rule, closer the interfering channel to the wanted channel, more responsive the receiver is to the interference. The interference is

a combination of the adjacent channel interference due to deficient suppression in the receiver band-pass filter and the co-channel interference from the wide-band noise of the closer interfering transmitter. Fig. 2 depicts some of the elements explained above.

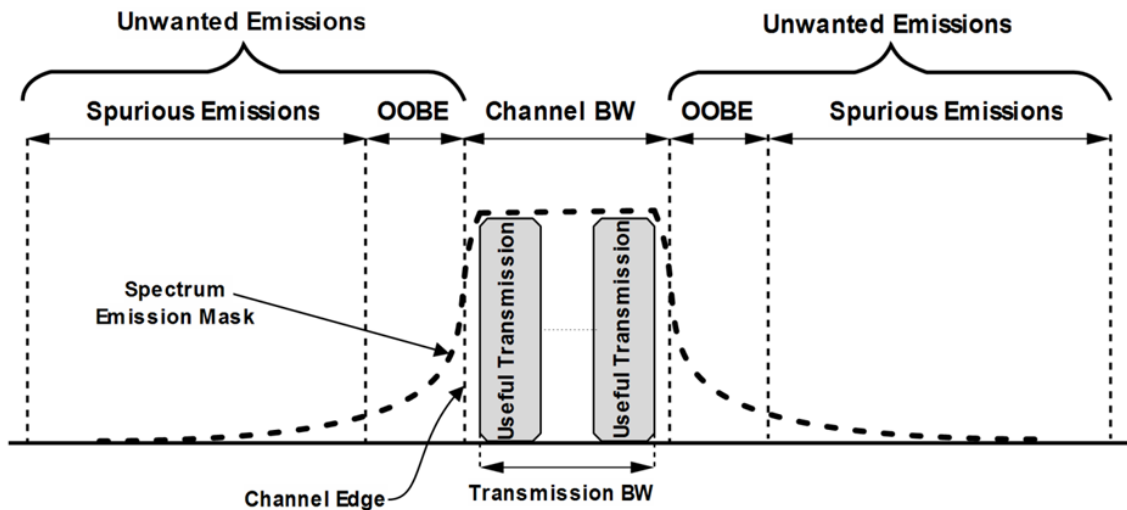


Figure 2. Key unwanted emissions in wireless systems

3. Interference from CDMA to W-CDMA Networks

In several countries in Asia, there has always been an argument as to whether Personal Communications Service (PCS) band (1850-1910/1930- 1990 MHz) allocations can coexist (fully or partially) with 2.1 GHz band allocations, also referred to as the UMTS band (1920-1980/2100-2170 MHz) and if so, what are the acceptable degrees of interference and the needed guard bands between these blocks. There is (60 + 60) MHz of spectrum openness if allocated either for PCS or for UMTS solely. See Fig. 3 for another potential sharing of the spectrum between PCS and 2.1 GHz band allocations (ITU-R M.2116-1 2010).

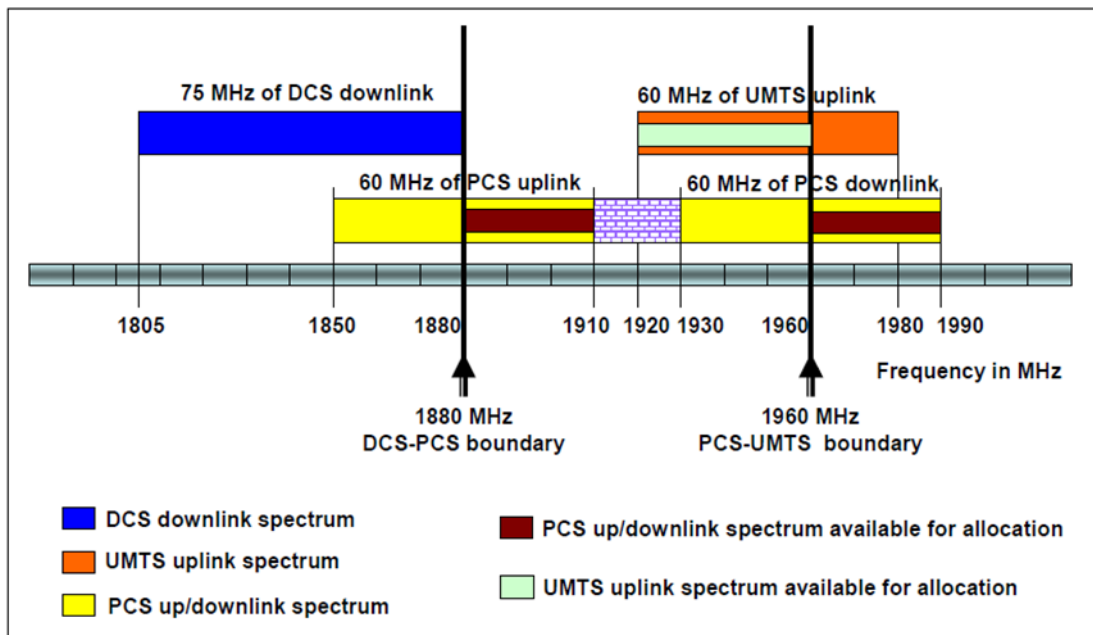


Figure 3. Potential spectrum sharing scenario between PCS and UMTS bands

In this band sharing plan, if the operators with PCS band allocation are the holders, then the matter will be to protect the holders while assigning more spectrums in the 2.1 GHz band. In some other situations, the conflict has been to rule out PCS allocations to protect forthcoming IMT-2000 core band allocations. In Fig. 3, it is depicted that the border between PCS and UMTS band allocations is 1960 MHz. With 5 MHz protection band (from 1955 to 1960 MHz) at the 1960 MHz margin, one gets (30+30) MHz for PCS and (35+35) MHz for UMTS operations. It is viable to move the border to either side of the spectrum (anywhere, from 1920 to 1980 MHz) to earmark various bandwidth sharing between PCS and UMTS operations.

The next section assesses potential coexistence issues between CDMA2000 network in the PCS band and WCDMA network in the 2.1 GHz band by scrutinizing the smallest guard band necessary between the two networks.

4. Formulation, Results and Discussion

In this subsection, the interference from CDMA2000 base station transmitter to the UMTS Node-B receiver is considered. The Thermal Noise Floor at the UMTS Node-B receiver is found as follows:

$$TN = -174 \frac{dBm}{Hz} \times K \times T + 10 \times \text{Log}(3.84) + NF_{Rx} \quad (2)$$

Where TN is the thermal noise of UMTS receiver, K is Boltzman's constant = 1.380×10^{23} T is receiver's temperature and NF_{Rx} is UMTS receiver noise floor. For receiver noise figure of 5 dB, $TN = -103$ dBm/3.84 MHz. As per 3GPP specification (3GPP TS 25.942), a maximum tolerable level of interference (I_{Rx}) from the intruder base station (Node-B) to the victim Node-B receiver is given as -110 dBm/3.84 MHz (3GPP TS 25.942). With TN value of -103 dBm/3.84 MHz for WCDMA Node-B receivers, a spurious signal at the maximum tolerable limit of -110 dBm/3.84 MHz would be 7 dB below the TN value and its presence would cause a receiver sensitivity degradation of 0.8 dB, which is calculated as per the formula shown below (Oudah et al. 2012):

$$S = 10 \times \text{Log} \left(1 + 10^{\frac{I_{Rx} - TN}{10}} \right) \quad (3)$$

Where S is the degradation in receiver sensitivity, I_{Rx} is maximum tolerable level of interference and TN is thermal noise. Similarly, 3GPP specifies a blocking requirement for WCDMA Node-B at a received spurious signal level of -115 dBm/3.84 MHz [10]. The processing gain (PG) of the WCDMA signal is 25 dB (i.e., $10 \times \log(3840000/12200)$) and for a target E_b/I_t of 5 dB, the Node-B receiver reference sensitivity would be of the order of -123 dBm/3.84 MHz (i.e., $TN + E_b/I_t - PG = -103 + 5 - 25 = -123$ dBm/3.84 MHz). In PCS band, the antenna gains are in the order of 15 dBi and the feeder cable losses would be around 3 dB. Hence, the equivalent isotropically radiated power (EIRP) of CDMA BS Tx would be around 55 dBm ($43 + 15 - 3 = 55$ dBm).

Assuming that the interfering CDMA2000 BS in PCS band has a vendor specific spurious emission limit of -75 dBc/30 kHz (20 dB better than the standards specified spurious emission limits of -55 dBc/30 kHz) (3GPP TS 25.942) for a carrier-to-carrier separation greater than 1.98 MHz, the amount of required isolation $I_{isolation}$ (dB) achieved through spatial separation of antennas to mitigate the interference due to the out-of-band spurious emissions can be found as follows:

$I_{isolation}$ (dB) = Tx EIRP (dBm) - Tx out-of-band spurious emission limit - $10 \times \log(30/1250)$ - 3GPP specified WCDMA maximum tolerable level of Interference - Rx feeder cable loss (dB) + Rx antenna gain (dB) - Tx filter attenuation (dB)

$$\begin{aligned} &= 55 \text{ (dBm)} - 75 + 10 \times \text{Log}(1250 / 30) - (-110 \text{ dBm}) - 3 \text{ (dB)} + 15 \text{ (dBi)} - Tx_{filter_attenuation} \\ &= (55 - 75 + 16 + 110 - 3 + 15) - Tx_{filter_attenuation} \text{ (dB)} \\ &= 118 - Tx_{filter_attenuation} \text{ (dB)} \end{aligned} \quad (4)$$

Therefore, with an extra filter of realistic complexity and cost in the CDMA BS Tx path with 60 dB attenuation, the required isolation due to physical separation of antennas " $I_{isolation}$ (dB)" = $118 - 60 = 58$ dB, which can very easily be realized by having about 40 meters of physical separation between CDMA2000 BS Tx and WCDMA Node-B Rx antennas. With some vertical separation of antennas, one can use lesser values for antenna gains and hence the burden on the required isolation further reduces.

As per 3GPP specification (3GPP TS 25.104), there is a 5 dB lower (more stringent) constraint on WCDMA Node-B tolerable blocking level compared to the maximum tolerable level for out-of-band Interference (-115 dBm versus -110 dBm), which translates directly into an extra 5 dB isolation requirement for blocking due to

Intermodulation. Therefore, the isolation requirement with physical separation of antennas to mitigate blocking would be around 65 dB. Hence, blocking would be the main issue for WCDMA Node-B.

It is worth mentioning that there are no clear requirements in band class 6 (IMT-2000) to protect WCDMA Node-B receiver in a co-area operation scenario from the spurious emissions caused by a CDMA2000 BS of another operator. But, most of the vendors would supply CDMA2000 BSs with spurious emissions that are most likely be considerably lower than the standards defined values, because a CDMA2000 BS needs to protect its receiver from its own transmitter.

As with CDMA2000 BS system, one can assume that the blocking performance of WCDMA Node-B receiver would most likely be expressively better than the standards stated value, because WCDMA Node-B system also has to protect its receiver from its own transmitter.

Using specified performance values for BS transmitters and receivers to establish isolation requirements also leads to overly pessimistic results, since they are not describing the real equipment performance. Accordingly, around 65 dB of isolation requirement is sufficient to meet the BS to Node-B spurious emission as well as intermodulation interference conditions, and it is quite possible to achieve such isolation requirement through physical separation of antennas with good antenna installation practices. 65 dB of antenna isolation is achievable with a site-to-site spacing of around 60 meters.

5. Conclusions

In this paper, deployment analysis of wireless systems has been discussed in details. The consequences of spurious emissions between two collocated systems and other related side-effects have also been discussed. It has been shown that analyses which looked into the coexistence of collocated CDMA2000 with WCDMA operating in the same geographic area are established on deterministic computations by supposing the worst-case scenario, i.e., BS /Node-B behavior according to minimum prerequisites stipulated in the related standard and maximum transmit power of BS /Node-B. Nevertheless, it is estimated that the real CDMA2000 BS out-of-band spurious emissions and blocking specifications of UMTS Node-B receivers are considerably better than the smallest requirements because CDMA2000 system needs to protect its receiver from its own transmitter. Consequently, the current equipment performance estimated to be at least 22 dB greater than the minimum performance specifications should be used. As a result, a minimum guard band of 5 MHz is adequate assuming actual filters and separations resulting from practical antenna separation. In reality, the guard band between the UMTS uplink and the PCS downlink is constantly more than 5 MHz as described below.

Acknowledgment

This research work has not been possible without the financial support from Universiti Malaysia Pahang (UMP) Vot. RDU 130387. Therefore, the authors would like to thank UMP for the continuous support.

References

- Ball, C., Hindelang, T., & Kambourov, I. (2008). *Spectral efficiency assessment and radio performance comparison between LTE and WiMAX* (pp. 0–5). Indoor and Mobile Radio.
- ITU-R M.2039-2. (2010). Characteristics of terrestrial IMT-2000 systems for frequency sharing/interference analyses.
- ITU-R M.2116-1. (2010). Characteristics of broadband wireless access systems operating in the land mobile service for use in sharing studies.
- Kazemitabar, S. J. (2010). *Coping with Interference in Wireless Networks* (1st ed.). Springer.
- Li, J., & Tatesh, S. (2009). Coexistence studies for 3GPP LTE with other mobile systems. *IEEE Communications Magazine*, 47(4), 60–65. <http://dx.doi.org/10.1109/MCOM.2009.4907408>
- Nguyen, M. A., & Zaghoul, A. I. (2007). On the Characterization of Cochannel Interference in an Aeronautical Mobile Environment. *IEEE Transactions on Vehicular Technology*, 56(2), 837–848. <http://dx.doi.org/10.1109/TVT.2007.891440>
- Oudah, A. (2013). Resource Element-Level Dimensioning of Long Term Evolution Networks. *Journal of Information and Communication Technology (JICT)*, 12, 189–205.
- Oudah, A., Rahman, T. A., & Seman, N. H. (2012). Coexistence and Sharing Studies of Collocated and Non-Collocated Fourth Generation Networks In the 2.6 GHz Band. *Journal of Theoretical and Applied Information Technology*, 43, 112–118.
- Roke Manor Research Ltd. (2008). Practical Compatibility and Coexistence Measures Analysis, Hampshire, UK.

Shamsan, Z. A., Rahman, T. B. A., & Al-Hetar, A. M. (2012). Point-Point Fixed Wireless and Broadcasting Services Coexistence with IMT-Advanced System. Progress in *Electromagnetics Research*, 122, 537-555. <http://dx.doi.org/10.2528/PIER11090305>

3GPP TS 25.104, Technical Specification. “BTS Radio transmission and Reception (FDD).”

3GPP TS 25.942, Technical Specification. “RF System Scenarios.”

3GPP2, Technical Specification. “Recommended Minimum Performance Standards for cdma2000 Spread Spectrum BSs”, C. S0010-B, V1.0.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

Equality of Google Scholar with Web of Science Citations: Case of Malaysian Engineering Highly Cited Papers

Nader Ale Ebrahim¹, Hadi Salehi², Mohamed Amin Embi³, Mahmoud Danaee⁴, Marjan Mohammadjafari⁵, Azam Zavvari⁶, Masoud Shakiba⁶ & Masoomeh Shahbazi-Moghadam⁷

¹ Research Support Unit, Centre of Research Services, Institute of Research Management and Monitoring (IPPP), University of Malaya, Malaysia

² Faculty of Literature and Humanities, Najafabad Branch, Islamic Azad University, Najafabad, Isfahan, Iran

³ Faculty of Education, Universiti Kebangsaan Malaysia (UKM), Bangi, 43600, Malaysia

⁴ Faculty of Agriculture, Roudehen Branch, Islamic Azad University, Roudehen, Iran

⁵ Department of Industrial Engineering, Faculty of Engineering, Science and Research Branch, Islamic Azad University, Kerman, Iran

⁶ Center for Software Technology and Management, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia

⁷ Perdana School of Science, Technology and Innovation Policy, Universiti Teknologi Malaysia

Correspondence: Hadi Salehi, Faculty of Literature and Humanities, Najafabad Branch, Islamic Azad University, Najafabad, Isfahan, Iran. E-mail: hadisalehi1358@yahoo.com

Received: June 1, 2014 Accepted: June 23, 2014 Online Published: August 6, 2014

doi:10.5539/mas.v8n5p63

URL: <http://dx.doi.org/10.5539/mas.v8n5p63>

Abstract

This study uses citation analysis from two citation tracking databases, Google Scholar (GS) and ISI Web of Science, in order to test the correlation between them and examine the effect of the number of paper versions on citations. The data were retrieved from the Essential Science Indicators and Google Scholar for 101 highly cited papers from Malaysia in the field of engineering. An equation for estimating the citation in ISI based on Google scholar is offered. The results show a significant and positive relationship between both citation in Google Scholar and ISI Web of Science with the number of versions. This relationship is higher between versions and ISI citations ($r = 0.395$, $p < 0.01$) than between versions and Google Scholar citations ($r = 0.315$, $p < 0.01$). Free access to data provided by Google Scholar and the correlation to get ISI citation which is costly, allow more transparency in tenure reviews, funding agency and other science policy, to count citations and analyze scholars' performance more precisely.

Keywords: bibliometrics, citation analysis, evaluations, equivalence, Google Scholar, High cited, ISI Web of Science, research tools, H-index

1. Introduction

Citation index as a type of Bibliometrics method traces the references in a published article. It shows that how many times an article has been cited by other articles (Fooladi et al., 2013). Citations are applied to evaluate the academic performance and the importance of information contained in an article (Zhang, 2009). This feature helps researchers get a preliminary idea of the articles and research that make an impact in a field of interest. The avenues to evaluate citation tracking have greatly increased in the past years (Kear & Colbert-Lewis, 2011). Citation analysis was monopolized for decades by the system developed by Eugene Garfield at the Institute for Scientific Information (ISI) now owned by Thomson Reuter Scientific (Bensman, 2011). ISI Web of Science is a publication and citation database which covers all domains of science and social science for many years (Aghaei Chadegani et al., 2013). In 2004, two competitors emerged, Scopus and Google Scholar (Bakkalbasi, Bauer, Glover, & Wang, 2006). Google Inc. released the beta version of 'Google Scholar' (GS) (<http://scholar.google.com>) in November 2004 (Pauly & Stergiou, 2005). These three tools, ISI from Thomson Reuters, Google Scholar (GS) from Google Inc. and Scopus from Elsevier are used by academics to track their citation rates. Access to ISI Web of Science is subscription-based service while GS provides a free alternative to retrieve the citation counts. Therefore, the researchers need to estimate their citation in ISI by knowing the GS

citation counts. On the other hand, publishing a research paper in a scholarly journal is necessary but not sufficient for receiving citations in the future (Nader Ale Ebrahim, 2013). The paper should be visible to the relevant users and authors in order to get citations. The visibility of the paper is defined by the number of paper versions which are available in the Google Scholar database. The number of citations will be limited to the versions of the published article on the web. The literature has shown increased visibility by making research outputs available through open access repositories, wider access results and higher citation impact (Nader Ale Ebrahim et al., 2014; Amancio, Oliveira Jr, & da Fontoura Costa, 2012; Antelman, 2004; Ertürk & Şengül, 2012; Hardy, Oppenheim, Brody, & Hitchcock, 2005). A paper has greater chance of becoming highly cited whenever it has more visibility (Nader Ale Ebrahim et al., 2013; Egghe, Guns, & Rousseau, 2013).

The objectives of this paper are two-fold. The first objective is to find the correlation between Google Scholar and ISI citation in the highly cited papers. The second objective is to find a relationship between the paper availability and the number of citations.

2. Google Scholar & Web of Science Citations

The citation facility of Google Scholar is a potential new tool for Bibliometrics (Kousha & Thelwall, 2007). Google Scholar, is a free-of-charge by the giant Google search engine, has been suggested as an alternative or complementary resource to the commercial citation databases like Web of Knowledge (ISI/Thomson) or Scopus (Elsevier) (Aguillo, 2011). Google Scholar provides Bibliometrics information on a wide range of scholarly journals, and other published material, such as peer-reviewed papers, theses, books, abstracts and articles, from academic publishers, professional societies, preprint repositories, universities and other scholarly organizations (Orduña-Malea & Delgado López-Cózar, 2014). GS also introduced two new services in recent years: Google Scholar Author Citation Tracker in 2011 and Google Scholar Metrics for Publications in April 2012 (Jasco, 2012). Perhaps some of these documents would not otherwise be indexed by search engines such as Google, so they would be "invisible" to web searchers, and clearly some would be similarly invisible to Web of Science users, since it is dominated by academic journals (Kousha & Thelwall, 2007). On the other hand, the Thomson Reuters/Institute for Scientific Information databases (ISI) or Web of Science database (actually there is ambiguity between different names of former ISI), include three databases: Science Citation Index/Science Citation Index Expanded (SCI/SCIE) (SCIE is the online version of SCI), Social Science Citation Index (SSCI) and Arts and Humanities Citation Index (AHCI) (Larsen & von Ins, 2010). Since 1964 the Science Citation Index (SCI) has been a leading tool in indexing (Garfield, 1972).

Few studies have been done to find a correlation between GS with WoS citations. Cabezas-Clavijo and Delgado-Lopez-Cozar (2013) found that the average h-index values in Google Scholar are almost 30% higher than those obtained in ISI Web of Science, and about 15% higher than those collected by Scopus. GS citation data differed greatly from the findings using citations from the fee-based databases such as ISI Web of Science (Bornmann et al., 2009). Google Scholar overestimates the number of citable articles (in comparison with formal citation services such as Scopus and Thomson Reuters) because of the automated way it collects data, including 'grey' literature such as theses (Hooper, 2012). The first objective of this study is to find the correlation between Google Scholar and ISI citation in the highly cited papers.

3. Visibility and Citation Impact

Nader Ale Ebrahim et al. (2014) based on a case study confirmed that the article visibility will greatly improve the citation impact. The journal visibility has an important influence on the journal citation impact (Yue & Wilson, 2004). Therefore, greater visibility caused higher citation impact (Zheng et al., 2012). In contrast, lack of visibility has condensed a significant citation impact (Rotich & Musakali, 2013). Nader Ale Ebrahim et al. (2013) by reviewing the relevant papers extracts 33 different ways for increasing the citations possibilities. The results show that the article visibility has tended to receive more download and citations. In order to improve the visibility of scholars' works and make them relevant on the academic scene, electronic publishing will be advisable. This provides the potential to readers to search and locate the articles at minimum time within one journal or across multiple journals. This includes publishing articles in journals that are reputable and listed in various databases and peer reviewed (Rotich & Musakali, 2013). Free online availability substantially increases a paper's impact (Lawrence, 2001a). Lawrence (2001a, 2001b) demonstrated a correlation between the likelihood of online availability of the full-text article and the total number of citations. He further showed that the relative citation counts for articles available online are on average 336% higher than those for articles not found online (Craig, Plume, McVeigh, Pringle, & Amin, 2007).

However, there are limited resources to explain the relationship between the paper availability and the number of citations (Craig et al., 2007; Lawrence, 2001b; McCabe & Snyder, 2013; Solomon, Laakso, & Björk, 2013).

None of them discussed about the relationship between the number of versions, and citation. The number of “versions” will be shown in any Google Scholar search result. Figure 1 shows 34 different versions of an article entitled “Virtual Teams: a Literature Review (Nader Ale Ebrahim, Ahmed, & Taha, 2009)” and number of citations. The second objective of this research is to find a relationship between the paper availability and the number of citations.

Virtual Teams: a Literature Review.

[NA Ebrahim, S Ahmed, Z Taha - Australian Journal of Basic ... , 2009 - search.ebscohost.com](#)

Abstract In the competitive market, virtual teams represent a growing response to the need for fast time-to-market, low-cost and rapid solutions to complex organizational problems.

Virtual teams enable organizations to pool the talents and expertise of employees and ...

Cited by 95 Related articles All 34 versions Cite Saved

Figure 1. The number of “versions” in the Google Scholar search result

4. Methodology

Highly cited papers from Malaysia in the field of engineering were retrieved from the Essential Science Indicators (ESI) which is one the Web of Knowledge (WoK) databases. ESI provides access to a comprehensive compilation of scientists’ performance statistics and science trend data derived from WoK Thomson Reuters databases. Total citation counts and cites per paper are indicators of influence and impact of each paper. There is a threshold to select highly cited papers according to the baseline data in ESI. This threshold is different from one discipline to another one. ESI rankings are determined for the most cited authors, institutions, countries, and journals (The Thomson Corporation, 2013). The paper must be published within the last 10-year plus four-month period (January 1, 2003-April 30, 2013) and must be cited above threshold level, in order to be selected. Essential Science Indicators data used in this research have been updated as of July 1, 2013.

Google Scholar which is a free online database was used for deriving the number of citations and versions of the ESI highly cited papers. The data from ESI was collected on 29 July 2013 and Google Scholar data was collected on 31 July 2013. The total numbers of 101 papers were listed in ESI as highly cited papers from Malaysia in the field of engineering. The lists of 101 papers were retrieved from ESI database and then were exported to an Excel sheet. A search engine was developed to get the number of citations and versions from Google Scholar. This gadget assisted the present researchers to collect the data more precisely and faster than searching for the papers one by one. The Statistical Package for the Social Sciences (SPSS) was used for analyzing the data. The results are illustrated in the following section.

5. Results and Discussion

The number of citations which were derived from Web of Knowledge platform hereafter are called ISI citation. To study the relationship among the number of citations in Google scholar and ISI and the number of versions, correlation coefficients were computed.

Table 1 shows descriptive statistics of the variables.

Table 1. Descriptive statistic of variables

	N	Minimum	Maximum	Mean	Std. Deviation
Versions	101	2	28	5.62	3.078
Cited in Google Scholar	101	4	348	80.76	71.718
ISI citation	101	5	189	43.15	36.076

As both numbers of citations in Google scholar and ISI were distributed normally, Pearson correlation coefficient (r) was used and the results showed a very high positive and significant association ($r = 0.932$, $P < 0.01$) between the number of citations in Google scholar and ISI for the articles that were published during 2006 to 2012 from Malaysia in the field of engineering. To study the relationship between both citation and the number of versions,

Spearman Rho was used due to the non-normal distribution of the versions. The results showed a significant and positive relationship between both citations in Google Scholar and ISI with the number of versions. This relationship was higher between versions and ISI citations ($r = 0.395, p < 0.01$) than between versions and Google Scholar citations ($r = 0.315, p < 0.01$). Linear regression was also applied to predict the number of citations in ISI based on Google Scholar citations. The results showed a very high predictability ($R^2 = 0.836$) for the linear model (see Figure 2) which was significant ($F = 511.63, p < 0.01$). Therefore, the final equation for estimating the citation in ISI based on Google Scholar is:

$$\text{ISI Citation} = 5.961 + 0.460 (\text{Google Scholar citation})$$

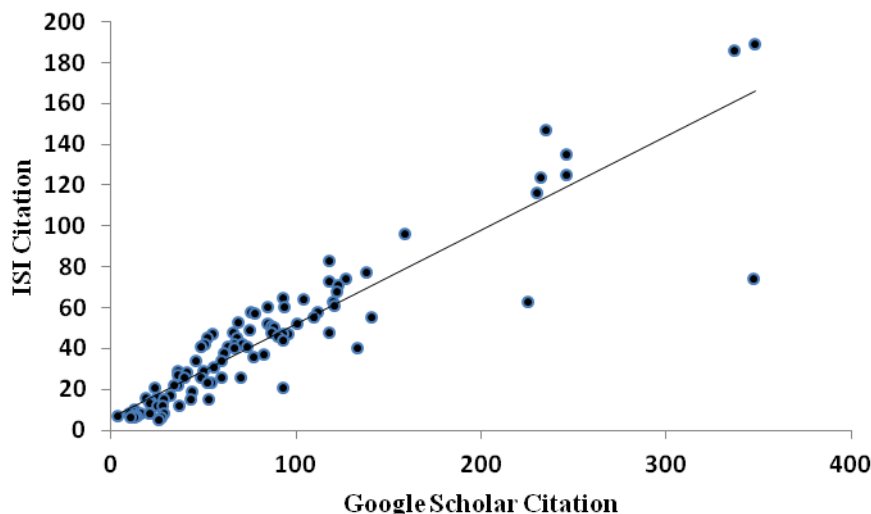


Figure 2. Scatter diagram between ISI citation and Google Scholar citation

To study the effect of the number of versions on both citations in Google Scholar and ISI, simple linear regression was applied. The results indicated that the number of versions had a significant positive effect on citations in both databases (see Table 2 and Table 3).

Table 2. Summary of regression analysis results

	R Square	F	β	t	p
Model^a	0.276	39.12**	0.532	6.255	<0.01
Model^b	0.272	38.316**	0.528	6.19	<0.01

Predictors: Versions

a: Dependent Variable: Cited in Google Scholar, b: Dependent Variable: ISI citation

Table 3. Descriptive statistics of variables - Year

Year	N	Versions		Cited in Google Scholar		ISI citation	
		Mean	SD	Mean	SD	Mean	SD
Before 2009	20	7.75	5.25	152.85	103.741	79.8	46.6
2009	26	6.08	1.695	101.19	38.948	56.96	20.577
2010	18	5.11	2.193	70.17	50.097	41.44	26.86
2011	16	4.31	1.352	49.25	33.66	21.31	12.015
2012	21	4.48	2.089	19.9	9.518	9.24	3.315

A comparison between Google Scholar and ISI citation for highly cited papers from Malaysia in the field of engineering (see Figure 3) shows that the citation counts in Google Scholar are always higher than the number of

citations in ISI.

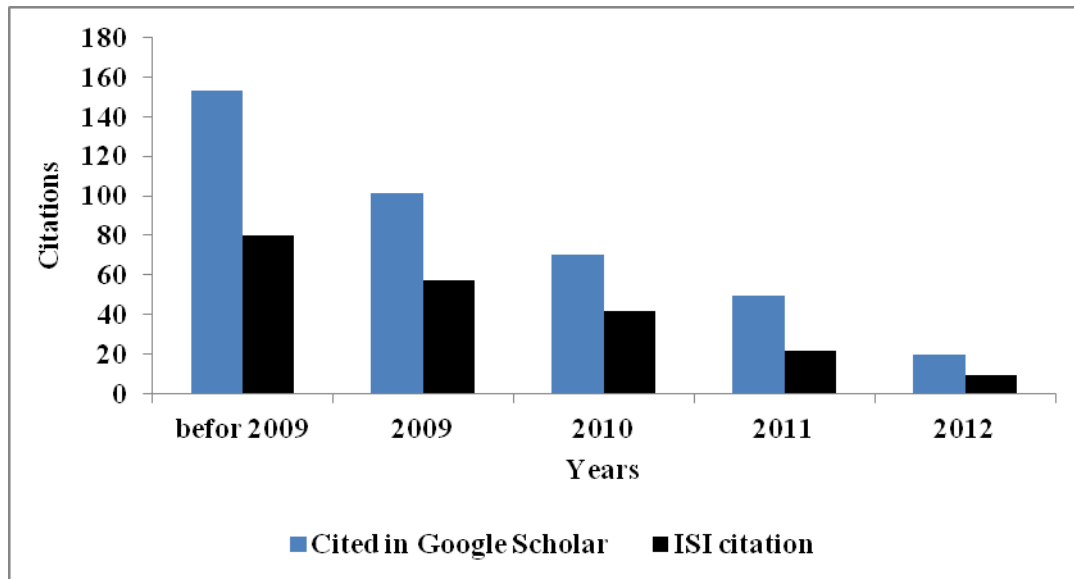


Figure 3. Comparison between Google Scholar and ISI citations

6. Conclusion

The number of publications and the number of citations in ISI Web of Science are used to measure the researchers' scientific performance and their research impact. However, these numbers are not freely available. Therefore, the offered equation can be used as a reference to convert the number of Google Scholar citations to ISI citations. On the other hand, the number of versions of both systems has a significant positive effect on the number of citations. This finding supports other researchers' (Amancio et al., 2012; Antelman, 2004; Egghe et al., 2013; Ertürk & Şengül, 2012; Hardy et al., 2005) findings related to the paper visibility. The results of this study indicate that there is a strong correlation between the number of citations in Google Scholar and ISI Web of Science. Therefore, the researchers can increase the impact of their research by increasing the visibility of their research papers (or paper versions). Future study is needed to determine the relationship between citation counts on the other databases such as Microsoft Academic Research, Scopus, SiteSeer index and ISI by considering journal article and conference papers.

Reference

- Aghaei Chadegani, A., Salehi, H., Yunus, M. M., Farhadi, H., Fooladi, M., Farhadi, M., & Ale Ebrahim, N. (2013). A Comparison between Two Main Academic Literature Collections: Web of Science and Scopus Databases. *Asian Social Science*, 9(5), 18-26. <http://dx.doi.org/10.5539/ass.v9n5p18>
- Aguillo, I. F. (2011). Is Google Scholar useful for Bibliometrics? A Webometric Analysis. In E. Noyons, P. Ngulube, & J. Leta (Eds.), *Proceedings of Issi 2011: The 13th Conference of the International Society for Scientometrics and Informetrics, Vols 1 and 2* (pp. 19-25). Leuven: Int Soc Scientometrics & Informetrics-Issi.
- Ale Ebrahim, N. (2013). Introduction to the Research Tools Mind Map. *Research World*, 10(4), 1-3. <http://dx.doi.org/10.5281/zenodo.7712>
- Ale Ebrahim, N., Ahmed, S., & Taha, Z. (2009). Virtual Teams: a Literature Review. *Australian Journal of Basic and Applied Sciences*, 3(3), 2653-2669. <http://dx.doi.org/10.6084/m9.figshare.1067906>
- Ale Ebrahim, N., Salehi, H., Embi, M. A., Habibi Tanha, F., Gholizadeh, H., & Motahar, S. M. (2014). Visibility and Citation Impact. *International Education Studies*, 7(4), 120-125. <http://dx.doi.org/10.5539/ies.v7n4p120>
- Ale Ebrahim, N., Salehi, H., Embi, M. A., Habibi Tanha, F., Gholizadeh, H., Motahar, S. M., & Ordi, A. (2013). Effective Strategies for Increasing Citation Frequency. *International Education Studies*, 6(11), 93-99. <http://dx.doi.org/10.5539/ies.v6n11p93>

- Amancio, D. R., Oliveira Jr, O. N., & da Fontoura Costa, L. (2012). Three-feature model to reproduce the topology of citation networks and the effects from authors' visibility on their h-index. *Journal of Informetrics*, 6(3), 427-434. <http://dx.doi.org/10.1016/j.joi.2012.02.005>
- Antelman, K. (2004). Do open-access articles have a greater research impact? *College & Research Libraries*, 65(5), 372-382. <http://dx.doi.org/10.5860/crl.65.5.372>
- Bakkalbasi, N., Bauer, K., Glover, J., & Wang, L. (2006). Three options for citation tracking: Google Scholar, Scopus and Web of Science. *Biomedical Digital Libraries*, 3(1), 7. <http://dx.doi.org/10.1186/1742-5581-3-7>
- Bensman, S. (2011). Anne-Wil Harzing: The publish or perish book: Your guide to effective and responsible citation analysis. *Scientometrics*, 88(1), 339-342. <http://dx.doi.org/10.1007/s11192-011-0388-8>
- Bornmann, L., Marx, W., Schier, H., Rahm, E., Thor, A., & Daniel, H. D. (2009). Convergent validity of bibliometric Google Scholar data in the field of chemistry-Citation counts for papers that were accepted by Angewandte Chemie International Edition or rejected but published elsewhere, using Google Scholar, Science Citation Index, Scopus, and Chemical Abstracts. *Journal of Informetrics*, 3(1), 27-35. <http://dx.doi.org/10.1016/j.joi.2008.11.r001>
- Cabezas-Clavijo, A., & Delgado-Lopez-Cozar, E. (2013). Google Scholar and the h-index in biomedicine: The popularization of bibliometric assessment. *Medicina Intensiva*, 37(5), 343-354. <http://dx.doi.org/10.1016/j.medin.2013.01.008>
- Craig, I. D., Plume, A. M., McVeigh, M. E., Pringle, J., & Amin, M. (2007). Do open access articles have greater citation impact?: A critical review of the literature. *Journal of Informetrics*, 1(3), 239-248. <http://dx.doi.org/10.1016/j.joi.2007.04.001>
- Egghe, L., Guns, R., & Rousseau, R. (2013). Measuring co-authors' contribution to an article's visibility. *Scientometrics*, 95(1), 55-67. <http://dx.doi.org/10.1007/s11192-012-0832-4>
- Ertürk, K., & Şengül, G. (2012). Self Archiving in Atılım University. In S. Kurbanoğlu, U. Al, P. Erdoğan, Y. Tonta, & N. Uçak (Eds.), *E-Science and Information Management* (Vol. 317, pp. 79-86): Springer Berlin Heidelberg. http://dx.doi.org/10.1007/978-3-642-33299-9_11
- Fooladi, M., Salehi, H., Yunus, M. M., Farhadi, M., Aghaei Chadegani, A., Farhadi, H., & Ale Ebrahim, N. (2013). Do Criticisms Overcome the Praises of Journal Impact Factor? *Asian Social Science*, 9(5), 176-182. <http://dx.doi.org/10.5539/ass.v9n5p176>
- Garfield, E. (1972). Citation analysis as a tool in journal evaluation. *Science*, 178, 471-479. <http://dx.doi.org/10.1126/science.178.4060.471>
- Hardy, R., Oppenheim, C., Brody, T., & Hitchcock, S. (2005). Open Access Citation Information.
- Hooper, S. L. (2012). Citations: not all measures are equal. *Nature*, 483(7387), 36-36. <http://dx.doi.org/10.1038/483036c>
- Jacso, P. (2012). Google Scholar Metrics for Publications The software and content features of a new open access bibliometric service. *Online Information Review*, 36(4), 604-619. <http://dx.doi.org/10.1108/14684521211254121>
- Kear, R., & Colbert-Lewis, D. (2011). Citation searching and bibliometric measures: Resources for ranking and tracking. *College & Research Libraries News*, 72(8), 470-474.
- Kousha, K., & Thelwall, M. (2007). Google Scholar citations and Google Web-URL citations: A multi-discipline exploratory analysis. *Journal of the American Society for Information Science and Technology*, 58(7), 1055-1065. <http://dx.doi.org/10.1002/asi.v58:7>
- Larsen, P. O., & von Ins, M. (2010). The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. *Scientometrics*, 84(3), 575-603. <http://dx.doi.org/10.1007/s11192-010-0202-z>
- Lawrence, S. (2001a). Free online availability substantially increases a paper's impact. *Nature*, 411(6837), 521-521. <http://dx.doi.org/10.1038/35079151>
- Lawrence, S. (2001b). Online or invisible. *Nature*, 411(6837), 521. <http://dx.doi.org/10.1038/35079151>
- McCabe, M. J., & Snyder, C. M. (2013). Does Online Availability Increase Citations? Theory and Evidence from a Panel of Economics and Business Journals: SSRN working paper.
- Orduña-Malea, E., & Delgado López-Cózar, E. (2014). Google Scholar Metrics evolution: an analysis according

- to languages. *Scientometrics*, 98(3), 2353-2367. <http://dx.doi.org/10.1007/s11192-013-1164-8>
- Pauly, D., & Stergiou, K. I. (2005). Equivalence of results from two citation analyses: Thomson ISI's Citation Index and Google's Scholar service. *Ethics in Science and Environmental Politics*, 5, 33-35.
- Rotich, D. C., & Musakali, J. J. (2013). *Publish or Perish: Remaining Academically Relevant and Visible In the Global Academic Scene through Scholarly Publishing*. Paper presented at the Conference and Programme Chairs, South Africa.
- Solomon, D. J., Laakso, M., & Björk, B.-C. (2013). A longitudinal comparison of citation rates and growth among open access journals. *Journal of Informetrics*, 7(3), 642-650. <http://dx.doi.org/10.1007/s11192-013-1164-8>
- The Thomson Corporation. (2013). Essential Science Indicators, Product Overview. Retrieved from http://www.esi.webofknowledge.com/help/h_whatish.htm
- Yue, W. P., & Wilson, C. S. (2004). Measuring the citation impact of research journals in clinical neurology: A structural equation modelling analysis. *Scientometrics*, 60(3), 317-332. <http://dx.doi.org/10.1023/B:SCIE.0000034377.93437.18>
- Zhang, C.-T. (2009). The e-Index, Complementing the h-Index for Excess Citations. *PLoS ONE*, 4(5), e5429. <http://dx.doi.org/10.1371/journal.pone.0005429>
- Zheng, J., Zhao, Z. Y., Zhang, X., Chen, D. Z., Huang, M. H., Lei, X. P., ... Zhao, Y. H. (2012). International scientific and technological collaboration of China from 2004 to 2008: A perspective from paper and patent analysis. *Scientometrics* 91(1), 65-80. <http://dx.doi.org/10.1007/s11192-011-0529-0>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

Ratio Estimators Using Coefficient of Variation and Coefficient of Correlation

Prayad Sangngam¹

¹ Department of Statistics, Faculty of Science, Silpakorn University, Nakorn Pathom, Thailand

Correspondence: Prayad Sangngam, Department of Statistics, Faculty of Science, Silpakorn University, Nakorn Pathom 73000, Thailand. E-mail: prayad@su.ac.th

Received: June 4, 2014 Accepted: June 10, 2014 Online Published: August 6, 2014

doi:10.5539/mas.v8n5p70 URL: <http://dx.doi.org/10.5539/mas.v8n5p70>

Abstract

This paper introduces ratio estimators of the population mean using the coefficient of variation of study variable and auxiliary variables together with the coefficient of correlation between the study and auxiliary variables under simple random sampling and stratified random sampling. These ratio estimators are almost unbiased. The mean square errors of the estimators and their estimators are given. Sample size estimation in both sampling designs are presented. An optimal sample size allocation in stratified random sampling is also suggested. Based on theoretical study, it can be shown that these ratio estimators have smaller MSE than the unbiased estimators. Moreover, the empirical study indicates that these ratio estimators have smallest MSE compared to the existing ones.

Keywords: ratio estimator, sample size estimation, coefficient of variation, coefficient of correlation

1. Introduction

Consider a population of N units with observations (x_i, y_i) for $i=1, 2, \dots, N$ where y_i is a value of study variable and x_i is a value of auxiliary variable. Under simple random sampling without replacement, an unbiased estimator of the population mean $\bar{Y} = \frac{1}{N} \sum_{i=1}^N y_i$ is the sample mean $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$. The variance of the unbiased estimator is

$$V(\bar{y}) = \frac{1-f}{n} S_y^2, \quad (1.1)$$

where $f = \frac{n}{N}$ and $S_y^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{Y})^2$.

A common ratio estimator is $\bar{y}_R = \bar{y} \frac{\bar{X}}{\bar{x}}$ where \bar{X} and \bar{x} are the population and sample means of the auxiliary variable, respectively. The efficiency of the ratio estimator depends on the coefficient of variation of auxiliary variable (C_x) and coefficient of variation of study variable (C_y). Murthy (1964) has suggested that if $\rho > \frac{C_x}{2C_y}$, the ratio estimator performs better than the unbiased estimator where ρ is the correlation coefficient between x and y . The approximate bias and mean square error (MSE) of the ratio estimator are as follows:

$$B(\bar{y}_R) = \frac{1-f}{n} \left(\frac{R}{\bar{X}} S_x^2 - \frac{1}{\bar{X}} S_{xy} \right), \quad (1.2)$$

$$MSE(\bar{y}_R) = \frac{1-f}{n} (S_y^2 + R^2 S_x^2 - 2RS_{xy}), \quad (1.3)$$

where $R = \frac{\bar{Y}}{\bar{X}}$, $S_x^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{X})^2$ and $S_{xy} = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{Y})(x_i - \bar{X})$. When the C_x is known,

Sissodia and Dwivedi (1981) has proposed a modified ratio estimator, $\bar{y}_{SD} = \bar{y} \frac{\bar{X} + C_x}{\bar{X} + C_x}$. The approximate bias and MSE of the estimator are

$$B(\bar{y}_{SD}) = \frac{1-f}{n} \left(\frac{R_x}{\bar{X}} S_x^2 - \frac{1}{\bar{X}} S_{xy} \right), \tag{1.4}$$

$$MSE(\bar{y}_{SD}) = \frac{1-f}{n} (S_y^2 + R_x^2 S_x^2 - 2R_x S_{xy}), \tag{1.5}$$

where $R_x = \frac{\bar{Y}}{\bar{X} + C_x}$. Sampath (2005) used the coefficient of variation of the study variable to improve the

unbiased estimator as $\bar{y}_s = \left(1 + \frac{1-f}{n} C_y^2\right)^{-1} \bar{y}$. The approximate bias and MSE of this estimator are

$$B(\bar{y}_s) = \left(1 + \frac{1-f}{n} C_y^2\right)^{-1} \bar{Y} - \bar{Y}, \tag{1.6}$$

$$MSE(\bar{y}_s) = \frac{1-f}{n} S_y^2 \left(1 + \frac{1-f}{n} C_y^2\right)^{-1}. \tag{1.7}$$

In addition, there are several authors, such as Upadhyaya and Singh (1999), Singh and Tailor (2003), who have developed various ratio estimators under simple random sampling.

If the study variable has different mean values in different subpopulations, it is advantageous to draw a sample by stratified random sampling. In stratified sampling, a population is partitioned into L strata. A stratum h contains N_h units with observations (x_{hi}, y_{hi}) where $h = 1, 2, \dots, L$ and $i = 1, 2, \dots, N_h$. An unbiased

estimator of \bar{Y} under stratified random sampling is given by $\bar{y}_{st} = \sum_{h=1}^L W_h \bar{y}_h$ where $W_h = \frac{N_h}{N}$ is the stratum weight and \bar{y}_h is the sample mean of the study variable in stratum h. The variance of the unbiased estimator is

$$V(\bar{y}_{st}) = \sum_{h=1}^L W_h^2 \gamma_h S_{y_h}^2, \tag{1.8}$$

where $\gamma_h = (1 - f_h)/n_h$, $f_h = \frac{n_h}{N_h}$ is sampling fraction in stratum h, n_h is a sample size in stratum h and $S_{y_h}^2$ is the variance of the study variable in stratum h. There are two types of ratio estimators in stratified random sampling, namely combined and separate ratio estimators.

The combined ratio estimator is given by $\bar{y}_{RC} = \frac{\bar{y}_{st}}{\bar{x}_{st}} \bar{X}$, where $\bar{x}_{st} = \sum_{h=1}^L W_h \bar{x}_h$ is an unbiased estimator of the population mean \bar{X} and \bar{x}_h is the sample mean of auxiliary variable in stratum h (Cochran, 1977). The approximate mean squared error of the combined ratio estimator is

$$MSE(\bar{y}_{RC}) = \sum_{h=1}^L W_h^2 \gamma_h (S_{y_h}^2 + R^2 S_{x_h}^2 - 2R S_{y_h x_h}), \tag{1.9}$$

where $R = \frac{\bar{Y}}{\bar{X}}$ is the population ratio, $S_{x_h}^2$ is the variance of auxiliary variable in stratum h and $S_{y_h x_h}$ is the covariance between auxiliary and study variables in stratum h. The approximate bias of the combined ratio estimator is

$$B(\bar{y}_{RC}) = \sum_{h=1}^L W_h^2 \gamma_h \left(\frac{R}{\bar{X}} S_{x_h}^2 - \frac{1}{\bar{X}} S_{y_h x_h} \right). \tag{1.10}$$

The separate ratio estimator is given by $\bar{y}_{RS} = \sum_{h=1}^L W_h \frac{\bar{y}_h \bar{X}_h}{\bar{x}_h}$. The approximate MSE of the separate ratio estimator can be given by

$$MSE(\bar{y}_{RS}) = \sum_{h=1}^L W_h^2 \gamma_h (S_{y_h}^2 + R_h^2 S_{x_h}^2 - 2R_h S_{y_h x_h}), \tag{1.11}$$

where $R_h = \frac{\bar{Y}_h}{\bar{X}_h}$ is the population ratio in stratum h. The approximate bias of the separate ratio estimator is

$$B(\bar{y}_{RS}) = \sum_{h=1}^L W_h \gamma_h \left(\frac{R_h}{\bar{X}_h} S_{xh}^2 - \frac{1}{\bar{X}_h} S_{xyh} \right). \tag{1.12}$$

Kadilar and Cingi (2003) have proposed several combined ratio estimators. The simplest one based on the

Sissodia and Dwivedi (1981) estimator is defined as $\bar{y}_{KC} = \bar{y}_{st} \frac{\sum_{h=1}^L W_h (\bar{X}_h + C_{xh})}{\sum_{h=1}^L W_h (\bar{X}_h + C_{xh})}$ where C_{xh} is the

coefficient of variation of the auxiliary variable in stratum h. The MSE and bias of this estimator are approximated as follows:

$$MSE(\bar{y}_{KC}) = \sum_{h=1}^L W_h^2 \gamma_h \left(S_{yh}^2 + R_{KC}^2 S_{xh}^2 - 2R_{KC} S_{xyh} \right), \tag{1.13}$$

$$B(\bar{y}_{KC}) = \sum_{h=1}^L W_h^2 \gamma_h \left(\frac{R_{KC}}{\bar{X}_{KC}} S_{xh}^2 - \frac{1}{\bar{X}_{KC}} S_{xyh} \right), \tag{1.14}$$

where $R_{KC} = \frac{\bar{Y}_{st}}{\bar{X}_{KC}} = \frac{\bar{Y}_{st}}{\sum_{h=1}^L W_h (\bar{X}_h + C_{xh})}$.

Kadilar and Cingi (2005) have improved the combined ratio estimator in stratified random sampling based on the estimator introduced by Prasad (1989). However, these estimators depend on several unknown parameters and therefore are difficult to use.

In Sections 2 and 3, the ratio estimators based on the coefficient of variation and correlation in simple random sampling and stratified random sampling are introduced, respectively. The approximate bias and MSE of the estimator are derived. An estimator of the MSE is given. The sample size estimation and an optimal allocation of sample size in stratified random sampling is presented. The comparison of the efficiency between the proposed estimator and unbiased estimator is theoretically provided. Hypothetical populations are used to compare the properties of the presented estimators with the existing ones.

2. Estimation in Simple Random Sampling

2.1 Parameter Estimation

Consider the following ratio estimator for the population mean of the study variable,

$$\bar{y}_c = \bar{y} \frac{\bar{X} + c}{\bar{X} + c} \tag{2.1}$$

where c is a real constant to be determined such that the $MSE(\bar{y}_c)$ is minimized. Note that when c is equal to 0, this estimator is reduced to the usual ratio estimator and when c is equal to C_x this estimator become

the estimator of Sissodia and Dwivedi (1981). To obtain the MSE and bias of the estimator (2.1), let $e_1 = \frac{\bar{y} - \bar{Y}}{\bar{Y}}$

and $e_2 = \frac{\bar{x} - \bar{X}}{\bar{X} + c}$. It can be shown that $E(e_1) = 0$, $E(e_2) = 0$, $E(e_1^2) = \frac{V(\bar{y})}{\bar{Y}^2}$, $E(e_1 e_2) = \frac{Cov(\bar{x}, \bar{y})}{\bar{Y}(\bar{X} + c)}$ and

$E(e_2^2) = \frac{V(\bar{x})}{(\bar{X} + c)^2}$. The estimator \bar{y}_c can be written as $\bar{y}_c = \bar{Y}(1 + e_1)(1 + e_2)^{-1}$. Using Taylor series

approximation, we obtain $\bar{y}_c = \bar{Y}(1 + e_1 - e_2 + e_2^2 - e_1 e_2 + \dots)$. When the terms of degree greater than two are ignored, we get the approximate bias of the estimator \bar{y}_c as

$$B(\bar{y}_c) = \left(\frac{1-f}{n} \right) \left[\frac{R_c S_x^2}{(\bar{X} + c)} - \frac{S_{xy}}{(\bar{X} + c)} \right], \tag{2.2}$$

where $R_c = \frac{\bar{Y}}{\bar{X} + c}$. Similarly, Taylor's formula can be used to approximate MSE of the estimator as

$$\text{MSE}(\bar{y}_c) = \frac{1-f}{n} (S_y^2 + R_c^2 S_x^2 - 2R_c S_{xy}) \quad (2.3)$$

Minimizing (2.3) with respect to c , we get the optimum value of c as $c = c^* = \frac{\bar{Y} S_x^2}{S_{xy}} - \bar{X}$. Substituting c^* for c in (2.1), (2.2) and (2.3) and using algebra, we obtain the optimum estimator, its bias and MSE as follows,

$$\bar{y}_{c^*} = \bar{y} \frac{\bar{X} C_x}{C_y \rho (\bar{x} - \bar{X}) + \bar{X} C_x} \quad (2.4)$$

$$B(\bar{y}_{c^*}) = 0 \quad (2.5)$$

$$\text{MSE}(\bar{y}_{c^*}) = \frac{(1-f)}{n} S_y^2 (1 - \rho^2) \quad (2.6)$$

Note that the optimum estimator is almost unbiased and its MSE is always smaller than the variance of the unbiased estimator. In addition, the optimum estimator can be applied for both populations with positive and negative coefficient of correlation. For a sample estimate of the MSE, one can substitute the sample estimate of S_y^2 which gives

$$\hat{\text{MSE}}(\bar{y}_{c^*}) = \frac{(1-f)}{n} s_y^2 (1 - \rho^2) \quad (2.7)$$

where s_y^2 is the sample variance of the study variable.

2.2 Sample Size Estimation

Sample size estimation is one of the important aspects in sample surveys. If the sample size is too small, the sampling error may be too large. However, too large sample size implies a waste of resources. We would like to specify a sample size that is sufficiently large to ensure a high probability that the estimate closes to the parameter. Under simple random sampling, the population mean of the study variable (\bar{Y}) is estimated with the optimum estimator \bar{y}_{c^*} . To obtain the desired sample size, one can specify the margin of error d and the probability α such that $P(|\bar{y}_{c^*} - \bar{Y}| \geq d) = \alpha$. Under some technical conditions as shown in Scott and Wu (1981) and Hajek (1960), we can show that \bar{y}_{c^*} is asymptotically normal distributed with mean \bar{Y} and variance $\text{MSE}(\bar{y}_{c^*})$. To obtain the absolute precision, we can find a value of n that satisfies $d / \sqrt{\text{MSE}(\bar{y}_{c^*})} = z_{\alpha/2}$ where $z_{\alpha/2}$ denotes the upper $\alpha/2$ point of the standard normal distribution. Solving for n , we have

$$n = \frac{n_0}{1 + \frac{n_0}{N}} \quad (2.8)$$

where $n_0 = \frac{z_{\alpha/2}^2 S_y^2 (1 - \rho^2)}{d^2}$. If the population size N is large relative to the sample size n , the formula of the sample size reduces to n_0 .

2.3 Comparison of Efficiency

In this section, the properties of the estimators in simple random sampling are compared. The relative efficiency of the optimum estimator and unbiased estimator is considered as follows:

$$e(\bar{y}, \bar{y}_{c^*}) = \frac{E(\bar{y} - \bar{Y})^2}{E(\bar{y}_{c^*} - \bar{Y})^2} = \frac{1}{1 - \rho^2} \quad (2.9)$$

This shows that the optimum estimator is always more efficient than the unbiased estimator because $0 < \rho^2 < 1$. The efficiency depends on the coefficient of correlation: if the coefficient of correlation increases then the efficiency also increases.

To compare the properties of the optimum estimator with the others, we consider hypothetical populations with vary characteristics. In this work, the coefficients of correlation in the populations are $\rho = 0.1, 0.2, \dots, 0.9$. In each population, the coefficients of variations are $C_x = 0.2, C_y = 0.2$ and the population means $\bar{X} = 5,000, \bar{Y} = 5,000$. With varying sample sizes, the biases and MSEs of the estimators are given in Table 1 and Table 2, respectively. The biases and MSEs are computed by the formulas in the previous sections. In Table 1, as expected, the absolute bias of the unbiased and optimum ratio estimators are always equal to 0. The estimator \bar{y}_s has the largest absolute bias among the compared estimators. The bias of the estimator \bar{y}_s is negative because the estimator is constructed by using a constant in which its value less than 1 multiplying the unbiased estimator. The bias of the estimator \bar{y}_s does not depend on the coefficient of correlation. Observe that the absolute biases of the estimator \bar{y}_{SD} are smaller than of the estimator \bar{y}_R . Given a sample size, when the coefficient of correlation increases the absolute bias of the two estimators \bar{y}_R and \bar{y}_{SD} decrease. Given a coefficient of correlation, the absolute bias of \bar{y}_s, \bar{y}_R and \bar{y}_{SD} decrease when the sample size increases. Table 2 shows that the optimum ratio estimator has smallest MSE among the compared estimators. The MSEs of the two estimators \bar{y} and \bar{y}_s do not depend on the coefficient of correlation. When $\rho > 0.5$ the MSEs of the two estimators \bar{y}_R and \bar{y}_{SD} are less than those of the unbiased estimator. Given a sample size, when the coefficient of correlation increases the MSEs of the three estimators \bar{y}_R, \bar{y}_{SD} and \bar{y}_{c^*} decrease. Given a coefficient of correlation, the MSEs of all estimators decrease when the sample size increases.

Table 1. Biases of the estimators in simple random sampling

n	ρ	$B(\bar{y})$	$B(\bar{y}_R)$	$B(\bar{y}_{SD})$	$B(\bar{y}_s)$	$B(\bar{y}_{c^*})$
30	0.1	0	5.9964	5.9961	-6.6538	0
	0.2	0	5.3301	5.3299	-6.6538	0
	0.3	0	4.6639	4.6636	-6.6538	0
	0.4	0	3.9976	3.9973	-6.6538	0
	0.5	0	3.3313	3.3311	-6.6538	0
	0.6	0	2.6651	2.6648	-6.6538	0
	0.7	0	1.9988	1.9985	-6.6538	0
	0.8	0	1.3325	1.3323	-6.6538	0
	0.9	0	0.6663	0.6660	-6.6538	0
50	0.1	0	3.5964	3.5962	-3.9928	0
	0.2	0	3.1968	3.1966	-3.9928	0
	0.3	0	2.7972	2.7970	-3.9928	0
	0.4	0	2.3976	2.3974	-3.9928	0
	0.5	0	1.9980	1.9978	-3.9928	0
	0.6	0	1.5984	1.5982	-3.9928	0
	0.7	0	1.1988	1.1986	-3.9928	0
	0.8	0	0.7992	0.7990	-3.9928	0
	0.9	0	0.3996	0.3994	-3.9928	0
100	0.1	0	1.7964	1.7963	-1.9952	0
	0.2	0	1.5968	1.5967	-1.9952	0
	0.3	0	1.3972	1.3971	-1.9952	0
	0.4	0	1.1976	1.1975	-1.9952	0
	0.5	0	0.9980	0.9979	-1.9952	0
	0.6	0	0.7984	0.7983	-1.9952	0
	0.7	0	0.5988	0.5987	-1.9952	0
	0.8	0	0.3992	0.3991	-1.9952	0
	0.9	0	0.1996	0.1995	-1.9952	0

Table 2. MSEs of the estimators in simple random sampling

n	ρ	MSE(\bar{y})	MSE(\bar{y}_R)	MSE(\bar{y}_{SD})	MSE(\bar{y}_s)	MSE(\bar{y}_{c^*})
30	0.1	33313.33	59964.00	59961.60	33269.00	32980.20
	0.2	33313.33	53301.33	53299.20	33269.00	31980.80
	0.3	33313.33	46638.67	46636.80	33269.00	30315.13
	0.4	33313.33	39976.00	39974.40	33269.00	27983.20
	0.5	33313.33	33313.33	33312.00	33269.00	24985.00
	0.6	33313.33	26650.67	26649.60	33269.00	21320.53
	0.7	33313.33	19988.00	19987.20	33269.00	16989.80
	0.8	33313.33	13325.33	13324.80	33269.00	11992.80
	0.9	33313.33	6662.67	6662.40	33269.00	6329.53
50	0.1	19980.00	35964.00	35962.56	19964.04	19780.20
	0.2	19980.00	31968.00	31966.72	19964.04	19180.80
	0.3	19980.00	27972.00	27970.88	19964.04	18181.80
	0.4	19980.00	23976.00	23975.04	19964.04	16783.20
	0.5	19980.00	19980.00	19979.20	19964.04	14985.00
	0.6	19980.00	15984.00	15983.36	19964.04	12787.20
	0.7	19980.00	11988.00	11987.52	19964.04	10189.80
	0.8	19980.00	7992.00	7991.68	19964.04	7192.80
	0.9	19980.00	3996.00	3995.84	19964.04	3796.20
100	0.1	9980.00	17964.00	17963.28	9976.02	9880.20
	0.2	9980.00	15968.00	15967.36	9976.02	9580.80
	0.3	9980.00	13972.00	13971.44	9976.02	9081.80
	0.4	9980.00	11976.00	11975.52	9976.02	8383.20
	0.5	9980.00	9980.00	9979.60	9976.02	7485.00
	0.6	9980.00	7984.00	7983.68	9976.02	6387.20
	0.7	9980.00	5988.00	5987.76	9976.02	5089.80
	0.8	9980.00	3992.00	3991.84	9976.02	3592.80
	0.9	9980.00	1996.00	1995.92	9976.02	1896.20

3. Estimation in Stratified Random Sampling

3.1 Parameter Estimation

In stratified random sampling, when \bar{X}_h , C_{xh} , C_{yh} and ρ_{xh} in stratum h are known, the separate ratio estimator can be modified as

$$\bar{y}_{RS-C} = \sum_{h=1}^L W_h \frac{\bar{y}_h \bar{X}_h C_{xh}}{C_{yh} \rho_h (\bar{x}_h - \bar{X}_h) + \bar{X}_h C_{xh}} \tag{3.1}$$

Since this estimator is constructed from the optimum ratio estimator, we call this estimator “optimum separate ratio estimator”. To obtain the MSE and bias of the optimum separate ratio estimator, applying the MSE and bias

of $\bar{y}_{c^*h} = \frac{\bar{y}_h (\bar{X}_h + C_{xh})}{\bar{X}_h + C_{xh}}$ under simple random sampling to draw in stratum h, yields

$$B(\bar{y}_{RS-C}) = 0, \tag{3.2}$$

$$MSE(\bar{y}_{RS-C}) = \sum_{h=1}^L W_h^2 \gamma_h S_{yh}^2 (1 - \rho_h^2), \tag{3.3}$$

For estimating the $MSE(\bar{y}_{RS-C})$, we substitute the sample estimates to obtain

$$\hat{MSE}(\bar{y}_{RS-C}) = \sum_{h=1}^L W_h^2 \gamma_h S_{y_h}^2 (1 - \rho_h^2). \quad (3.4)$$

Note that the bias of the optimum separate ratio estimator is the cumulative bias of an optimum ratio estimate in each stratum which closes to zero. In addition, we found that the MSE of the estimator is also smaller than the variance of the unbiased estimator in (1.8).

3.2 Optimum Sample Size Allocation

Given a total sample size n and using the optimum separate ratio estimator, one may choose how to allocate the sample size among the L strata. In this section, the allocation scheme which minimizes the MSE of the estimator by fixing the total sample size is considered. That is, we need the values of n_1, n_2, \dots, n_L which minimize

$MSE(\bar{y}_{RS-C}) = \sum_{h=1}^L W_h^2 \gamma_h S_{y_h}^2 (1 - \rho_h^2)$ subject to the condition $n = \sum_{h=1}^L n_h$. The sample size allocated to each stratum is

$$n_h = n \frac{N_h S_h \sqrt{1 - \rho_h^2}}{\sum_{h=1}^L N_h S_h \sqrt{1 - \rho_h^2}} \quad ; h = 1, 2, \dots, L. \quad (3.5)$$

Thus, the optimum scheme allocates larger sample sizes to strata with larger variances and larger stratum sizes but smaller sample sizes to strata with larger coefficients of correlation.

3.3 Sample Size Estimation

The formula (3.5) gives n_h in terms of n , but in practice, we do not yet know what value of n is. This section presents a formula for the determination of n under the optimum sample size allocation. It is assumed that the optimum separate ratio estimate has a specified mean squared error M . When n_h, N_h and $N_h - n_h$ are all sufficient large and the technical conditions in Scott and Wu (1981) hold, we can show that the estimator \bar{y}_{RS-C} has also the asymptotic normal distribution. If the margin of error d has been given, then

$M = (d / z_{\alpha/2})^2$. Let $n_h = n w_h$, where $w_h = \frac{W_h S_h \sqrt{1 - \rho_h^2}}{\sum_{h=1}^L W_h S_h \sqrt{1 - \rho_h^2}}$. So, the mean square error of \bar{y}_{RS-C} is

$$M = \frac{1}{n} \sum_{h=1}^L \frac{W_h^2 S_{y_h}^2 (1 - \rho_h^2)}{w_h} - \frac{1}{N} \sum_{h=1}^L W_h S_{y_h}^2 (1 - \rho_h^2) = \left(\frac{d}{z_{\alpha/2}} \right)^2.$$

Solving for n , we have

$$n = \frac{n_0}{1 + \frac{z_{\alpha/2}^2}{Nd^2} \sum_{h=1}^L W_h S_h^2 \sqrt{1 - \rho_h^2}}, \quad (3.6)$$

$$\text{where } n_0 = \frac{z_{\alpha/2}^2 \left(\sum_{h=1}^L W_h S_h \sqrt{1 - \rho_h^2} \right)^2}{d^2}.$$

3.4 Comparison of Efficiency

In this section, we compare the properties of the proposed optimum separate ratio estimator with the existing ones in stratified random sampling. The relative efficiency of the optimum separate ratio estimator and unbiased estimator is

$$e(\bar{y}_{st}, \bar{y}_{RS-C}) = \frac{\sum_{h=1}^L W_h^2 \gamma_h S_{y_h}^2}{\sum_{h=1}^L W_h^2 \gamma_h S_{y_h}^2 (1 - \rho_h^2)}. \quad (3.7)$$

This shows that the optimum separate ratio estimator is always more efficient than the unbiased estimator. The efficiency depends on the coefficient of correlation in stratum. If the correlation coefficient increases then the efficiency also increases.

To compare the properties of the optimum separate ratio estimator with the other estimators, we consider the following hypothetical populations. Each population consists of $N = 50,000$ units and is divided into $L = 2$ strata of which sizes are $N_1 = 20,000$ and $N_2 = 30,000$. The coefficients of variations are $C_{x_1} = C_{y_1} = 0.2$ and $C_{x_2} = C_{y_2} = 0.3$. The population means are given by $\bar{X}_1 = \bar{Y}_1 = 500$ and $\bar{X}_2 = \bar{Y}_2 = 1,000$. The coefficients of correlations are $\rho_h = 0.1, 0.2, \dots, 0.9$. We set the total sample size $n = 150$ with three allocations, namely equal allocation $n_h = \frac{n}{L}$, proportional allocation $n_h = \frac{nN_h}{N}$ and optimal allocation as in (3.5). The biases and MSEs of the estimators are given in Table 3 and Table 4, respectively. The biases and MSEs are computed by the formulas in the above sections.

In Table 3, the absolute biases of the unbiased and optimum separate ratio estimators are always equal to 0. The absolute bias of the estimator \bar{y}_{RS} is smallest among the compared estimators. The absolute bias of the estimator \bar{y}_{KC} is smaller than that of the estimator \bar{y}_{RC} . Given an sample size allocation, when the coefficient of correlation increases the absolute biases of the three estimators \bar{y}_{RC} , \bar{y}_{RS} , and \bar{y}_{KC} decrease. Using the optimum allocation, the absolute biases of the estimators \bar{y}_{RC} , \bar{y}_{RS} and \bar{y}_{KC} are smallest among the three allocations.

Table 4 presents that the optimum separate ratio estimator gives the smallest MSE among the compared estimators. Observe that the MSE of the unbiased estimator \bar{y}_{st} does not depend on the coefficient of correlation because it does not use the information of the auxillary variable. When $\rho_h > 0.5$, the MSEs of the estimators \bar{y}_{RC} , \bar{y}_{RS} and \bar{y}_{KC} are less than that of the unbiased estimator. Given an sample size allocation, when the coefficient of correlation increases, the MSEs of the estimators \bar{y}_{RC} , \bar{y}_{RS} , \bar{y}_{KC} and \bar{y}_{RS-C} decrease. The MSEs of all estimators are smallest under the optimum allocation.

Table 3. Biases of the estimators in stratified random sampling

Allocation	n_1	n_2	ρ_1	ρ_2	ρ	$B(\bar{y}_{st})$	$B(\bar{y}_{RC})$	$B(\bar{y}_{RS})$	$B(\bar{y}_{KC})$	$B(\bar{y}_{RS-C})$
equal	75	75	0.1	0.1	0.56	0	0.5087	0.4261	0.5083	0
			0.2	0.2	0.61	0	0.4522	0.3787	0.4518	0
			0.3	0.3	0.66	0	0.3957	0.3314	0.3953	0
			0.4	0.4	0.71	0	0.3391	0.2841	0.3388	0
			0.5	0.5	0.75	0	0.2826	0.2367	0.2823	0
			0.6	0.6	0.80	0	0.2261	0.1894	0.2258	0
			0.7	0.7	0.85	0	0.1696	0.1420	0.1693	0
			0.8	0.8	0.90	0	0.1130	0.0947	0.1128	0
			0.9	0.9	0.95	0	0.0565	0.0473	0.0563	0
proportion	60	90	0.1	0.1	0.56	0	0.4337	0.3709	0.4334	0
			0.2	0.2	0.61	0	0.3855	0.3297	0.3852	0
			0.3	0.3	0.66	0	0.3373	0.2885	0.3371	0
			0.4	0.4	0.71	0	0.2891	0.2473	0.2889	0
			0.5	0.5	0.75	0	0.2409	0.2060	0.2407	0
			0.6	0.6	0.80	0	0.1928	0.1648	0.1925	0
			0.7	0.7	0.85	0	0.1446	0.1236	0.1444	0
			0.8	0.8	0.90	0	0.4337	0.3709	0.4334	0
			0.9	0.9	0.95	0	0.3855	0.3297	0.3852	0
optimum	27	123	0.1	0.1	0.56	0	0.3617	0.3421	0.3614	0
			0.2	0.2	0.61	0	0.3215	0.3041	0.3213	0
			0.3	0.3	0.66	0	0.2813	0.2661	0.2811	0
			0.4	0.4	0.71	0	0.2411	0.2281	0.2409	0

0.5	0.5	0.75	0	0.2009	0.1900	0.2007	0
0.6	0.6	0.80	0	0.1608	0.1520	0.1606	0
0.7	0.7	0.85	0	0.1206	0.1140	0.1204	0
0.8	0.8	0.90	0	0.0804	0.0760	0.0802	0
0.9	0.9	0.95	0	0.0402	0.0380	0.0400	0

ρ is the coefficient of correlation in the whole population.

Table 4. MSE of the estimators in stratified random sampling

Allocation	n_1	n_2	ρ_1	ρ_2	ρ	$MSE(\bar{y}_{st})$	$MSE(\bar{y}_{RC})$	$MSE(\bar{y}_{RS})$	$MSE(\bar{y}_{RC})$	$MSE(\bar{y}_{RS,c})$
equal	75	75	0.1	0.1	0.56	452.17	813.91	813.91	813.65	447.65
			0.2	0.2	0.61	452.17	723.48	723.48	723.24	434.09
			0.3	0.3	0.66	452.17	633.04	633.04	632.84	411.48
			0.4	0.4	0.71	452.17	542.61	542.61	542.43	379.83
			0.5	0.5	0.75	452.17	452.17	452.17	452.03	339.13
			0.6	0.6	0.80	452.17	361.74	361.74	361.62	289.39
			0.7	0.7	0.85	452.17	271.30	271.30	271.22	230.61
			0.8	0.8	0.90	452.17	180.87	180.87	180.81	162.78
			0.9	0.9	0.95	452.17	90.43	90.43	90.41	85.91
proportion	60	90	0.1	0.1	0.56	385.51	693.91	693.91	693.69	381.65
			0.2	0.2	0.61	385.51	616.81	616.81	616.61	370.09
			0.3	0.3	0.66	385.51	539.71	539.71	539.53	350.81
			0.4	0.4	0.71	385.51	462.61	462.61	462.46	323.83
			0.5	0.5	0.75	385.51	385.51	385.51	385.38	289.13
			0.6	0.6	0.80	385.51	308.41	308.41	308.31	246.72
			0.7	0.7	0.85	385.51	231.30	231.30	231.23	196.61
			0.8	0.8	0.90	385.51	154.20	154.20	154.15	138.78
			0.9	0.9	0.95	385.51	77.10	77.10	77.08	73.25
optimum	27	123	0.1	0.1	0.56	321.51	578.71	578.71	578.52	318.29
			0.2	0.2	0.61	321.51	514.41	514.41	514.24	308.65
			0.3	0.3	0.66	321.51	450.11	450.11	449.96	292.57
			0.4	0.4	0.71	321.51	385.81	385.81	385.68	270.07
			0.5	0.5	0.75	321.51	321.51	321.51	321.40	241.13
			0.6	0.6	0.80	321.51	257.21	257.21	257.12	205.76
			0.7	0.7	0.85	321.51	192.90	192.90	192.84	163.97
			0.8	0.8	0.90	321.51	128.60	128.60	128.56	115.74
			0.9	0.9	0.95	321.51	64.30	64.30	64.28	61.09

ρ is the coefficient of correlation in the whole population.

4. Discussion

In simple random sampling, the optimum ratio estimator and its variance estimate depend on the coefficient of variation, the coefficient of correlation and the mean of the auxiliary variable in the whole population. Similarly, the optimum separate ratio estimator and its variance estimate are in terms of the coefficients of variation, the coefficient of correlation and the means of the auxiliary variable in all strata. In practice, sample estimates of these parameters may be used to substitute in the formulas of these estimates.

In simple random sampling, the relative efficiency of the optimum ratio estimator and the unbiased estimator depends on the coefficient of correlation ρ . When the coefficient of correlation between the study and auxiliary

variable is weak, then the relative efficiency will be low so that the unbiased estimator is almost as good as the ratio estimators. For example, if $\rho \leq 0.2$, then $e(\bar{y}, \bar{y}_{c^*}) \leq 1.042$. This means that if we increase the sample size about 4.2%, the unbiased estimator will have the most efficiency among the estimators in the class of $\left\{ \bar{y}_c = \bar{y} \frac{\bar{X} + c}{\bar{X} + c} : -\infty < c < \infty \right\}$. Therefore, in case of $\rho \leq 0.2$ we suggest using the unbiased estimator because it uses only the information of the study variable and we do not need to collect the data of the auxiliary variables. In stratified random sampling, when $\text{Max}\{\rho_h\} < 0.2$, we also recommend the unbiased estimator. For future studies, we can consider applying the ratio estimator in adaptive sampling schemes as suggested by Thompson (1990) and Sangngam (2013) for.

Acknowledgements

This work was supported by the Faculty of Science, Silpakorn University, Nakorn Pathom, Thailand.

References

- Cochran, W. G. (1977). *Sampling Techniques*. Newyork: Jonh Wiley and Sons.
- Hajek, J. (1960). Limiting distributions in simple random sampling from a finite population. *Publication of the Mathematical Institute of the Hungarian Academy of Scienees*, 5, 361-374.
- Kadilar, C., & Cingi, H. (2003). Ratio estimators in stratified random sampling. *Biometrical Journal*, 45(2), 218–225. <http://dx.doi.org/10.1002/bimj.200390007>
- Kadilar, C., & Cingi, H. (2005). A new ratio estimator in stratified random sampling. *Communication in Statistics Theory and Method*, 34, 597-620. <http://dx.doi.org/10.1081/STA-200052156>
- Murthy, M. N. (1964). Product method of estimation. *Sankhya*, 26, 69-74.
- Prasad, B. (1989). Some improved ratio type estimators of population mean and ratio infinite population sample surveys. *Communication in Statistics Theory and Method*, 18(1), 379–392. <http://dx.doi.org/10.1080/03610928908829905>
- Sampath, S (2005). *Sampling Theory and Methods*. Alpha Science International Ltd. Harrow, U.K.
- Sangngam, P. (2013). Unequal Probability Inverse Adaptive Cluster Sampling. *Chiang Mai Journal of Science*. 40(4), 736-742.
- Scott, A., & Wu, C. (1981). On the asymptotic distribution of ratio and regression estimators. *Journal of American Statistical Assocation*, 76(373), 98-102. <http://dx.doi.org/10.1080/01621459.1981.10477612>
- Singh, H. P., & Tailor, R. (2003). Use of known correlation coefficient in estimating the finite population mean. *Statistics in Transition*, 6, 555–560.
- Sisodia, B. V. S., & Dwivedi, V. K. (1981). A modified ratio estimator using coefficient of variation of auxiliary variable. *Journal of Indian Society of Agricultural Statistics*, 33, 13–18.
- Thompson, S. K. (1990). Adaptive cluster sampling. *Journal of the American Statistical Association*, 85, 1050-1059.
- Upadhyaya, L. N., & Singh, H. P. (1999). Use of transformed auxiliary variable in estimating the finite population mean. *Biometrical Journal*, 41(5), 627–636. [http://dx.doi.org/10.1002/\(SICI\)1521-4036\(199909\)41:5<627::AID-BIMJ627>3.0.CO;2-W](http://dx.doi.org/10.1002/(SICI)1521-4036(199909)41:5<627::AID-BIMJ627>3.0.CO;2-W)

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

Impact of Landuse Change on River Floodplain Using Public Domain Hydraulic Model

Abolghasem Akbari¹, Golamali Mozafari², Mohsen Fanodi² & Maliheh Sadat Hemmesy³

¹ Faculty of Civil Engineering & Earth Resources, University Malaysia Pahang, Kuantan, Malaysia

² Department of Geography, Yazd University, Iran

³ Department of Geography, University of Lorestan, Khoramabad, Iran

Correspondence: Abolghasem Akbari, Faculty of Civil Engineering & Earth Resources, University Malaysia Pahang, Lebuhraya Tun Razak 26300 Gambang, Kuantan, Malaysia. E-mail: akbariinbox@yahoo.com

Received: May 5, 2014

Accepted: May 8, 2014

Online Published: August 17, 2014

doi:10.5539/mas.v8n5p80

URL: <http://dx.doi.org/10.5539/mas.v8n5p80>

Abstract

Floodplains are land areas adjacent to rivers and streams that are subjected to recurring inundation. Owing to their continually changing nature, floodplains and other flood-prone areas need to be examined in the light of how they might affect or be affected by landuse change. In this research, the effect of land use changes on floodplain is investigated. Major landuse change has occurred in Azaran watershed during the investigation period. Irrigated farmland has decreased by about 52%. However, bare lands, dry farm lands, and rangeland have increased by 12.40%, 17.25% and 14.46%, respectively. The extent of floodplain was determined based on the annual maximum instantaneous flood for different return periods using Hydrologic Engineering Centers River Analysis System (HEC-RAS) coupled with HEC-GeoRAS which is a tool for processing geospatial data in ArcGIS. Water surface profile data and velocity data exported from HEC-RAS simulations were processed by HEC-GeoRAS for floodplain mapping. It was found that the floodplain has increased due to land use change from 1956 to 2007. This study showed that floodplain areas in irrigated farmlands have increased by 151.99% and 68.63% for return period of 25 and 50 years, respectively.

Keywords: HEC-RAS, flood, floodplain, azaran

1. Background of the Study

Floodplains are land areas adjacent to rivers and streams that are subject to recurring inundation. Owing to their continually changing nature, floodplains and other flood-prone areas need to be examined in the light of how they might affect or be affected by landuse change. There are different factors that affect flood frequency including topographic characteristics, river morphology, environmental structures, and human activities. One of the most important effects of human activities on flood occurrence is landuse change and its inconformity with the land capabilities. In this regard, the importance of floodplain mapping, which has much utilization in floodplain management, is emphasized. In particular, hazard zonation in rural areas gives orientation to planners and settlements to stabilize their activities as well as economic development (Eftekhari, 2009). Recently major development has been made in river engineering using GIS tools and techniques. As a result, several commercial software packages are available that integrates GIS capabilities in spatial analysis with hydraulic modeling. The U.S. Army Corps of Engineers Hydrologic Engineering Center River Analysis System (HEC-RAS) is a public domain software that links with the ArcGIS software through HEC-GeoRAS (USACE, 2000). Numerous researches have been done to address effects of this phenomenon in different aspects and objectives. Neal et al. (2012) have comprehensively described the physical complexity needed to model floodplains. Several other researches have been reported substantial changes to catchment runoff and floodplain due to conversion of forest to pasture or the afforestation of grassed catchments. There are many reviews of these experiments in the literature, notably those of Zhang et al. (1999), Best et al. (2003), Siriwardena et al. (2006), Brath et al. (2006), Nirupama and Simonovic (2007), Wheatera and Evansb (2009), Sriwongsitanon and Taesombat (2011), Theiling and Burant (2013) and Wagner et al. (2013). Some of these researches including Wernera et al. (2005), Cook and Merwade (2009), Merwade et al. (2008) and more recently Alexakis et al. (2014) employed GIS and remote sensing techniques for the assessment of land use change impact on flood runoff. HEC-RAS is a well know and public domain software for flood modeling and is widely used by different

researchers including Sanders (2007), Hagen et al.(2010) and Croke et al. (2014). The purpose of this study is to identify the areas located in floodplain based on two flood frequency interval. The second objective of this study is to assess the changes in floodplain boundaries due to landuse changes in 50 years between 1956 and 2007. This period was selected because the land use data and observed peak discharge are available for this period.

2. Material and Methodology

2.1 Study Area

The Azaran watershed is the unit of consideration for this research and is located northwest of Esfahan, Iran (see Fig.1). Geographically it extends from 33° 39' 57"N to 33° 44' 45"N and 50° 59' 46"E to 51° 15' 07"E while its area encompasses 9601(ha). Its elevation is range from 2030 m to 3400 m based on mean sea level. The predominant land use of Azaran is rangeland.

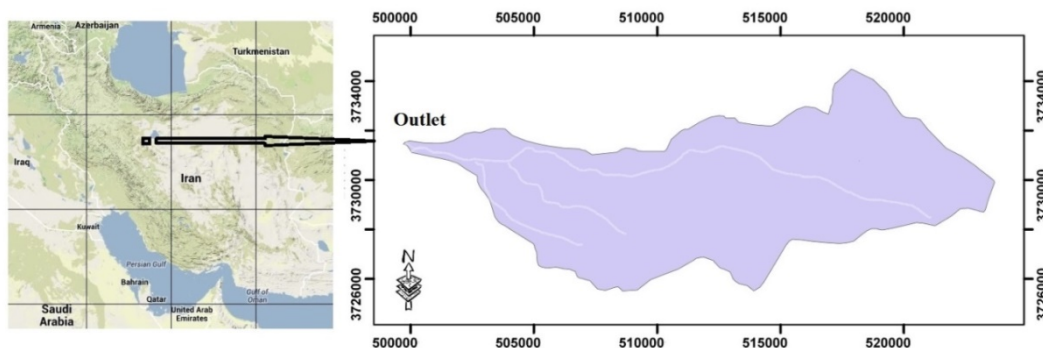


Figure 1. Layout of the study area

According to the Forest, Range and Watershed Organization of Esfahan, a landuse map of year 1956 had been generated using aerial photos at the scale of 1:55000 (Fig.2-a). A subsequent landuse map for 2007 was generated using Landsat satellite images (Fig.2-b). Roughness coefficients for the main channels were estimated according to the field survey of main rivers in the study area.

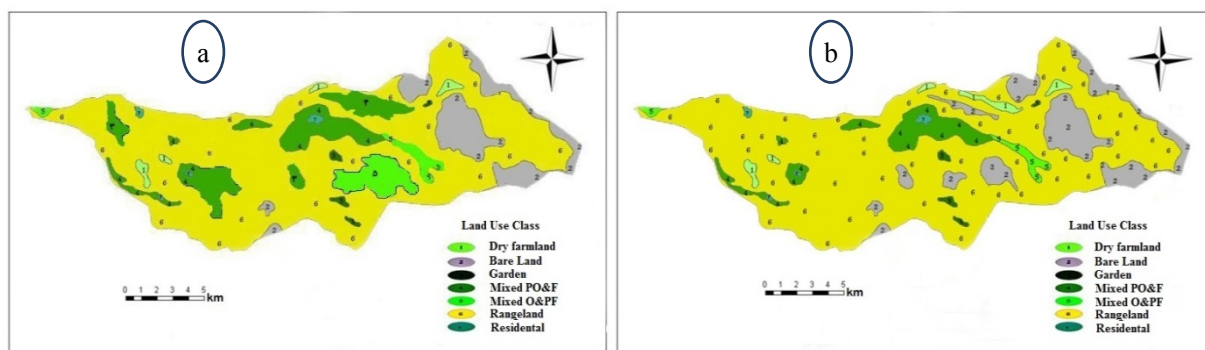


Figure 2. Representative landuse map of Azaran watershed for 1956 (a) and 2007 (b)

Table 1. Distribution landuse classes for Azaran watershed (1956 and 2007)

Landuse class	1956	2007	Landuse changes
	Area(ha)	Area(ha)	
Dry farm	289.79	339.79	17.25
Bare lands	1209.45	1359.45	12.40
Gardens *	315.97	146.11	-53.76
Mixed predominate Gardens and Farm *	799.96	351.91	-56.01
Mixed predominate Farm and Gardens*	937.91	489.85	-47.77
Fair to poor Rangelands	5988.48	6854.45	14.46
Residential area	59.44	59.44	0.00

*Irrigated Farmland.

As is evident in table 1, there was no change in residential area. This is because there is no major population in this area. There are only a few small villages in which no development has been made due to long term drought and lack of water resources during the last 15 years.

2.2 Maximum Instantaneous Flow Rate

The maximum instantaneous flow rate for 34 years dataset were analyzed to determine the maximum flow rate for various return periods using the Stormwater Management and Design Aid (SMADA) software. The Log Pearson Type III distribution depicts a minimum standard deviation in comparison to other statistical measures. Instantaneous flow rates for different return periods are provided in table 2.

Table 2. Estimated maximum flow rate based on Log Pearson Type III distribution

Non- exceedance Probability	Return Period	Q (m ³ /s)
0.5	2	17.71
0.8	5	35.43
0.9	10	57.09
0.96	25	100.40
0.98	50	147.23
0.99	100	216.54
0.995	200	358.28

2.3 Generation of Digital Terrain Model (DTM)

The first step in the pre-processing stage is to create a Digital Elevation Model (DEM) of the river system in a Triangulated Irregular Network (TIN) format. The TIN must be constructed with special care in order to provide accurate analyses. Elevation data for each cross section is extracted from the TIN (Sredojevic and Simonovic, 2009). The TIN also serves to determine floodplain boundaries and calculation of inundation depths. The DEM is a representation of the topographical surface in terms of regularly spaced x, y, z, coordinates. The DEM can be developed from a number of sources including ground survey, cartography, photogrammetry, surface sensing Lidar Data, Satellite data such as SRTM and ASTER-GDEM. The TIN-based model has a vector-based data structure, but it can be converted into grid cells. In the TIN model, each point has defined x, y, and z coordinates. The coordinate z represents the height. These points are connected by their edges to form a network of overlapping triangles (finite surfaces) that represent the terrain surface (Lo and Yeung, 2005). The basis of TIN-based DTM is that a large series of these finite surfaces, sharing common horizontal edges, can be linked together and used to interpolate the XYZ coordinate of any point, even though actual measurements have not been obtained at that point. In this research a Digital Elevation Model (DEM) was generated for the investigated river domain by using topographic map at a scale of 1:25000 collected from the National Cartographic Center of Iran, Esfahan branch.

2.3.1 Creating River Center Line

The river centerline layer is very important, because it represents the river network for HEC-RAS. Digitizing of the stream centerline starts with selecting the sketch tool from the Editor Toolbar, and digitization proceeds in the direction of river flow

2.3.2 Creating River Banks

The bank lines layer is used to define the river channel from overbank areas. This definition is important because Manning's n values are different for channel and for floodplain areas. Usually, the overbank areas have higher values of Manning's n due to vegetation or presence of residential areas.

2.4 Manning Roughness Coefficient

Manning roughness coefficients (n) represent the resistance to flood flows in channels and floodplains. The results of Manning's formula, an indirect computation of stream flow, have applications in floodplain management, flood insurance studies, and the design of bridges and highways across floodplains. Many researches have been done to determine Manning roughness coefficients. Among those, the research done by Gharib (2006), yan and zhong (2002) as well as Wong (2008) are notably comprehensive. To determine the Manning's n, field survey conducted and during the fieldwork representative pictures were taken from the left

and right overbanks and main channels. Then, the Manning coefficient values were determined by field survey and justifying the taken photos with the above mentioned references.

2.5 Pre-Processing HEC-RAS Results

The use of the unsteady flow simulation features in the U.S. Army Corps of Engineers Hydrologic Engineering Center's HEC-RAS software program requires the use of geometry pre-processor to complete the hydraulic calculations. The correct selection of geometric parameters and review of the pre-processor output is an important factor in the successful execution of an unsteady flow simulation. The geometry pre-processor in HEC-RAS calculates a series of curves that describe the geometric and hydraulic properties of cross-sections, bridges, culverts, and weirs. The HEC-RAS modeler can modify several variables that affect the calculation of these curves. GeoRAS generates three additional data sets prior to using HEC-RAS. They are the 3D Centerline theme, the 3D Cross Section theme, and the RAS GIS Import File (Snead and Maidment, 2000). Upon successful implementation of the simulation, the HEC-RAS output is exported to HEC-Geo RAS for post-processing of the output. Post-processing facilitates the automated floodplain delineation based on the data contained in the RAS output file and the original terrain TIN. Using HEC-GeoRAS functionalities, the imported HEC-RAS results are processed with the TIN of the region to generate the flood water surface extents and the flood water depth files for return periods of 25 and 50 years. The water surface profile data is used to develop a water surface TIN, and the intersection of the water surface TIN with the terrain model TIN provides flood visualization. The results can be shown in 2-D or 3-D views.

2.6 HEC-RAS Basic Concepts and Equations

HEC-RAS is an integrated software system, designed for interactive use in a multi-tasking environment and used to perform one-dimensional water surface calculations (Sredojevic and Simonovic, 2009). HEC-RAS is comprised of a graphical user interface, separate hydraulic analysis components, data storage and management capabilities, and graphing as well as reporting facilities (USACE, 2010). The computation engine of HEC-RAS is based on the solution of the one-dimensional energy equation. Energy losses are evaluated by friction (Manning's formula), contraction, and expansion. In cases where the water surface profile varies rapidly, use of the momentum equation is necessary. These cases include: mixed flow regime calculations, bridge hydraulic calculations and evaluation of profiles at river confluence. Water surface is calculated from one cross section to the next by solving the energy equation written as:

$$h_e = L\bar{S}_f + C \left| \frac{\alpha_2 V_2^2}{2g} - \frac{\alpha_1 V_1^2}{2g} \right| \quad (1)$$

Where,

V_1 and V_2 = average velocities (m²/s)

g = gravitational acceleration

α_1 and α_2 = velocity weighting coefficients (dimensionless) and

h_e = energy head loss (m).

L = reach length between the adjacent cross sections

\bar{S}_f = friction slope between the two sections and

C = expansion or contraction loss coefficient (dimensionless).

The magnitude of α depends upon the channel characteristics. Typical values of α are shown in Table 3.

Table 3. Magnitude of α after Debo and Reese (2002)

Type of channel	Value of α		
	Min.	Avg.	Max.
Regular Channel	1.10	1.15	1.20
Natural Channel	1.15	1.30	1.50
Natural Channel-flooded overbanks	1.50	1.75	2.00

3. Result and Discussion

Surface water profile was derived based on flood frequency analysis as well as geometric and hydraulic specification of river cross sections. Validation of derived profile for return periods 50-yr was made based on the identified watermarks for the flood occurred in June 27, 2007. Based on this watermark, the peak discharge is estimated to be $120\text{m}^3/\text{s}$, which is 16.6% is lower than the estimated flood by Pearson Type III distribution. Moreover, according to local inhabitant's reports, a similar flood has occurred in June 20, 1956. Therefore that flood can be considered with 50-yr return period.

Based on this classification the floodplain was generated by intersecting the floodplain in 1956 and 2007. Delineation of floodplain due to landuse changes is vital for residential and agricultural landuse. The HEC-HMS was set for estimated flood flow for two return periods and water surface profile was determined. The Manning roughness was determined based on the field survey and available references. Floodplain maps were provided for 25-yr and 50-yr (see Fig.3). A comparison was made for this type of land use and presented in Table 4. It is because of the law made by local organizations specifically insurance companies that they insure only the farm and garden lands which are out of the buffer delineated by 50-yr flood in urban areas and 25-yr in rural areas.

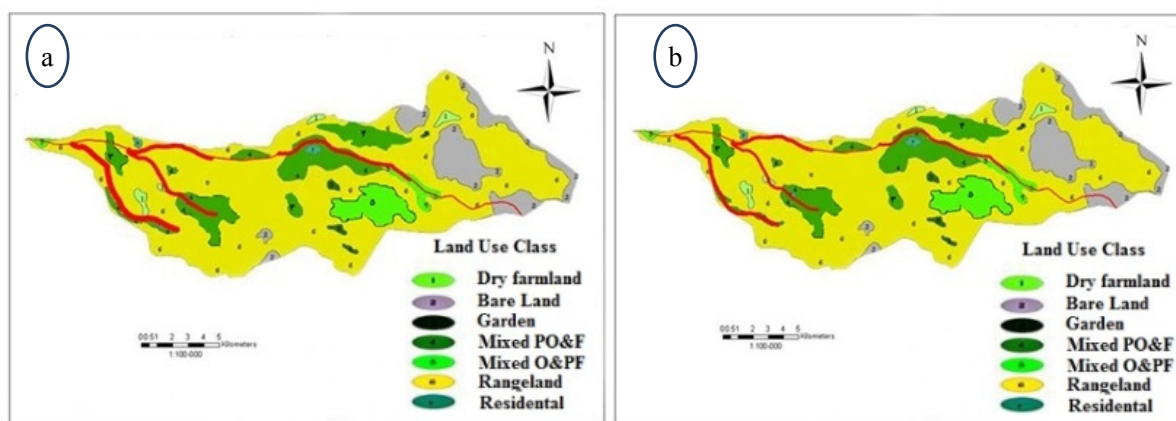


Figure 3. Floodplain area (Red color) for return period of 25-yr (a) and 50-yr (b) in Azaran watershed

Table 4. Irrigated farmlands inundated by flood with different return period

	Return period	
	25	50
Floodplain area of irrigated farmlands for land use 1956 (ha)	35.12	53.37
Floodplain area of irrigated farmlands for land use 2007 (ha)	88.50	90.00
Difference (ha)	53.38	36.63
Percentage of extent	151.99	68.63

Some relationships were found between the floodplain and return period in Azaran watershed. Figure 4 illustrates the logarithmic relationship between maximum instantaneous flood and floodplain area with strong correlation coefficient of 0.94 for irrigated farmlands in Azaran watershed. This can be due to the fact that irrigate farmlands are in a floodplain with low slope ($<10\%$). Therefore, the floodplain area will rapidly increase with small increase in flood depth. It is evident that floodplain is exponentially decreased by increasing the return period. This trend enhances the importance of topography in floodplain in irrigated farmlands of Azaran watershed.

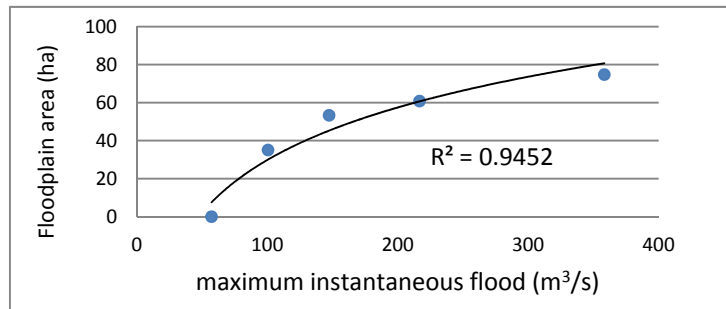


Figure 4. Relationship between maximum instantaneous flood and floodplain area in Azaran watershed

4. Conclusion

As there is no quantified data on the inundation depths for such a high magnitude flood hazard in the study region, the visualization and the quantification of the flood risks, as facilitated by this approach, can assist the decision makers to mitigate the catastrophic effects of floods. Combining the derived maximum floods by flood frequency analysis with surveyed watermarks gives a fair and reasonable confidence to the floodplain analysis and results. The Land use change analysis shows 52% decrease in irrigated farmlands. However, bare lands, dry farm lands, and rangeland have increased by 12.40%, 17.25% and 14.46% respectively. The analysis of water surface profiles shows that floodplain areas have increased in irrigated farmlands by 151.99% and 68.63% for return periods of 25-yr, and 50-yr respectively. Decreasing the floodplain by increasing the return period demonstrates the important role of topography in floodplain of Azaran's irrigated farmlands.

References

- Alexakis, D. D., Grillakis, M. G., Koutroulis, A. G., Agapiou, A., Themistocleous, K., Tsanis, I. K., ... Hadjimitsis, D. G. (2014). GIS and remote sensing techniques for the assessment of land use change impact on flood hydrology: The case study of Yialias basin in Cyprus. *Natural Hazards and Earth System Sciences*, *14*, 413–426. <http://dx.doi.org/10.5194/nhess-14-413-2014>
- Brath, A., Montanari, A., & Moretti, G. (2006). Assessing the effect on flood frequency of land use change via hydrological simulation (with uncertainty). *Journal of Hydrology*, *324*, 141–153. <http://dx.doi.org/10.1016/j.jhydrol.2005.10.001>
- Cook, A., & Merwade, V. (2009). Effect of topographic data, geometric configuration and modeling approach on flood inundation mapping. *Journal of Hydrology*, *377*, 131–142. <http://dx.doi.org/10.1016/j.jhydrol.2009.08.015>
- Croke, J., Reinfelds, I., Thompson, C., & Roper, E. (2014). Macrochannels and their significance for flood-risk minimisation: examples from southeast Queensland and New South Wales, Australia. *Stochastic Environmental Research and Risk Assessment*, *28*, 99–112. <http://dx.doi.org/10.1007/s00477-013-0722-1>
- Debo, T. N., & Reese, A. (2002). *Municipal Stormwater Management, Second Edition*.
- Eftekhari, A. (2009). The evaluation of flood risk zoning villages using HEC-GEORAS models in GIS: Case study: Gorganroud villages. *Rural Development Journal*, *1*, 157–182.
- Gharib, M. (2006). Risk Hazard Zonation and Flood management assessment. Case Study: Gharehchai River area in Rummian Suburb. *Journal of Faculty of Natural Resources*, *60*, 785–797.
- Hagen, E., Shroder, J. F., & Teuferta, J. F. (2010). Reverse engineered flood hazard mapping in Afghanistan: A parsimonious flood map model for developing countries. *Quaternary International*, *226*, 82–91. <http://dx.doi.org/10.1016/j.quaint.2009.11.021>
- Merwade, V., Cook, A., & Coonrod, J. (2008). GIS techniques for creating river terrain models for hydrodynamic modeling and flood inundation mapping. *Environmental Modelling & Software*, *23*, 1300–1311. <http://dx.doi.org/10.1016/j.envsoft.2008.03.005>
- Neal, J., Villanueva, I., Nigelwright, Willis, T., Fewtrell, T., & Bates, P. (2012). How much physical complexity is needed to model flood inundation? *HYDROLOGICAL PROCESSES*, *26*, 2264–2282. <http://dx.doi.org/10.1002/hyp.8339>
- Nirupama, N., & Simonovic, S. P. (2007). Increase of Flood Risk due to Urbanisation: A Canadian Example.

- Natural Hazards*, 40, 25–41. <http://dx.doi.org/10.1007/s11069-006-0003-0>
- Sanders, B. F. (2007). Evaluation of on-line DEMs for flood inundation modeling. *Advances in Water Resources*, 30, 1831–1843. <http://dx.doi.org/10.1016/j.advwatres.2007.02.005>
- Siriwardena, L., Finlayson, B. L., & McMahon, T. A. S. (2006). The impact of land use change on catchment hydrology in large catchments: The Comet River, Central Queensland, Australia. *Journal of Hydrology*, 326, 199–214. <http://dx.doi.org/10.1016/j.jhydrol.2005.10.030>
- Snead, D., & Maidment, D. R. (2000). *Floodplain Visualization Using HEC-GeoRAS*. Texas at Austin.: The University of Texas at Austin. Retrieved from <http://www.crwr.utexas.edu/gis/gishydro01/class/exercises/georas.html> 2013
- Sredojevic, D., & Simonovic, S. P. (2009). City of London: Vulnerability of Infrastructure to Climate Change Background.
- Sriwongsitanon, N., & Taesombat, W. (2011). Effects of land cover on runoff coefficient. *Journal of Hydrology*, 410, 226–238. <http://dx.doi.org/10.1016/j.jhydrol.2011.09.021>
- Theiling, C. H., & Burant, J. T. (2013). FLOOD INUNDATION MAPPING FOR INTEGRATED FLOODPLAIN MANAGEMENT: UPPER MISSISSIPPI RIVER SYSTEM. *River Research and Applications*, 29, 961–978.
- USACE (2010). HEC-RAS, River Analysis System, Hydraulic User Manual, Version 4.1. In CENTER, U. H. E. (Ed.). USACE.
- Wagner, P. D., Kumar, S., & Schneider, K. (2013). An assessment of land use change impacts on the water resources of the Mula and Mutha Rivers catchment upstream of Pune, India. *Hydrol. Earth Syst. Sci.*, 17, 2233–2246. <http://dx.doi.org/10.5194/hess-17-2233-2013>
- Werner, M. G. F., Hunter, N. M., & Bates, P. D. (2005). Identifiability of distributed floodplain roughness values in flood extent estimation. *Journal of Hydrology*, 314, 139–157. <http://dx.doi.org/10.1016/j.jhydrol.2005.03.012>
- Wheatera, H., & Evansb, E. (2009). Land use, water management and future flood risk. *Land Use Policy*, 26, S251–S264. <http://dx.doi.org/10.1016/j.landusepol.2009.08.019>
- Wong, T. S., & F.Asce (2008). Optimum Rainfall Interval and Manning's Roughness Coefficient for Runoff Simulation. *Journal of Hydrologic Engineering*, 13, 1097–1102. [http://dx.doi.org/10.1061/\(ASCE\)1084-0699\(2008\)13:11\(1097\)](http://dx.doi.org/10.1061/(ASCE)1084-0699(2008)13:11(1097))
- Yan, W. L., & Zhong, H. J. (2002). Composite Manning roughness coefficient of ice-covered flow. *Journal of Wuhan University of Hydraulic and Electric Engineering*.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

Detection of Cassava Leaves in Multi-Temporally Acquired Digital Images of a Cassava Field Under Different Brightness Levels by Simultaneous Binarization of the Images Based on Indices of Redness/Greenness

Mallika Srisutham^{1,2}, Ryoichi Doi¹, Anan Polthanee² & Masaru Mizoguchi¹

¹ Department of Global Agricultural Sciences, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan

² Faculty of Agriculture, Khon Kaen University, Khon Kaen, Thailand

Correspondence: Ryoichi Doi, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo 113-8657, Japan. Tel: 81-358-415-022. E-mail: roird2000@yahoo.com/roird@aeiou.pt

Received: May 17, 2014

Accepted: June 17, 2014

Online Published: August 16, 2014

doi:10.5539/mas.v8n5p87

URL: <http://dx.doi.org/10.5539/mas.v8n5p87>

The research is financed by the Japan Society for the Promotion of Science (grant-in-aid 23255014).

Abstract

Plant leaf area reveals various types of abnormalities which can enable appropriate plant/crop management actions. The quantification of plant leaf area is now feasible using commonly available digital photographing tools. Changes in brightness, however, make it difficult to compare leaf areas in digital photographs acquired at multiple time points. This difficulty could be overcome by employing an index of redness/greenness (R/G), which was suggested to be one of the best indices to discriminate between plant leaves and other objects such as soils. R/G and other indices were examined when discriminating cassava leaves from other objects in a field. A surveillance camera captured digital photographs on a daily basis. Of these, 183 photographs were stored. They were pasted into a single image file and simultaneously analyzed. The International Commission on Illumination color model's a^* was the best index in the discrimination, with a distinctiveness score of 1.36. R/G was the second best, with a distinctiveness score of 0.70. The percentage of leaf-likely pixels followed sigmoidal patterns with time, resulting in great coefficients of determination of 0.981 (a^*) and 0.965 (R/G). The percentage of leaf-likely pixels and cassava leaf weight had a real-time response relationship. The range of the 95% confidence limit was narrowed from -16 to +14% of a predicted value of 98% leaf-likely pixels for R/G to $\pm 12\%$ for a^* . Thus, the simultaneous binarization and the detection of leaf-likely pixels in the photographs acquired under different brightness levels was enabled with improved discrimination accuracy by employing a^* .

Keywords: *Manihot esculenta*, red-green-blue and $L^*a^*b^*$ color models, remote sensing, surveillance camera, temporal variation of brightness

1. Introduction

Plant leaf area indicates crop nutrient availability (Uhart & Andrade, 1995; Heege, Reusch, & Thiessen, 2008) drought stress (Alves & Setter, 2000), physical damage of canopy (Olthof, King, & Lautenschlager, 2003), land degradation/rehabilitation (Doi & Ranamukhaarachchi, 2013), and other factors. As such, the consecutive observation of leaf area is an effective measure that can be employed to find abnormalities in plant growth within a certain period. For leaf area quantification of crops and other plant species, discriminating plant leaves from other objects is essential but difficult. Some indices such as the green normalized difference vegetation index (Gitelson, Kaufman, & Merzlyak, 1996) have been developed to discriminate between leaves and other objects from earth observation satellites. Meanwhile, on the ground, Zheng, Shi, & Zhang (2010) succeeded in discriminating between vegetable leaves and soils photographed by a commonly available digital camera. However, the discrimination may be significantly hampered by changes in brightness (Pettorelli, 2013). Thus, because of temporal changes in brightness, the consecutive discrimination between leaves and other objects has

proven difficult.

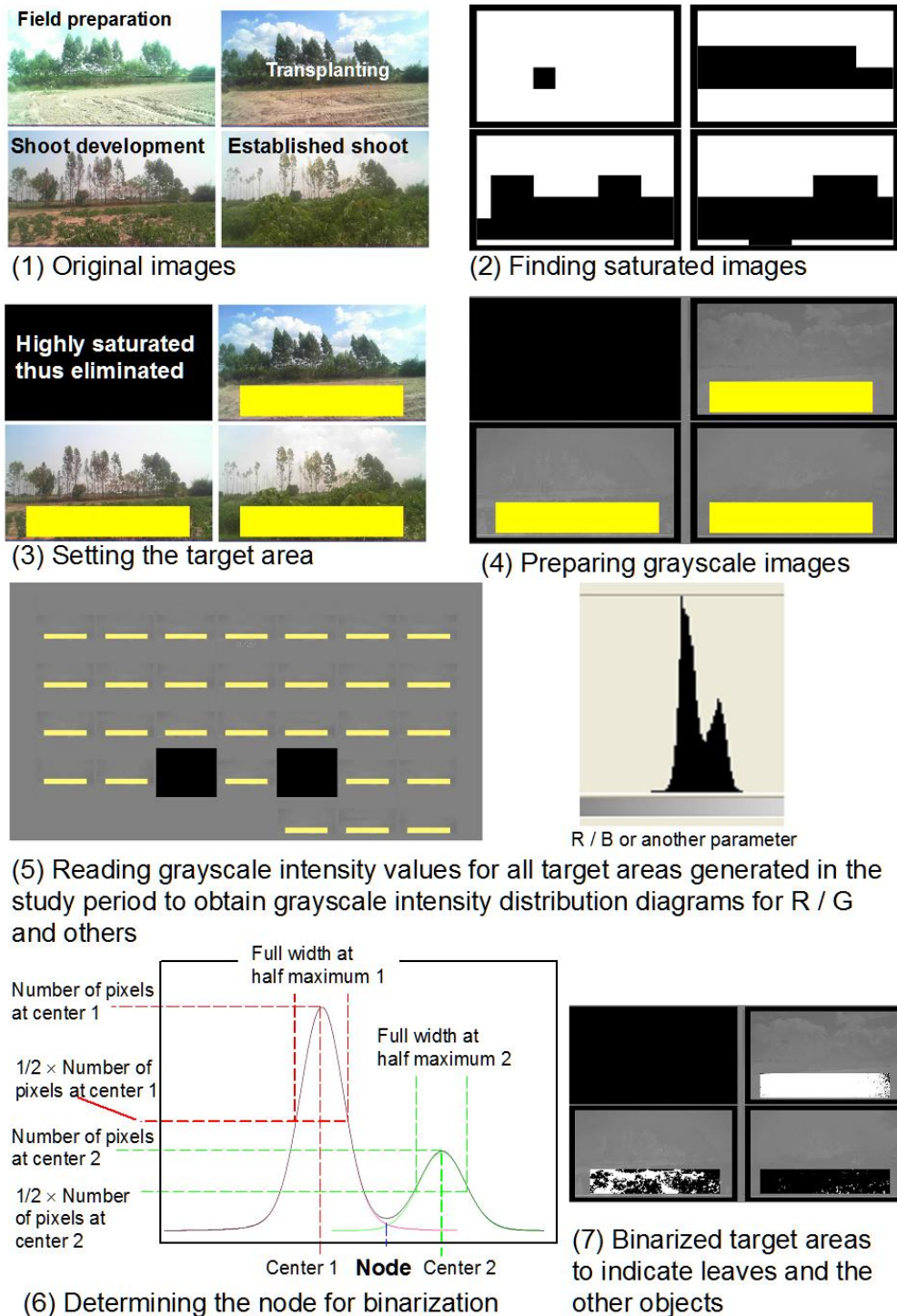


Fig. 1. Procedures of image processing for discriminating between cassava leaves and other objects in the current field in Khon Kaen province, Thailand

Among the various indices to discriminate between leaves and other objects, redness/greenness (R/G) was demonstrated to be one of the best (Yang, Willis, & Mueller, 2008). Although soil colors can vary, most plant leaves are greenish. Greenish soils and other non-leaf objects are very rare. In this context, when discriminating

between leaves and other objects, R/G must be favored. Despite this, R/G has been used very rarely in the discrimination of plant leaves and other objects (Yang, Willis, & Mueller, 2008, Doi, Arif, Setiawan, & Mizoguchi, 2013). The most likely reason for this is that data on R/G do not include those on other spectra that help in the detection/observation of areas other than agricultural fields. On the other hand, in agricultural fields, the distribution patterns of objects are simpler. Soil and plant shoots comprise the majority and the number of other objects is negligible. Hence, R/G was considered to be sufficiently useful for the discrimination between leaves and other objects in agricultural fields. Furthermore, today, data on R/G in an area is quite easily attained through digital photography and digital image processing. The index does not require complicated calculation procedures that rely on sophisticated and expensive sensors and other equipment. Therefore, the index is widely available.

The index is expected to be less affected by changes in brightness than some other indices (Pettorelli, 2013). In multi-temporally acquired digital images, redness and greenness have significant correlations with brightness. Redness, greenness, and brightness in multi-temporally acquired images change proportionally (Doi, 2012). Thus, R/G could be an index to discriminate between leaves and other objects. The proportionality must compensate for the disadvantage (Yang, Willis, & Mueller, 2008), i.e., the index must involve no infrared reflectance, which is helpful in finding leaves (Jordan, 1969). In addition, in the brightness adjustment of the green and red intensity values of the entire area of a single color digital photograph, each of the green and red grayscale images is individually brightness-adjusted and then combined to prepare a red and green intensity-adjusted color digital image (Doi, 2014). This laborious brightness adjustment of multiple digital photographs could be skipped if R/G was used because of the proportionality among R, G, and brightness. Discrimination between plant leaves and other objects could be possible through the chronological placement of color digital photographs of a target photographed area and the simultaneous binarization of all pixels in the target areas to visualize pixels that represent the leaves.

In this study, the applicability of R/G in discriminating between cassava leaves and other objects was evaluated. Cassava cuttings were planted in a field in Thailand. The field was digitally photographed daily. The digital images were pasted into a single digital image, and the best R/G value for discriminating between the leaves and other objects was statistically determined. The discriminatory power was evaluated by determining the fitness of the increase in leaf-likely pixels to a sigmoidal curve as the typical pattern of increase in cassava leaf biomass (Alves & Setter, 2004). Also, the authors examined greenness per luminosity when luminosity was given as a measure of brightness. Likewise, yellowness (redness + greenness)/luminosity was used because of its consistency as well as change in brightness (Doi, 2012; Doi, 2014). The International Commission on Illumination color model's a^* (Adobe Systems, 2002) was also investigated as another index of R/G.

2. Methods

2.1 Site Description

The field was located in Khon Kaen city, Khon Kaen province, Thailand (16° 34' 11" N, 102° 40' 78" E). In Khon Kaen, the mean annual temperature is 27°C, and the annual precipitation is 1,248 mm. The climate is classified as savanna according to Köppen (1931).

Two cultivars of cassava (*Manihot esculenta*), KU 50 and Rayong 11, were planted on 26 November 2012. Cassava cuttings, 15 to 20 cm in length, were planted at 60 cm × 100 cm spacing. No irrigation water was available, and the growth was thus dependent on precipitation and soil moisture.

2.2 Digital Photography and Handling Digital Photographs

In this study, digital photographs of a cassava field were used. The photographs were captured by a surveillance camera, UCAM-DLO130, Elecom Co., Ltd., Osaka, Japan. The digital photographs were captured daily between 12:00 and 12:30, and stored on a hard disk between 26 November 2012 and 31 May 2013. The number of pixels in the field photograph was 8712. One hundred and eighty-three digital photographs were acquired. When the photograph was captured, the data on the values of the red-green-blue (RGB) color intensity were generated. The Adobe RGB 1998 color space was chosen as one of the RGB color spaces (Adobe Systems, 2002). Some extremely bright/saturated photographs were generated possibly due to computer software glitches. These saturated photographs were eliminated by adopting the following criteria (Figure 1) using Adobe Photoshop 7.0 as one of the tools for digital image processing. A luminosity grayscale image to indicate the brightness of each pixel was obtained as described later. In the luminosity grayscale image, 1089 pixels were merged. Then, the mean brightness for the pixels was determined. A luminosity value of 200 was used as the threshold to find dark-enough merged pixels. When only a single pixel or no merged pixels were darker than the luminosity value of 200, the original photograph was regarded to be highly saturated and eliminated in the

following processes. Forty-five photographs were eliminated and the remaining 138 photographs were used.

The 138 RGB color digital photographs were pasted into a new file window of Adobe Photoshop 7.0. In the photograph of each day, in another layer overlapping the file window, a target area was set on the area that included the cassava cuttings and soil (Figure 1). The number of analyzed pixels in the target area was 1350. Using Adobe Photoshop 7.0, the grayscale images that show the intensity values of R, G, and B, and the value of a^* were prepared (Doi, 2013). The R + G (RGB yellow) grayscale image was prepared by merging the R and G grayscale images at the same weight (Doi, 2014). The luminosity grayscale image was prepared by merging the R, G, and B grayscale images at weights of 0.51, 1.00, and 0.19, respectively (Handschuh, Schwaha, & Metscher, 2010). These grayscale images were further used to prepare those that show the values of G/luminosity, (R+G)/luminosity, and R/G.

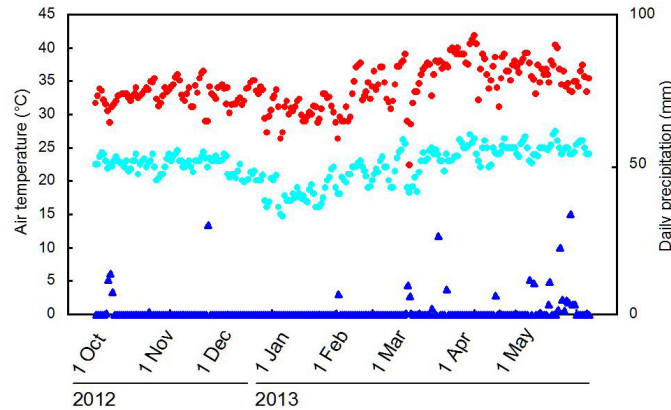


Fig. 2. Meteorological conditions of the region in the study period. The symbols indicate the maximum daily temperature (●), the minimum daily temperature (●), and daily precipitation (▲).

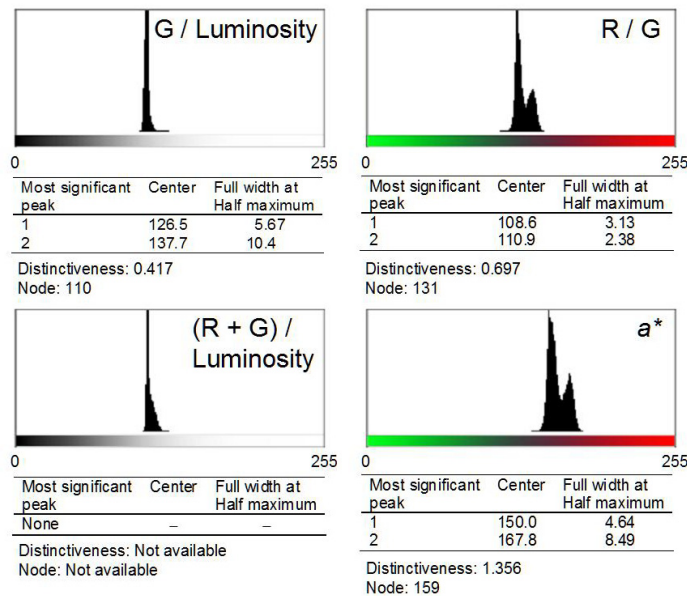


Fig. 3. Distribution patterns of the intensity values of grayscale of G/luminosity and other color indices in the target areas in the grayscale images derived from the 138 digital photographs captured in the period of this study

2.3 Statistical Performance

To find the threshold value of grayscale intensity to discriminate between leaves and other things such as soil in the target area, all 138 target areas were analyzed simultaneously (Figure 1). In the target frames in each grayscale image, the intensity values of G/luminosity, (R+G)/luminosity, R/G, or a^* were read for the 186,300 pixels (= 138 target areas \times 1350 pixels/target area). A diagram was prepared to show the distribution pattern of the grayscale intensity values for the 186,300 pixels. The intensity values were provided as digital numbers between 0 and 255. The most significant two peaks were regarded to represent the leaf and other pixels. Then, the value as the node to separate the two peaks was statistically determined using Fityk 0.8.0 (Marcin Wojdyr) as a tool for peak separation. The distinctiveness in the separation was determined using the following equation (Figure 1):

$$\text{Distinctiveness} = \frac{|\text{Center of peak 1} - \text{Center of peak 2}|}{(\text{Full width at half maximum of peak 1} + \text{Full width at half maximum of peak 2})} \quad (1)$$

The value of the node was used to generate binary images to discriminate between the leaves and other objects in the target area. After the binarization, leaf-likely pixels were shown as black pixels. The percentage of leaf-likely pixels in the target area was determined using Adobe Photoshop 7.0.

Using the statistical software SPSS 10.0.1 (SPSS Inc.), a sigmoidal relationship between the percentage of leaf-likely pixels and the number of days after transplanting the cassava cuttings was examined.

3. Results

The closest weather station of Khon Kaen reported the amount of precipitation within the study period to be 186 mm, and the air temperature ranged between 14.7 (1 January 2013) and 41.8°C (5 April 2013, Figure 2). During the period of this study, the mean air temperature was 28.4°C.

Figure 3 shows the distribution patterns of the grayscale intensity values between 0 and 255 in the 138 target frames. Values of R/G and a^* were converted to those between green (0) and red (255). According to the values of distinctiveness, among the indices, a^* most distinctively separated the two most significant peaks on the scale. R/G was the second best for separating the primary two peaks. The comparative relationships among the indices were also visually perceivable. The diagram for a^* or R/G indicates two main peaks but those for G/luminosity and (R+G)/luminosity did not show clear multiple peaks.

Given these results, the values of the node for R/G and a^* were used in the following processes. After eliminating saturated photographs, the target area was binarized by adopting the node value (Figure 3) as the threshold of the grayscale intensity (Figure 1). Pattern increases in leaf-likely pixels were examined if they followed sigmoidal patterns with time (Gray, 2000). The authors found sigmoidal patterns in the increase of leaf-likely pixels for R/G and a^* (Figure 4). The great values of the coefficient of determination ($R^2 > 0.964$) demonstrate the relevance of the leaf-likeness of the black pixels generated by the binarization (Figure 1). The fitness was better ($R^2 = 0.981$) for a^* than R/G ($R^2 = 0.965$). This relationship was also visually recognized as a smaller delineation of outliers from the predicted values for the a^* diagram (Figure 4). The lower and upper 95% confidence limits were -16% and +14% of a predicted value of 98% leaf-likely pixels in May 2013, respectively, for R/G, while the values were improved to be $\pm 12\%$ for a^* (Figure 4).

Relationships between the increases in leaf-likely pixels in the target area and cassava growth are shown in Figure 5. Leaf-likely pixels (%) in the target area did not respond quickly to the increases in cassava height (Figure 5a). On the other hand, the increase in leaf-likely pixels and that in cassava leaf fresh weight showed a real-time relationship until the target area was filled with leaf-likely pixels between 13 March and 13 April (Figure 5b).

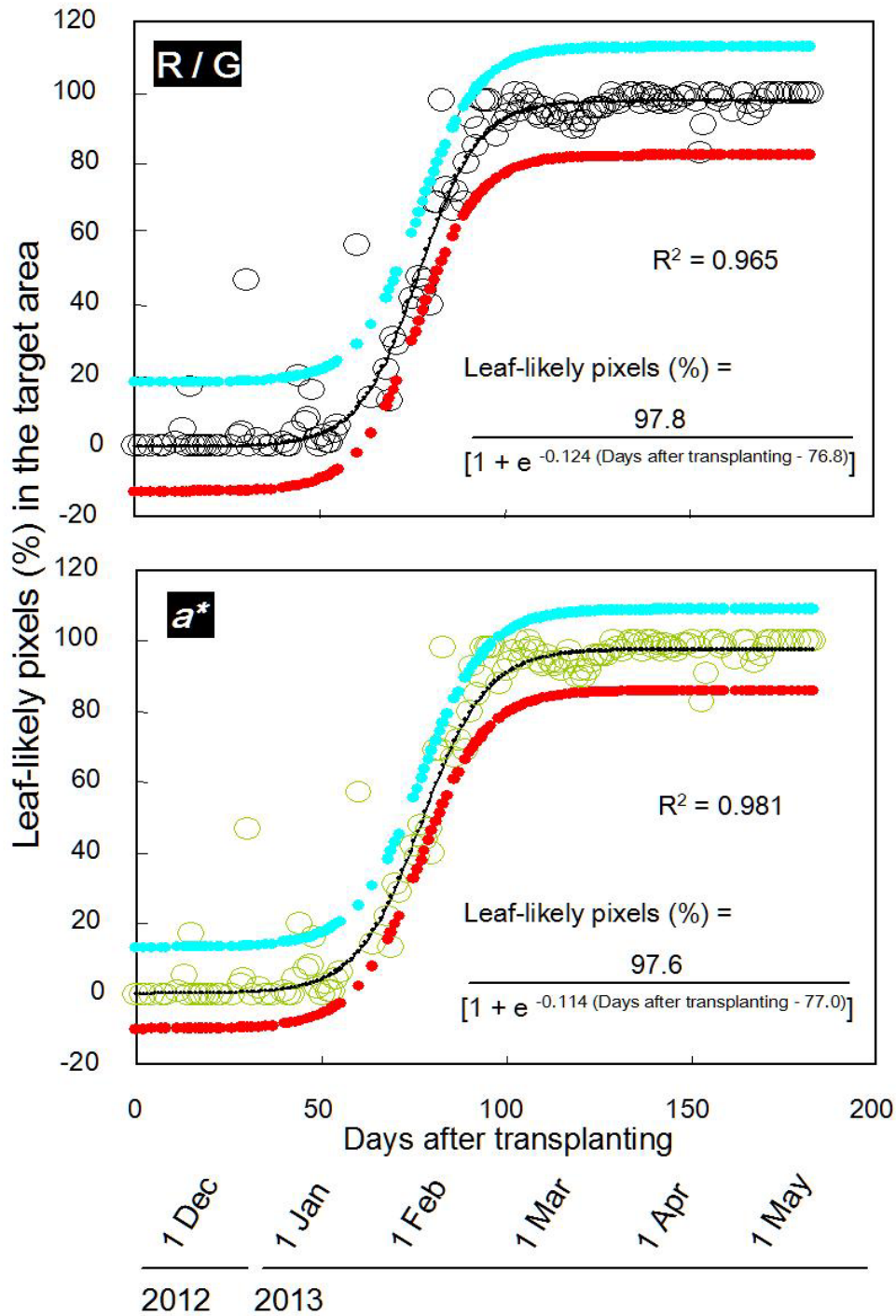


Fig. 4. Changes in leaf-likely pixels' percentage for the target area based on the R / G (top) and the a^* (bottom) grayscale images. The symbols indicate observed (\circ R / G; \circ a^*) and predicted (—) values, respectively. The light blue (\bullet) and the red (\bullet) dots indicate upper and lower 95% confidence limits, respectively.

4. Discussion

The study period was meteorologically characterized according to the aridity index (de Martonne, 1926). For the study period, the value was 4.84 ($= 186 / [28.4 + 10]$), comparable to values of the arid lands. Because of the hot and dry weather, however, the cassava seedlings grew by relying on soil moisture stored in the previous months during the rainy season (Doi & Ranamukhaarachchi, 2009a).

In this study, the relative superiority of R/G and a^* compared to others in the discrimination between the cassava leaves and other objects was shown. The superiority is based on their ranges in a color space. G/luminosity or (R+G)/luminosity describes colors between black (grayscale value = 0) and the most colorful green or RGB yellow (grayscale value = 255). On the other hand, R/G or b^* shows colors between the most colorful red and green via gray (Moriizumi, Nakashima, Okumura, & Yamanoi, 2009), at which the intensity values of red and green are the same (Figure 3) (Lakio, Heinamaki, & Yliruusi, 2010). This nature of R/G and a^* is advantageous to discriminate between greenish and reddish objects. Both the intensity values of R and G are affected by brightness and, as such, normalization is required to compare the information carried by digital photographs acquired over a period of time. However, for R/G, the effects of brightness changes in the study period were minimized because the intensity values of R and G change proportionally with that of luminosity as a parameter of brightness (Doi, 2012). Likewise, a^* , which has a close relationship with R/G, showed great consistency in discriminating between the leaves and other objects. Thus the color component a^* was found to be another feasible index for discriminating between leaves and other objects in digital photographs.

When changes in the percentage of leaf-likely pixels in the R/G and a^* grayscale images were compared (Figure 4), a significant correlation was found ($R^2 = 0.996$, $P < 0.001$). This high correlation supports the hypothesis that a^* , as an index of redness/greenness, consistently showed differences between the leaves and other objects within the conditions in this study. Half of the target area was occupied by leaf-likely pixels 77 days after transplanting (Doi & Ranamukhaarachchi, 2009b) (Figure 4). The time point was mid-February when rapid increases in leaf-likely pixels were indicated. The smooth sigmoidal curves with high R^2 values indicate that the binarized black and white pixels in the target areas were very likely to represent the cassava leaves and other objects, respectively.

The delayed increase in the percentage of leaf-likely pixels in the target areas compared with the increase in cassava height (Figure 5a) means that the cassava cuttings gained height before expanding their leaves. This time lag was possibly caused by the dryness of the upper layer of soil in which the cassava cuttings were developing roots, which later went down to much moister layers (Alves & Setter, 2004). On the other hand, the real-time response of the percentage of leaf-likely pixels to mean cassava leaf fresh weight (Figure 5) again indicates the leaf-likeness of the black pixels in the binarized target areas (Figure 1).

A profitable finding of this study was that the color component a^* , in addition to R/G, consistently indicated leaf-likely pixels by adopting the node value despite brightness changes in the study period (Figure 4). Furthermore, a^* was better than R/G in the discrimination between cassava leaves and other objects in the field (Figures 3, 4). The International Commission on Illumination's $L^*a^*b^*$ color model can describe all visible colors. Meanwhile, the color space of the current RGB color model is included in the $L^*a^*b^*$ color gamut (Adobe Systems, 2002). In the RGB color model, the darkest and the most colorful red or green colors are described by the intensity values of 0 and 255, respectively. These extreme RGB colors are still those in between 0 and 255 in the $L^*a^*b^*$ description. The most colorful RGB green ($G = 255$) has an a^* value of 49, whereas the most colorful RGB red ($R = 255$) has an a^* value of 209 when all visible colors have values of a^* between 0 and 255. This nature of the $L^*a^*b^*$ color model which describes all visible colors was thought to be the advantage that enhanced the discrimination between the cassava leaves and other objects.

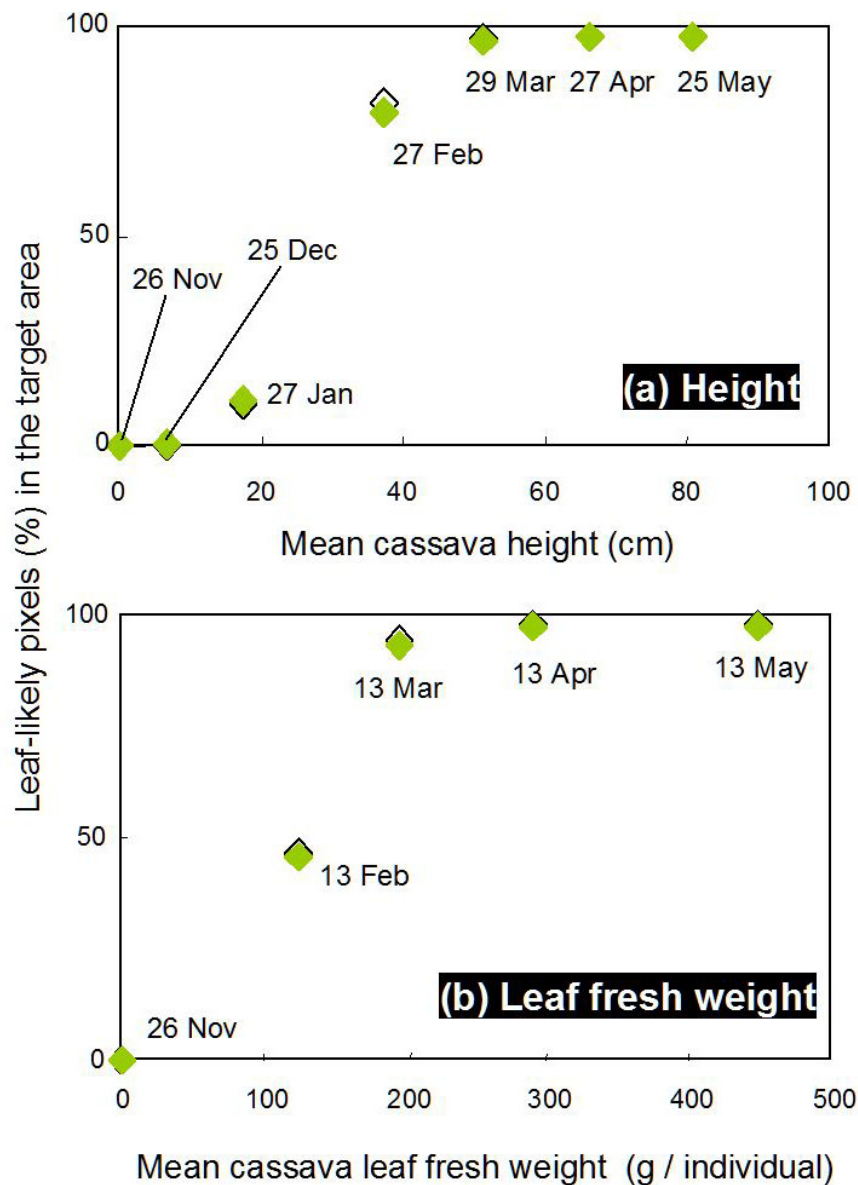


Fig. 5. Relationships between cassava height (a) and cassava leaf weight (b) and changes in leaf-likely pixels' percentage for the target area based on the R / G (◆) and the a^* (◇) grayscale images

Today, digital cameras and various types of software for digital photograph processing/analysis are quite commonly available. As such, the current method enables the discrimination between leaves and other objects by simultaneous binarization of multiple digital photographs multi-temporally acquired under different brightness levels. In this context, other related gadgets such as balloons may offer practical alternatives for the consecutive quantification of leaf area in agricultural fields and other areas (Verhoeven, 2009; Zhang & Kovacs, 2012). Thus, the use of R/G and a^* as the indices for the determination of plant leaf area is worthy of consideration.

Acknowledgments

The authors acknowledge the financial support by the Japan Society for the Promotion of Science (grant-in-aid 23255014).

References

- Adobe Systems. (2002). *Adobe Photoshop 7.0 Classroom in a Book*. Adobe Press Inc., San Jose, CA. Adobe Photoshop 7.0 Classroom in a Book.
- Alves, A. A., & Setter, T. L. (2000). Response of cassava to water deficit: Leaf area growth and abscisic acid. *Crop Sci.*, *40*, 131-137. <http://dx.doi.org/10.2135/cropsci2000.401131x>
- Alves, A. A. C., & Setter, T. L. (2004). Response of cassava leaf area expansion to water deficit: Cell proliferation, cell expansion and delayed development. *Ann. Bot.*, *94*, 605-613. <http://dx.doi.org/10.1093/aob/mch179>
- de Martonne, E. (1926). Areism and aridity index. *Comptes Rendus Hebdomadaires Des Seances De L Academie Des Sciences*, *182*, 1395-1398.
- Doi, R. (2012). Simple luminosity normalization of greenness, yellowness and redness/greenness for comparison of leaf spectral profiles in multi-temporally acquired remote sensing images. *J. Biosci.*, *37*, 723-730. <http://dx.doi.org/10.1007/s12038-012-9241-3>
- Doi, R. (2013). Discriminating crop and other canopies by overlapping binary image layers. *Opt. Eng.*, *52*, 020502. <http://dx.doi.org/10.1117/1.OE.52.2.020502>
- Doi, R. (2014, in press). Red-and-green-based pseudo-RGB color models for the comparison of digital images acquired under different brightness levels. *J Modern Opt.* <http://dx.doi.org/10.1080/09500340.2014.935506>
- Doi, R., Arif, C., Setiawan, B., & Mizoguchi, M. (2013). Quantitative use of luminosity-normalized grayscale images of greenness, redness and yellowness of a rice canopy derived from multi-temporally acquired digital photographs. *Int. J. Agric. Biol.*, *15*, 707-712. http://www.fspublishers.org/published_papers/91820_.pdf
- Doi, R., & Ranamukhaarachchi, S. L. (2009a). Community-level physiological profiling in monitoring rehabilitative effects of *Acacia auriculiformis* plantation on degraded land in Sakaerat, Thailand. *Silva Fenn.*, *43*, 739-754. <http://www.metla.fi/silvafennica/full/sf43/sf435739.pdf>
- Doi, R., & Ranamukhaarachchi, S. L. (2009b). Correlations between soil microbial and physicochemical variations in a rice paddy: Implications for assessing soil health. *J. Biosci.*, *34*, 969-976. <http://dx.doi.org/10.1007/s12038-009-0111-6>
- Doi, R., & Ranamukhaarachchi, S. L. (2013). Slow restoration of soil microbial functions in an *Acacia* plantation established on degraded land in Thailand. *Int. J. Environ. Sci. Technol.*, *10*, 623-634. <http://dx.doi.org/10.1007/s13762-012-0165-0>
- Gitelson, A. A., Kaufman, Y. J., & Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sens. Environ.*, *58*, 289-298.
- Gray, V. M. (2000). A comparison of two approaches for modelling cassava (*Manihot esculenta* Crantz.) crop growth. *Ann. Bot.*, *85*, 77-90. <http://dx.doi.org/10.1006/anbo.1999.0999>
- Handschuh, S., Schwaha, T., & Metscher, B. (2010). Showing their true colors: A practical approach to volume rendering from serial sections. *BMC Dev. Biol.*, *10*, 41. <http://dx.doi.org/10.1186/1471-213X-10-41>
- Heege, H. J., Reusch, S., & Thiessen, E. (2008). Prospects and results for optical systems for site-specific on-the-go control of nitrogen-top-dressing in Germany. *Precision Agriculture*, *9*, 115-131. <http://dx.doi.org/10.1007/s11119-008-9055-3>
- Jordan, C. F. (1969). Derivation of leaf-area index from quality of light on the forest floor. *Ecology*, 663-666. <http://dx.doi.org/10.2307/1936256>
- Köppen, W. P. (1931). Grundriss der klimakunde.
- Lakio, S., Heinamaki, J., & Yliruusi, J. (2010). Colorful drying. *Aaps Pharmscitech*, *11*, 46-53. <http://dx.doi.org/10.1208/s12249-009-9351-x>
- Moriizumi, M., Nakashima, S., Okumura, S., & Yamanoi, Y. (2009). Color-change processes of a plinian pumice and experimental constraints of color-change kinetics in air of an obsidian. *Bulletin of Volcanology*, *71*, 1-13.

<http://dx.doi.org/10.1007/s00445-008-0202-5>

Olthof, I., King, D. J., & Lautenschlager, R. (2003). Overstory and understory leaf area index as indicators of forest response to ice storm damage. *Ecol. Ind.*, 3, 49-64. [http://dx.doi.org/10.1016/S1470-160X\(03\)00010-4](http://dx.doi.org/10.1016/S1470-160X(03)00010-4)

Pettorelli, N. (2013). *The Normalized Difference Vegetation Index*, Oxford University Press.

Uhart, S. A., & Andrade, F. H. (1995). Nitrogen deficiency in maize: I. Effects on crop growth, development, dry matter partitioning, and kernel set. *Crop Sci.*, 35, 1376-1383. <http://dx.doi.org/10.2135/cropsci1995.0011183X003500050020x>

Verhoeven, G. J. J. (2009). Providing an Archaeological Bird's-eye View - an Overall Picture of Ground-based Means to Execute Low-altitude Aerial Photography (LAAP) in Archaeology. *Archaeological Prospection*, 16, 233-249. <http://dx.doi.org/10.1002/arp.354>

Yang, Z., Willis, P., & Mueller, R. (2008). Impact of band-ratio enhanced AWIFS image to crop classification accuracy. *Proc. Pecora*, 17. Retrieved from <http://www.asprs.org/a/publications/proceedings/pecora17/0041.pdf>

Zhang, C., & Kovacs, J. M. (2012). The application of small unmanned aerial systems for precision agriculture: A review. *Precision Agriculture*, 13, 693-712. <http://dx.doi.org/10.1007/s11119-012-9274-5>

Zheng, L., Shi, D., & Zhang, J. (2010). Segmentation of green vegetation of crop canopy images based on mean shift and fisher linear discriminant. *Pattern Recog. Lett.*, 31, 920-925. <http://dx.doi.org/10.1016/j.patrec.2010.01.016>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

Brain Activation and Psychophysiological Interaction in Association with a Phonological Working Memory Task

Ahmad Nazlim Yusoff¹, Hanani Abdul Manan¹, Siti Zamratol-Mai Sarah Mukari², Khairiah Abdul Hamid^{1,3} & Elizabeth A. Franz⁴

¹ School of Diagnostic and Applied Health Sciences, Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia

² School of Rehabilitation Sciences, Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia

³ School of Health Sciences, KPJ Healthcare University College, Negri Sembilan, Malaysia

⁴ Department of Psychology and fMRIotago, University of Otago, Dunedin, New Zealand

Correspondence: Ahmad Nazlim Yusoff, Diagnostic Imaging & Radiotherapy Program, School of Diagnostic and Applied Health Sciences, Faculty of Health Sciences, Universiti Kebangsaan Malaysia, Jalan Raja Muda Abdul Aziz, 50300 Kuala Lumpur, Malaysia. Tel: 603-2687-8084. E-mail: nazlimtrw@ukm.edu.my

Received: June 10, 2014

Accepted: June 16, 2014

Online Published: August 16, 2014

doi:10.5539/mas.v8n5p97

URL: <http://dx.doi.org/10.5539/mas.v8n5p97>

The research is financed by the UKM-GUP-SK-07-20-205 research grant.

Abstract

Brain activation within, and psychophysiological interaction between, significantly activated regions in the brain obtained from a phonological working memory experiment on a single participant were studied. Given that working memory and speech processing are key functions of human behaviour, this type of investigation is of fundamental importance to our understanding of brain-behaviour relationships. The study objectives were to determine the areas that respond significantly to a phonological working memory task and to investigate the influence of babble noise on their activation and the psychophysiological interactions (PPI) between the source region and those activated areas. Three conditions were used during functional magnetic resonance imaging (fMRI) scans which were working memory in quiet (WMQ), working memory in noise (WMN) and listening to babble noise (N). More voxels are activated in the right temporal lobe than in the left during N condition due to the non-speech stimulus. However, a higher mean stimulus efficacy (ϵ) of the point of maximum intensity in the left temporal lobe causes its signal intensity to be higher than in the right temporal lobe. Both the WMQ and WMN conditions resulted in similar activated regions in the brain but with a higher number of activated voxels (NOV) during WMQ for the right hemispheric areas in association with the working memory task. This is due to the sensitivity of those regions in perceiving and performing the phonological working memory task in quiet to a level that actually exceeds the activation enhancement commonly associated with the performance of working memory task in noise. This is supported by the PPI results that performing the working memory task is less influenced by noise for that particular brain region.

Keywords: functional magnetic resonance imaging, statistical parametric mapping, working memory, brain activation, psychophysiological interaction

1. Introduction

In spite of extensive research conducted on human brain processes, our knowledge about specific functions associated with storing and manipulating information (i.e., working memory and reasoning), and how background noise influences this special ability, is still in a rapid growth period (Brzezicka et al. 2011; Gathercole 1999; Karlsgodt et al. 2005; Toepper et al. 2010). Working memory is different from short term memory. The distinction between the two is that the latter does not involve the processing of a cognitive task such as mental calculation or decision making. In contrast to short-term memory, working memory can be defined as the capacity of the brain to retain information for a brief period of time for the purpose of manipulating and processing of cognitive information. A novel model on working memory was proposed by Baddeley and Hitch (1974) and has become a gold standard since then. The model comprises a three-component system consisting of the central executive, phonological loop and visuospatial sketchpad. Twenty two years after

its original development, the fourth component was introduced into the original model (Baddeley 1996) and was named the episodic buffer. While the present study focuses on specific processes of working memory of speech-related information (under the influence of background noise), the interested reader can refer to some exceptional reviews on the development of working memory models and theories as applied across a vast number of studies (Baddeley 2002; Gathercole 1999).

Known brain processes involved in working memory processing include cortical regions of the posterior parietal cortex, Broca's area, premotor and supplementary motor areas (Brodmann Area (BA) 40, 44 and 6) in the left hemisphere for phonological loop function. In the right hemisphere are the right inferior prefrontal cortex, right anterior occipital cortex, right posterior parietal cortex and right premotor cortex (BA 47, 19, 40 and 6) for the visuospatial sketchpad. Bilateral prefrontal cortex (BA 9, 10, 44, 45 and 46) is also known to be associated with central executive function (Gathercole 1999).

One interesting sub-cortical region involved in working memory processing is the hippocampus, a paired structure located inside the medial temporal lobe beneath the cortical surface. The hippocampus has been found to mediate working memory processing such as encoding the change in locations (Toepper et al. 2010), maintenance and retrieval (Karlsgodt et al. 2005) in addition to its known role in spatial learning and long term memory. In an fMRI study using a Corsi Block-Tapping test (CBT), it has also been found (Toepper et al. 2010) that other cortical and subcortical areas were activated during working memory which included the superior frontal gyrus, precentral gyrus, superior parietal lobule, inferior parietal lobule, middle occipital gyrus, inferior occipital gyrus, insula, putamen, caudate and cerebellum.

The present study, which investigates the effects of babble noise on working memory processing of speech stimuli is based on the foundation that working memory plays an important role in maintaining, manipulating, processing and storing information during cognitive processes. Brain activation that results during the execution of a task will peak normally between 5 to 10 s after stimulus delivery (Amaro & Barker 2006; Belin et al. 1999; Bernal & Altman 2001; Hall et al. 1999; Hall et al. 2000) with the dispersion of the peak around 2 to 3 s. In order for this brief response of the brain to be captured without being interfered by the scanner noise, a short-acquisition time (TA) of 2 s with a long-repetition time (TR) of 16 s was applied in a sparse temporal sampling functional magnetic resonance imaging (STS-fMRI) protocol (Gaab et al. 2003). The data were analysed using Statistical Parametric Mapping (SPM8) from which the spatial and height extent of activation in the brain were compared and contrasted between the execution of working memory task in quiet and the same task in a noisy background. Another objective of this work is to investigate and measure the psychophysiological interaction (PPI) between the input area and the areas that are activated when the participant performed the working memory task. PPI is defined as the responses in an activated cortical area in terms of an interaction between the influence of another activated area and an experimental factor (Friston et al. 1997). In this work, the experimental factor is the babble noise itself. What this study aimed to discover is whether babble noise has some measurable effects on the slope of a regression line when the response in one region is regressed on the activity in the second region.

This study was conducted on a single male participant. Thus, the inferences made about the brain activation (data) evoked by the execution of working memory task in different environments are only valid for this particular participant with a variance given by σ_w^2/n . The subscript w denotes within participant and n is the number of measurements. The error variance was estimated on a scan-by-scan basis with the assumption that each scan represents an independent observation with no serial correlation. Assessing brain activations obtained from a single participant has the advantage that they exhibit the true effects of that particular participant at a relatively high threshold and the effects observed are not influenced by between participants variability which sometimes can be cascaded down into the average effects of group activation. On the other hand, inferences about the population are not possible given there is no test of reproducibility of the data. However, this study will serve as a baseline for future multiple participant studies in a similar context in the attempt to gain a better understanding of the brain functional and structural cortical connections during the execution of working memory and how such processes are mediated by experimental factors such as noise.

2. Methods

2.1 Participant

A phonological working memory experiment was conducted on a single right-handed Malay male participant using functional magnetic resonance imaging (fMRI). The participant is a 27-year old native Malay male speaker with no reported history of psychiatric or neurological disorder and no current use of any psychoactive medications. The participant had no hearing impairment and no history of long time exposure to loud noise to be

inappropriate for auditory stimulus presentation. The participant's hearing levels for both ears are not greater than 30 dB (HL) in the frequency range of 250 – 8,000 Hz. The participant was confirmed to be right-handed with laterality index of 82.14 (in the range of 5th right). The participant agreed to participate by filling up the informed consent and screening forms and signing them, after full explanation of the nature and risks of the research, as required by the Institutional Ethics Committee (IEC) (Reference no: FF-205-2006). The participant was then interviewed on his health condition prior to the scanning session and tested for middle ear conditions (Tympanometer Model Grason Stadler Inc. GSI33) and hearing level (Pure tone audiometer Model Grason Stadler Inc. GSI61) in the frequency range of 250 Hz to 8,000 Hz, by a qualified Audiologist. Prior to the fMRI scans, the participant's handedness was tested using the Edinburgh Handedness Inventory (Oldfield 1971). The participant was told not to move his head during the scan and was instructed to attend and respond to the given stimulus accordingly.

2.2 Working Memory Paradigm

A sparse temporal sampling functional magnetic resonance imaging (STS - fMRI) was used in this study due to its relatively high sensitivity in detecting brain activation and in order to avoid the interference of the scanner sound with the auditory paradigm (Mueller et al. 2011). There were altogether 122 functional volumes in the whole imaging session. The first two volumes were dummies and the images were not acquired. Each functional volume consisted of 32 axial slices that were acquired in 2-s acquisition time (TA) (one image slice in 62.5 ms) with an inter-scan interval (TR) of 16 s. The imaging time for the whole functional scan is 1,920 s (32 minutes) which produces $120 \times 32 = 3,840$ images in total.

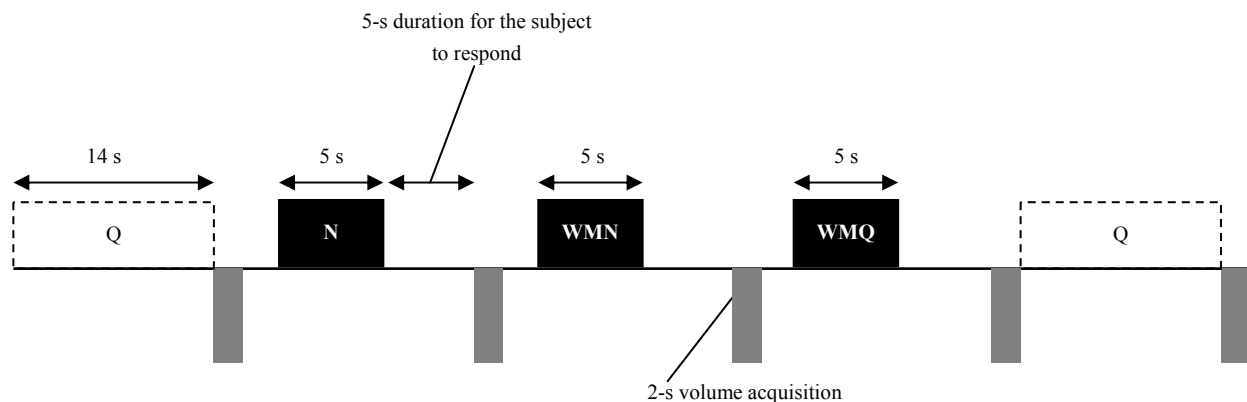


Figure 1. Schematic illustration of a cycle of the sparse temporal sampling used in this study; Q = quiet condition, N = babble noise condition, WMN = working memory performed with the presence of babble noise, and WMQ = working memory performed in quiet

The working memory stimulus was presented for 5 s during the 14-s gap of scanner silence, at the 4th second after each successive acquisition, using a digital playback system. Fig. 1 shows a schematic illustration of stimulus presentation during the STS. The stimulus consisted of a series of 5 Malay nouns (e.g. buku, kayu, kain, bola, meja) which were verbally presented at random. The participant was instructed to listen to the given words, memorise and then verbally 'rephrase' all the words backwardly immediately after hearing the last word. The participant was given 5 s to respond to the given stimuli and was requested to remain focused throughout the entire scanning session. Three conditions were used in this study; working memory in quiet (WMQ), working memory in babble noise (WMN) and listening to babble noise (N). In WMQ, the participant was required to rephrase backward all the words that he heard without being distracted. Similar stimulus presentation was used for WMN but the words were embedded in babble noise (Stimulus-to-noise ratio: SNR = 5dB) as distractor. The babble noise was recorded in a session where five multi-talkers read the same paragraph of text in an unsynchronised tempo. A total of 20 trials for each of the three N, WMN and WMQ conditions were alternately presented to the participant. A number of 20 functional image volumes were also acquired during which the participant listened only to the noise (N). Additionally, 60 volume acquisitions during the quiet condition (Q) were also conducted. All stimuli were presented binaurally at the intensity level of 70 dB via a headphone which was connected to a digital playback system. The stimuli were transmitted to both ears via air tubes.

2.3 fMRI Scans

Functional magnetic resonance imaging (fMRI) scans were performed using a 1.5 tesla magnetic resonance imaging (MRI) system (Siemens Magnetom Avanto) equipped with functional imaging option, echo planar imaging capabilities and radiofrequency (RF) head coil used for signal transmission and reception. Gradient Echo – Echo Planar Imaging (GRE-EPI) pulse sequence with the following parameters were applied to produce high resolution T2*-weighted images: Repetition time (TR) = 16,000 ms, acquisition time (TA) = 2,000 ms, echo time (TE) = 50 ms, field of view (FOV) = 192 × 192 mm, flip angle (α) = 90°, voxel size = 3 mm³ isotropic, slice thickness = 3 mm. In addition, anatomical images of the entire brain were obtained using a T1-weighted multiplanar reconstruction (MPR) pulse sequence with the following parameters: TR = 1620 ms, FOV = 250 × 250 mm, α = 90°, voxel size = 3 mm³ isotropic and slice thickness = 1 mm.

2.4 Behavioral Study

The participant's performance in both WMQ and WMN conditions was scored as the percentage of the series of words that have been correctly rephrased. A qualitative interpretation on the participant's performance during WMQ and WMN is made in relation to the fMRI results.

2.5 Image Processing

The first two scans were discarded to eliminate the effects of magnetic saturation. The fMRI data were analysed using MATLAB 7.4 – R2008a (Mathworks Inc. MA, USA) and Statistical Parametric Mapping (SPM8) (Functional Imaging Laboratory, Wellcome Department of Imaging Neuroscience, Institute of Neurology, University College of London) programming software. The participant's structural T1 and functional T2 weighted images were normalised based on the standard templates produced by the Montreal Neurological Institute (MNI). Conventional fixed-effects (FFX) and conjunction analyses based on the general linear model (GLM) were used to generate brain activation in the regions of interest using the *T*-statistic for each voxel at α = 0.05. Due to the spatial correlation that is cascaded down from the smoothed functional images, the brain activations obtained from GLM were inferred at corrected significant level (α_{FWE}) of 0.05 using random field theory (Friston et al. 1996). The fact that this study deals with volume or family of voxel statistics, the related risk of error is known as the family-wise error (FWE) rate which is defined as the likelihood that a particular family of voxel could have arisen by chance if the null hypothesis of no effects (activation) is rejected. The cortical regions which are found to be significantly activated during the phonological working memory task are defined using the Anatomy toolbox (Eickhoff et al. 2005) at α_{FWE} = 0.05, corrected for multiple comparisons.

The hemodynamic properties of the area of interest were also estimated using SPM to determine the mean efficacy of the stimulus (ϵ) in generating BOLD responses (Friston et al. 2000). This is achieved by extracting the hemodynamic parameters from a 4-mm radius sphere centered at the point of maximum intensity of each respective area. Spearman's correlation analysis was then conducted to determine the existence of any relationship between the effect size (*t* value) and the mean stimulus efficacy at α = 0.05 (95% CI) for each particular area of interest.

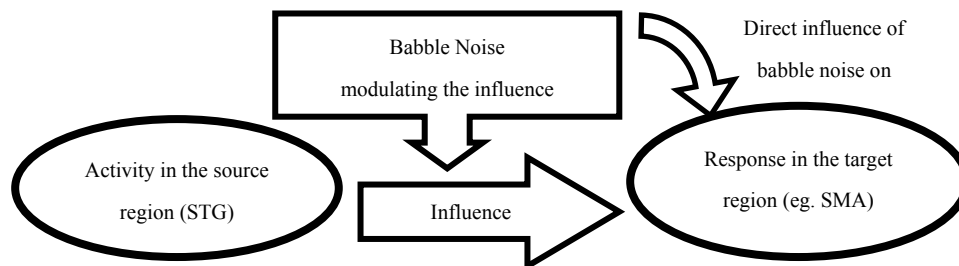


Figure 2. PPI model used in psychophysiological interaction analysis

Psychophysiological interaction (PPI) was used in this study to investigate the responses in one region of interest (ROI), or source region, in terms of an interaction between the influence of another ROI, target regions, and the babble noise (Friston et al. 1997). A 4-mm radius sphere with the point of maximum intensity as the centre was drawn for each of the source and target regions. The PPI responses for the target regions were then plotted as a function of the activity in the source region, assumed to be interacting with an experimental factor (babble noise), which in turn influences the response in the target area. The psychophysiological interaction can be understood in

terms of looking for a significant difference in the regression slope of the activity in the source region on response in a target region with and without the interaction with the experimental factor. Illustration of the PPI model used in this study is shown in Fig. 2 and will be explained in the Discussion section.

3. Results

3.1 Behavioural Results

The results obtained from the behavioural data indicated that 90% (18/20) of the participant's responses were correct in the noisy background as compared to 85% (17/20) in quiet. However, the significance of the difference was not able to be determined from these single-participant data.

3.2 Brain Activation Characteristics

Fig. 3 shows the brain areas that are activated when the participant was listening to babble noise. Two activation clusters survived the spatial threshold of 5 voxels and the height threshold of $t = 5.21$ ($p_{FWE} < 0.05$ for each voxel, corrected for multiple comparisons). Three local maxima more than 8.0 mm apart were observed in each cluster. The coordinates of the first two maxima in one hemisphere are the mirror image of the ones in the other hemisphere. However, the x and z coordinates for the third maxima in the two hemispheres are almost the same but differ largely in the y direction. Some important data characterising the brain activation in Fig. 3 are tabulated in Table 1.

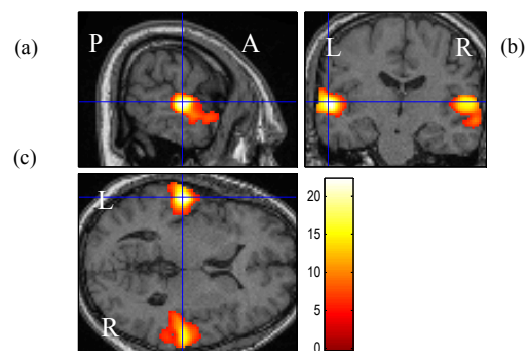


Figure 3. Brain activation resulting from listening to babble noise, overlaid onto the subject's (a) sagittal, (b) coronal and (c) axial normalised T1-weighted images (P = Posterior, A = Anterior, L = Left, R = Right). The t -values for the activated voxels are scaled in colors (red to white) as shown

Table 1. Characteristics of brain activation shown in Fig. 3

Cluster	Local maximum	Region	K_E	x	y	z	t -value	p_{FWE}	ϵ/s^{-1}
1	1	Left STG	1,489	-58	-22	6	22.21	< 0.05*	3.65
	2	Left MTG		-64	-14	-6	12.97	< 0.05*	1.32
	3	Left MTG		-68	-34	12	12.35	< 0.05*	1.26
2	1	Right STG	1,759	62	-22	6	15.80	< 0.05*	2.95
	2	Right STG		68	-14	-4	12.30	< 0.05*	1.84
	3	Right STG		64	-6	-8	10.26	< 0.05*	1.16

*Significant at set, cluster and voxel level (Friston et al. 1996).

Listening to babble noise or 'N' task resulted in widespread activation in the middle temporal gyrus (MTG) and superior temporal gyrus (STG) in the left and right temporal lobes. The activation appears to be spatially symmetrical but the number of activated voxels (K_E) is higher in the right temporal lobe as compared to the left temporal lobe. The height extent of activation observed at the coordinates of maximum intensity in each cluster, however, was determined to be higher in the left temporal lobe (Table 1). This characteristic of activation in the left temporal lobe is accentuated in the t -statistics for the three maxima and the more intense activation pattern as shown in Table 1 and Fig. 3 respectively. The stimulus efficacy (ϵ) was also determined to be higher for the

maximum intensity voxel in the left hemisphere (Table 1) as compared to the maximum intensity voxel in the right hemisphere.

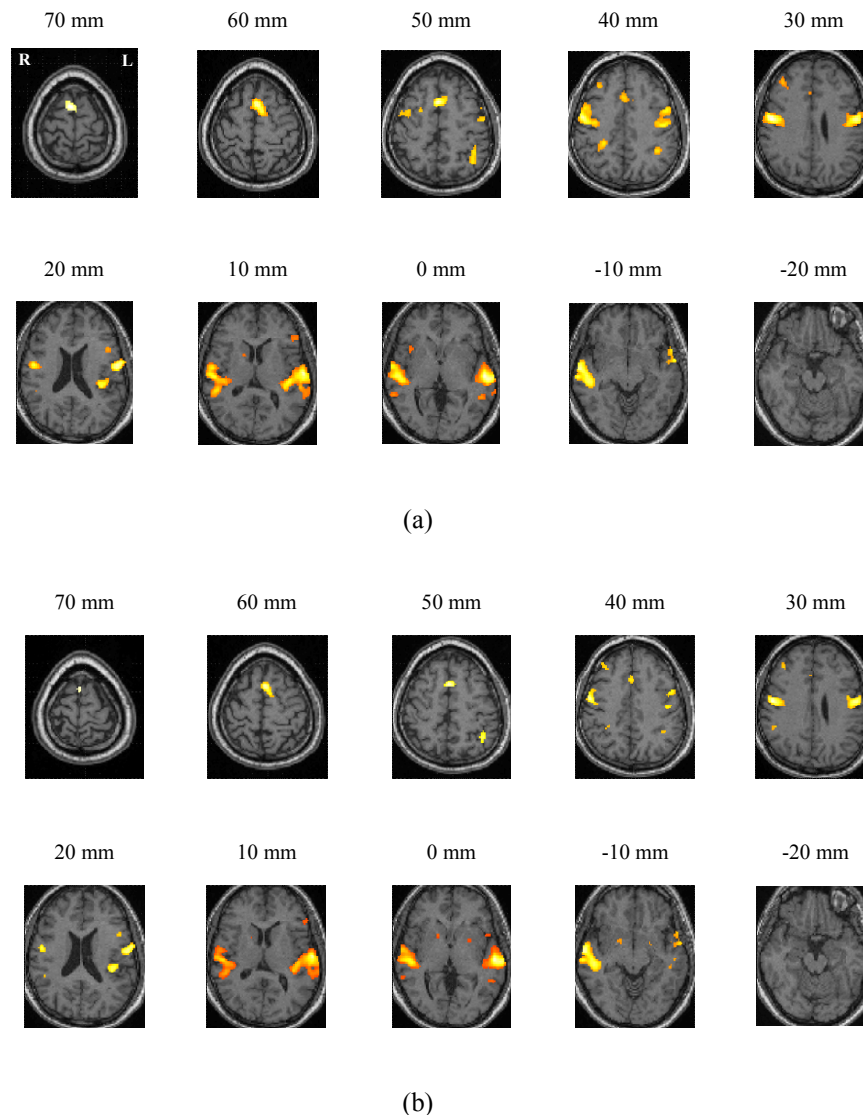


Figure 4. Activated brain regions when the subject performs (a) working memory task in quiet (WMQ) and (b) in noise (WMN); (R = Right, L = Left). The statistical images (colored) were overlaid decendingly onto the participant's axial ($z = 70$ to $z = -20$ mm) normalised T1-weighted images. The scale for the t -value is similar as in Fig. 3.

Figs. 4(a) and (b) show, respectively, the brain activations that have been captured during the execution of working memory in quiet (WMQ) and in the presence of babble noise (WMN). The images are presented neurologically from which one should note that the left-hand side of the images represent the left side of the brain. The images were obtained after thresholding the height extent of activation at $t = 5.21$ ($p_{FWE} < 0.05$ for each voxel, corrected for multiple comparisons). The spatial extent threshold was not used on these images to avoid losing important clusters of voxels that might be involved in, or associated with, execution of the working memory task. By comparing Figs. 4(a) and (b), it can be said that WMQ and WMN has resulted in similar activated clusters of voxels or regions in the brain. The spatial extent of activation for the clusters is about the same from the 20-mm level down to -20 -mm level. However, the activation from the 20-mm level to 70-mm level is relatively wider (containing more activated voxels) during WMQ as compared to WMN. A detailed analysis result for WMQ and WMN is given in Table 2 and described in the following paragraphs.

Table 2. Characteristics of brain activation for (a) WMQ and (b) WMN as shown in Fig. 4. BA = Brodmann Area; TE1.1 = primary auditory area; hIP1 = inferior parietal sulcus; IPC = Inferior parietal cortex; IFG = Inferior frontal gyrus.

Table 2 (a).

Cluster	Local maximum	Region	K_E	x	y	z	t -value	$p_{FWE-corr.}$	ϵ/s^{-1}
1	1	Left MTG	3,521	-56	-20	4	14.31	< 0.05*	3.33
	2	Left BA3a		-54	-10	24	13.33	< 0.05*	2.06
	3	Left TE1.1		-36	-36	16	11.51	< 0.05*	1.67
2	1	Right BA3b	3,423	56	-10	32	12.12	< 0.05*	2.48
	2	Right STG		58	-22	4	11.46	< 0.05*	3.59
	3	Right STG		64	-10	-4	11.35	< 0.05*	1.61
3	1	Right SMA	1,025	2	8	58	11.86	< 0.05*	2.49
	2	Right BA6		6	12	52	9.85	< 0.05*	1.84
	3	Left BA6		-4	0	62	9.11	< 0.05*	2.00
4	1	Right MTG	176	40	34	34	7.80	< 0.05*	2.14
5	1	hIP1	226	-38	-50	44	7.69	< 0.05*	1.44
	2	IPC (PFm)		-38	-62	48	6.52	< 0.05*	1.50
	3	BA2		-42	-42	48	6.01	< 0.05*	0.87
6	1	-	111	32	-42	40	7.30	< 0.05*	1.02
7	1	Left IFG	73	-42	14	22	6.93	< 0.05*	0.69
8	1	Right Putamen	37	18	8	6	6.91	< 0.05*	0.90
9	1	Left MTG	69	-52	-50	0	6.76	< 0.05*	0.80
10	1	Left IFG	52	-46	28	8	6.13	< 0.05*	0.68
11	1	-	21	28	2	48	6.08	< 0.05*	0.77

*Significant at set, cluster and voxel level (Friston et al. 1996).

Table 2 (b).

Cluster	Local maximum	Region	K_E	x	y	z	t -value	$p_{FWE-corr.}$	ϵ/s^{-1}
1	1	Left STG	2,868	-58	-18	6	16.49	< 0.05*	5.33
	2	Left BA3a		-54	-10	26	10.95	< 0.05*	3.58
	3	Left TE1.1		-36	-36	14	10.88	< 0.05*	3.13
2	1	Right STG	2,510	54	-24	2	12.40	< 0.05*	3.48
	2	Right STG		64	-10	-4	11.58	< 0.05*	3.70
	3	Right BA6		58	-6	36	9.78	< 0.05*	2.58
3	1	Right SMA	516	2	8	58	9.87	< 0.05*	2.31
	2	Left BA6		-4	0	62	6.91	< 0.05*	2.00
	3	Right SMA		14	8	76	6.89	< 0.05*	1.75
4	1	Right MTG	135	40	34	36	7.26	< 0.05*	2.51
5	1	IPC (PFm)	64	52	-44	26	6.97	< 0.05*	1.11
6	1	Right Putamen	92	18	10	4	6.97	< 0.05*	0.78
	2	-		16	0	-6	5.92	< 0.05*	0.73
	3	Right Amygdala		22	2	-12	5.61	< 0.05*	1.16
7	1	Right IFG	36	-44	12	22	6.61	< 0.05*	1.27
8	1	hIP1	132	-40	-52	46	6.54	< 0.05*	2.03
9	1	Left MTG	33	-52	-50	0	6.22	< 0.05*	1.43
10	1	BA45	27	-50	28	8	6.16	< 0.05*	1.17
11	1	-	62	-22	-2	-6	6.11	< 0.05*	0.92

*Significant at set, cluster and voxel level (Friston et al. 1996).

There are altogether 18 activated clusters of voxels for WMQ and 15 for WMN that survived the height threshold of $t = 5.21$ ($p_{FWE} < 0.05$ for each voxel, corrected for multiple comparisons) from which the respective anatomical regions for 11 highly activated clusters are shown in Table 2 (a) and (b). By comparing and contrasting WMQ and WMN, the first four clusters of voxels with the highest t -values appear to be similar. However, a marked increase in the number of activated voxels can be seen for WMQ in contrast to WMN. The coordinates of the maxima of first two clusters for both WMQ and WMN are quite similar to one another. These two clusters exhibited widespread activation in the temporal and frontal regions. In addition, a relatively intense activation is also observed in the supplementary motor area (SMA) and the number of activated voxels for WMQ in this region is almost twice the number of activated voxels for WMN. The rest of the brain areas that

are activated with lower t -value and less number of activated voxels are shown in Table 2.

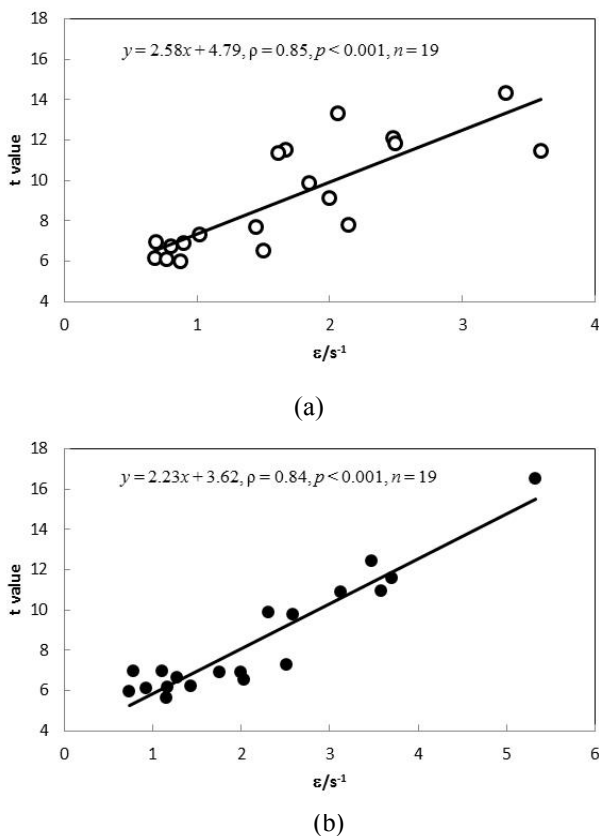


Figure 5. Linear correlation between the t value and the mean stimulus efficacy for a) WMQ and b) WMN conditions

Figure 5(a) and (b) illustrate the relationship between the effect size (t value) and the mean stimulus efficacy (ϵ) calculated for the respective areas as tabulated in Table 2 for both WMQ and WMN conditions. There exists a linear, significant ($p < 0.001$), positive relationship between the t value and the mean stimulus efficacy for both WMQ and WMN conditions with a steeper slope (m) for WMQ ($m = 2.58$ s) as compared to WMN ($m = 2.23$ s). The Spearman's correlation coefficient for WMQ ($\rho = 0.85$) is, however, not different from that of WMN ($\rho = 0.84$).

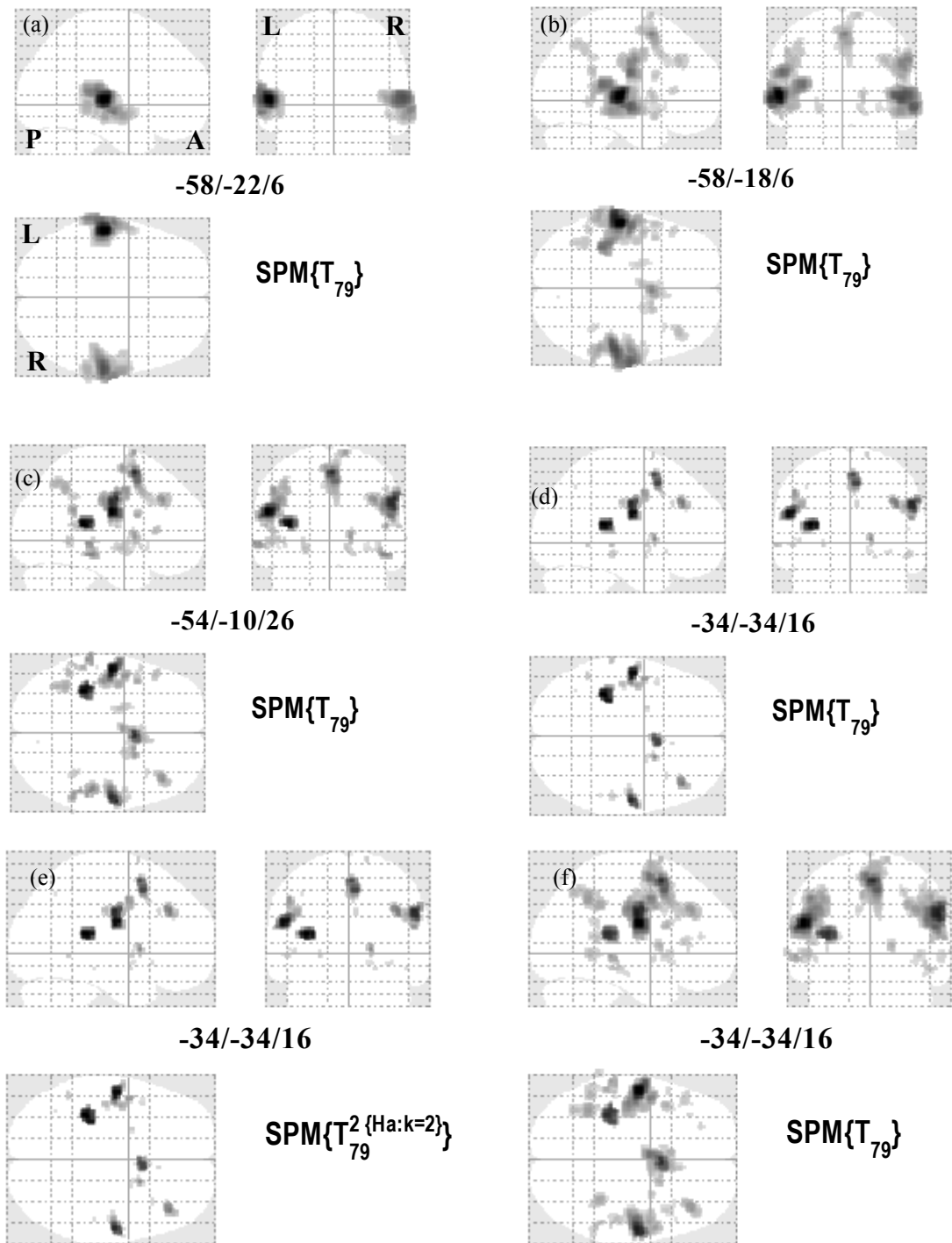


Figure 6. The MIP of the brain activation for the effects of (a) N, (b) WMN, (c) WMN - N, (d) WMN > N, (e) conjunction of c and d and (f) WMQ - N. (P = Posterior, A = Anterior, L = Left, R = Right)

Another way of visualising brain activation is by using maximum intensity projection (MIP) or glass images. Figs. 6(a and b) show the MIPs for N and WMN conditions. Note that while still being in the same region, the coordinates of maximum intensity have shifted from -58/-22/6 to -58/-18/6. To obtain the effects due to working memory; thus two MIPs were used to subtract Fig. 6(a) from Fig. 6(b) (WMN minus N). The results are shown in Fig. 6(c). Comparing Figs. 6(b) and 6(c), one notes that the effects of babble noise have been subtracted from Fig. 6(b) and the coordinates of maximum intensity, previously at -58/-22/6 have shifted to -54/-10/26. What is

left in Fig. 6(c) is assumed to be the true effects of the working memory task in noise (to be compared with Fig. 6(b)). Fig. 6(c) can be compared with Fig. 6(d) which shows the results of the contrast WMN>N. The contrast WMN>N is used to search for voxels that are more active in the WMN condition than in the N condition at a certain significance level. It can be clearly seen that Fig. 6(d) is comparable to Fig. 6(c) in terms of the most prominent clusters but with less background noise in the activation. One can see that the coordinates of maximum intensity between Figs. 6(c) and (d) were also different.

Fig. 6(e) is the conjunction of the effects of Figs. 6 (c) and (d). Conjunction analysis was performed to confirm the activation areas that solely belong to the true effects of working memory in noise. No significant difference was found between Fig. 6(d) and (e). Fig. 6(e) therefore represents the spatial extent of activation (or the true effects) in the brain when the participant performs the working memory task in the noisy (babble noise) condition. Fig. 6(f) depicts results of the working memory in quiet (WMQ) minus N. Comparing Fig. 6(e) and Fig. 6(f) and contrasting Fig. 6(c) and Fig. 6(f) it appears that all areas involved in the working memory task in noise are also involved in the same task in quiet; moreover, the working memory task in quiet activates a wider extent of brain areas compared to the working memory performed in a noisy background.

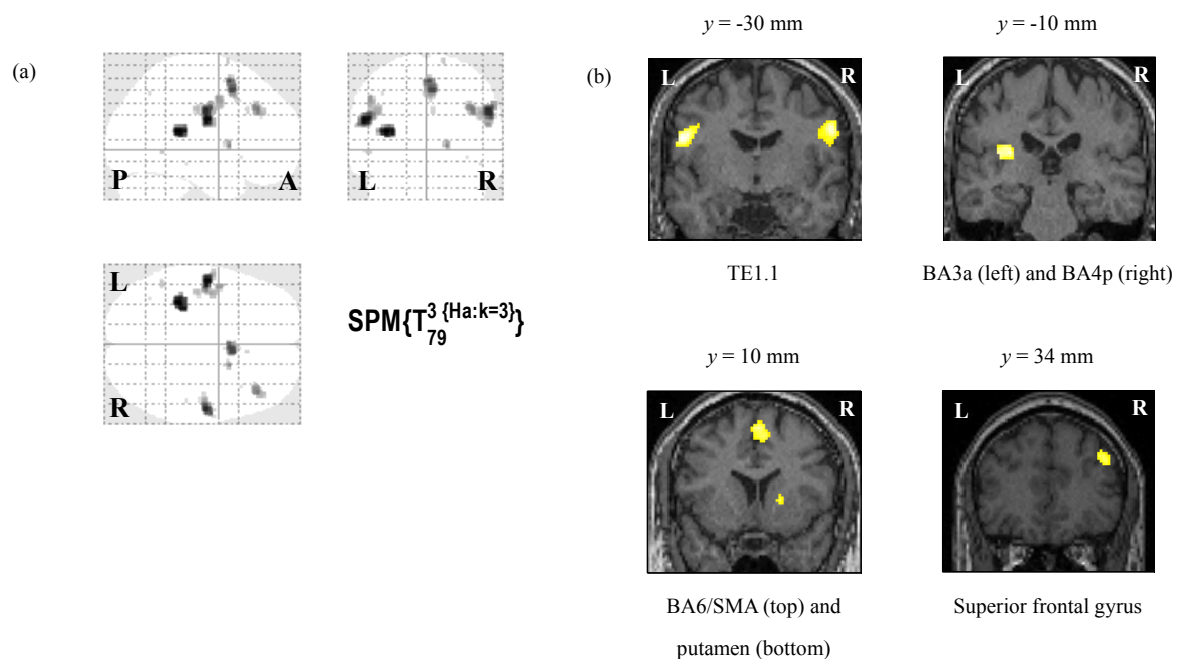


Figure 7. The MIP for the conjunction of the effects of WMQ, WMN and MN>N minus N and (b) Participant's coronal T1-weighted images (a) at four y positions showing the region associated with working memory. The scale for the *t*-value is similar as in Fig. 3 (P = Posterior, A = Anterior, L = Left, R = Right)

Fig. 7(a) illustrates the true effects of the working memory task, regardless of whether the task is performed in quiet or noisy (babble noise) conditions. The MIPs are the results obtained from the conjunction of the effects of WMQ, WMN and WMN>N, minus the effects of N. Fig. 7(b) is Fig. 7(a) overlaid onto the participant's coronal slices. The four slices at different y locations clearly indicate the role of TE 1.1, BA3a, BA4p, SMA, putamen and SFG when the participant was performing the working memory task. It is evident from Fig. 7(a) that TE1.1 (-34/-34/16) is the cluster that shows the highest intensity as compared to the others, indicating the important role of this area as a processing centre in the execution of the working memory task used in this study.

3.3 Psychophysiological Interaction (PPI)

The activated areas that have been obtained from conjunction analysis of the effects of the working memory task in the left and right hemispheres are the left and right superior temporal gyrus (STG) as the source region, the left and right TE1.1, left BA3a, right BA4p, right superior frontal gyrus (SFG), right putamen and SMA. Due to the midsagittal activation for the SMA, it is considered that the same SMA is involved in the connectivity in the left and right hemispheres, during both WMQ and WMN. The goal is to use PPI to examine the change in

effective connectivity between the STG as the input region and all the other activated regions as the target regions in the left and right hemisphere while the participant is performing the working memory task under the experimental treatment of less attention (WMQ) and more attention (WMN).

Table 3. Statistics obtained from regression analysis on PPI data between the target region and the region of interest

	<i>b</i>	β	r^2	<i>p</i>	95% CI
<i>Left hemisphere areas in quiet</i>					
Left TE1.1 vs. left STG	0.504	0.632	0.400	<0.001	0.391 – 0.616
Left BA3a vs. left STG	0.491	0.637	0.405	<0.001	0.383 – 0.600
SMA vs. left STG	0.479	0.568	0.322	<0.001	0.352 – 0.605
<i>Left hemisphere areas in noise</i>					
Left TE1.1 vs. left STG	0.552	0.661	0.437	<0.001	0.438 – 0.667
Left BA3a vs. left STG	0.519	0.653	0.427	<0.001	0.409 – 0.629
SMA vs. left STG	0.506	0.591	0.349	<0.001	0.380 – 0.631
<i>Right hemisphere areas in quiet</i>					
Right BA4p vs. right STG	0.429	0.568	0.322	<0.001	0.316 – 0.543
Left SFG vs. right STG	0.588	0.541	0.292	<0.001	0.421 – 0.755
SMA vs. right STG	0.471	0.600	0.361	<0.001	0.357 – 0.586
Right putamen vs. right STG	0.202	0.535	0.286	<0.001	0.144 – 0.260
<i>Right hemisphere areas in noise</i>					
Right BA4p vs. right STG	0.424	0.574	0.330	<0.001	0.314 – 0.534
Left SFG vs. right STG	0.532	0.500	0.250	<0.001	0.364 – 0.700
SMA vs. right STG	0.437	0.572	0.328	<0.001	0.323 – 0.551
Right putamen vs. right STG	0.163	0.426	0.181	<0.001	0.100 – 0.226

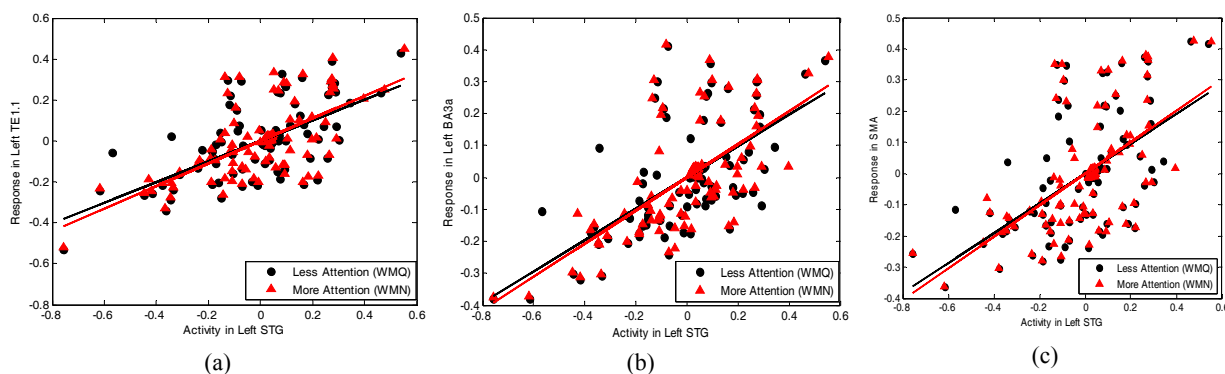


Figure 8. Psychophysiological interaction between left STG and (a) TE1.1, (b) BA3a and (c) SMA without (dark) and with (red) the presence of babble noise

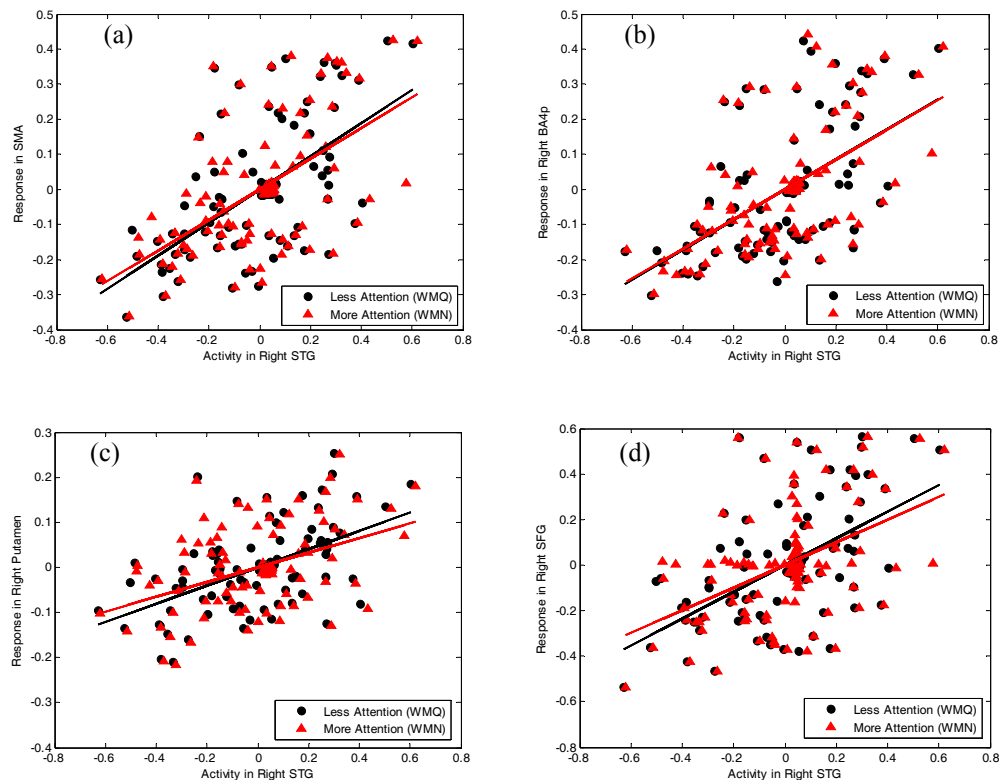


Figure 9. Psychophysiological interaction between right STG and (a) right SMA, (b) BA4p, (c) right putamen and (d) right SFG without (dark) and with (red) the presence of babble noise

Figs. 8 illustrate the response in the left TE1.1, BA3a and SMA plotted as a function of the activity in the left STG obtained from psychophysiological interaction (PPI) analyses. It can be said that the increase in activity in the left STG is associated with increases in the responses in TE1.1, BA3a and SMA both in noise and in quiet. The PPI during execution of the working memory task seemed to be increased in noise (more attention) as compared to in quiet (less attention) as shown by the steeper slope of the red lines. Similar relationships are obtained in the right hemisphere between the right STG and BA4p, SMA, SFG and putamen as shown in Fig. 8. However, these right hemisphere regions indicated an increase in the slope during less attention (dark lines) as compared to more attention except for the BA4p – STG relationship in which the two lines have almost equal slopes.

The statistics characterising all of the slopes shown in Fig. 8 and 9 are tabulated in Table 3 (a) – (d) as obtained from simple linear regression analysis. All the data fulfill the assumptions of normality, linearity and homoscedasticity for the predicted values and residuals. It can be said that the results obtained from simple linear regression analysis have shown that there is a significant, positive and linear relationship between the responses in all the target regions and the activity in the STG ($n = 120$, $p < 0.001$, 95% CI). However, the proportion of the variation in the response of the target regions that can be explained by the activity in the STG, according to the linear regression model is less than 50% (18 – 44%) due to a large dispersion of the data.

4. Discussion

The MNI x , y , z coordinates shown in Table 1 have been found to be located in the auditory associated areas, mainly in the middle and superior temporal gyri of the left and right temporal lobes, according to Specht & Reul (2003) in which brain activations due to attentive listening to tones, sounds and words were comprehensively recorded. The two maxima in the left and right temporal lobes that mirrored each other show that the processing centres for attentive listening to babble noise are the same in both hemispheres, at least for this particular participant. The auditory cortex can be functionally and structurally divided into three main parts which are the primary, secondary and tertiary auditory cortices, forming concentrically around one another. The most inner part is the primary auditory cortex and the most outer part is the tertiary auditory cortex. Widespread activation

observable in the bilateral temporal lobes (Fig. 2) can be explained as due to the commonly known tonotopic characteristic of the auditory cortex especially in the primary auditory cortex. Tonotopic characteristic means that certain cells in the auditory cortex are sensitive to specific frequencies. The babble noise spectrum used in this study consists of a wide range of frequencies which in turn excite many cell regions that are responsible for each particular frequency, resulting in widespread activation. The tonotopic characteristic for the auditory cortex was evident in Langers et al. (2007) and Langers & van Dijk (2012). In their study, a prominent low-to-high frequency tonotopic gradient was observed in the lateral to medial direction of the transverse temporal gyrus (TTG), in the primary auditory field and planum temporale. It has been demonstrated (Specht & Reul 2003) that the primary auditory cortex responded to a broad range of auditory stimuli while the secondary auditory cortex responded to stimuli with complex spectral dynamics. Similar temporal lobe activations (average) have also been found in our previous study using a white noise stimulus on five right handed male participants (Yusoff et al. 2011). The absence of brain activations in the frontal and parietal lobe regions indicate that listening to babble noise does not involve any higher cognitive function over and above baseline comparisons. The activation in the auditory cortex may also indicate that the participant's auditory processing is working well in conjunction with the participant's good hearing condition which has been confirmed by the auditory test performed on the participant's left and right ears.

Interestingly, a higher number of activated voxels in the right auditory cortex is accompanied by a lower t statistic at the point (voxel) of maximum intensity as compared to the left auditory cortex which has fewer activated voxels but with higher t statistic at the point of maximum intensity (Table 1). Not only did the number of activated voxels and the t statistic for the two hemispheres differ in values, but the significance of these effects was very large. Based on previous findings (Hwang et al. 2005), more voxels have been found to be activated in the right temporal lobe as compared to the left temporal lobe when noise is used as stimulus, due its non-speech property (Hwang et al. 2005). Speech stimulus, on the other hand, will most likely activate the left temporal lobe relatively higher than the right temporal lobe especially when perception of words comes into play and only if the stimulus task is linguistically demanding (Hwang et al. 2005). The t value for the voxel of maximum intensity is at odds with the observed number of activated voxels in both temporal lobes but in good agreement with the mean stimulus efficacy (ϵ) value. The mean stimulus efficacy is defined as the potency of the stimulus in eliciting a neuronal response (Friston et al. 2000). It reflects the ability of the ensuing synaptic activity to induce a signal. It seems that the left hemisphere temporal lobe has a stronger ability in the processing of babble noise as compared to the right temporal lobe. One possible explanation for the increase in the mean stimulus efficacy is that the babble noise used in this study was still in the range intelligible to the participant who was young and healthy and was able to be processed by the left hemisphere temporal lobe. It is perhaps this mean stimulus efficacy of the left hemisphere temporal lobe that causes the signal intensity to be higher in the left temporal lobe, hence the higher t value. The fact that signal intensity increases correspondingly with the mean stimulus efficacy is evident in Fig. 5. The underlying mechanism will be discussed in the following paragraph.

While maintaining the activation in the cortical auditory regions, performing a phonological working memory task in quiet (WMQ) and noisy (WMN) conditions has evoked responses in many other cortical regions as tabulated in Table 2 (a) and (b) and can be clearly seen in Figs. 4(a) and (b). Among others are TE1.1 which is known as the primary auditory area, supplementary motor area (SMA), Brodmann Area (BA) 3a and 3p, inferior parietal cortex (IPC) and putamen. These regions have been found to be activated during working memory tasks in several previous studies (Gaab et al. 2003; Karlsgodt et al. 2005; Toepper et al. 2010). Nevertheless, the differences in the brain activated regions during WMQ and WMN can hardly be seen. This could be due to the same regions in the brain responding to the phonological working memory task used in this study whether performed in quiet or in noise. The result is indirectly supported by the behavioural results which indicate similar participant's performance during WMQ and WMN. The spatial extent of activation that seems to be wider from the 70-mm level down to 20-mm level during WMQ, involves the first four activation clusters shown in Table 2 (a). A larger activation area means that a larger number of brain cells are recruited in performing the task in quiet as compared to in noise. In a finger tapping task (Jancke et al. 1998), it was found that a larger brain area was involved in performing the tapping at 3 Hz as compared to at 1 Hz. They concluded that faster movement involves the recruitment of more motor units and will therefore activate a greater number of voxels. Their findings were later reconfirmed (Lutz et al. 2005). In this study, a higher number of activated voxels reflects the higher sensitivity of the brain areas in perceiving the stimulus and later performing the task in quiet. During WMN, it can be said that the participant used a lesser volume of the respective brain regions in accomplishing the working memory task due to the distraction from the noise. Brain areas that are supposed to be involved in processing the phonological working memory task used in this study are less sensitive when the stimulus is

embedded in noise. In other words their ability to perceive what is delivered is relatively lower in noise than in quiet. The stimulus-to-noise ratio (SNR) used in this study was 5 dB. Furthermore, a lower noise intensity (I_N) as compared to stimulus intensity (I_S) used in this study did not enhance the spatial extent of activation as found in previous studies on multiple participants (Wong et al. 2009; Wong et al. 2008). Enhancement of spatial activation occurs in certain brain regions when the participants performed working memory (Abd Hamid et al. 2011) or speech perception (Wong et al. 2009; Wong et al. 2008) tasks in noisy backgrounds (SNR = + 3dB and -5 dB respectively) due to an increase in participants' effort in attending and performing the task while being distracted by noise. One possible explanation for the effects observed for this particular participant is that the sensitivity of the particular brain region in perceiving and performing the phonological working memory task in quiet has exceeded the activation enhancement in performing the task in noisy background. The +5dB SNR used in the present study may not be distracting enough for the participant who is relatively young and has good hearing.

This argument is indirectly supported by the steeper slope (m) of the t value vs. ε plot for WMQ, $m = 2.58$ s, as compared to the less steeper slope of the t value vs. ε plot for WMN, $m = 2.23$ s, as shown by Figs. 5(a) and (b). A steeper slope for WMQ means that the blood oxygenation level dependent (BOLD) signal intensities, hence the hemodynamic responses, across the activated regions change relatively larger during WMQ as compared to during WMN due to a better efficiency of the ensuing synaptic activity to induce the signal in quiet. The mean stimulus efficacy thus represents the increase in perfusion signal elicited by neuronal activity, expressed in terms of number of evoked transients per second (Friston et al. 2007). From the plots, it appears that the extent of signal change across activated regions is larger in quiet as compared to in noise for the same change in the mean stimulus efficacy. However, the distributions of data for the two conditions are about the same, resulting in similar values of the Spearman coefficient for both plots.

The participant-specific brain activation characteristic discussed above may differ from participant to participant. However, in view of the brain activation obtained from WMQ and WMN conditions that have been averaged over 20 measurements, the brain activation results can be said to be valid, at least for this particular participant due to the high sensitivity of sparse imaging technique used in this study in capturing the change in the magnetic resonance signal intensity during the task performance (Gaab et al. 2007; Mueller et al. 2011).

Further analyses conducted using contrast comparison and conjunction analysis on our participant revealed two additionally important findings reflecting 1) the true effects of phonological working memory task performance in babble noise and in quiet and 2) that the brain areas involved in performing the working memory task are similar regardless of whether the task was performed in quiet or in noise. The six brain areas found to be significantly activated at corrected $\alpha = 0.05$ are left TE1.1, left BA3a, right BA4p, BA6/SMA, putamen and SFG. These six brain areas appear to be reproducible using WMN - N, WMN > N and WMQ - N contrasts, using conjunction of the effects of WMN - N and WMN > N and the conjunction of the effects of WMQ, WMN and WMN > N, minus the effects of N. The difference between WMN - N and WMN > N contrasts is that the former is looking for voxels or brain areas that are activated during WMN after subtracting it from areas that are activated during N, while the latter is looking for voxels that are more (significantly) activated in WMN than in N. What is revealed by taking the conjunction between these two contrasts should be the areas that are involved in processing the phonological working memory task stimulus delivered to the participant during the scan.

The conjunction of the effects of WMQ, WMN and WMN > N, minus the effects of N confirmed the earlier results and is reconfirmed by the WMQ - N contrast. Conjunction analysis permits the determination of the areas in the brain for a specific function that are common in all tasks under study. It allows for the identification of areas that share a specific function in a participant. For the WMQ - N contrast, even though the effects of N are not inherited in WMQ, the activation obtained can be compared with WMN - N and WMN > N contrasts because what is taken out from the WMQ - N contrast are voxels that are involved in listening or, in particular, in babble noise, leaving behind voxels that do not belong to the listening task when performed during N condition.

The exact location of the six brain areas found to be involved in the phonological working memory task can be observed in Fig. 7 from which the statistical images are overlaid coronally onto the participant's normalised brain images. Out of the six areas, the left TE1.1, left BA3a, BA6/SMA and right BA4p have shown an increase in the number of activated voxels when the task is performed in quiet as compared to in noise, see Table 2. In relation to the above discussion, it can be said that the sensitivity of these particular brain regions in perceiving and performing the phonological working memory task in quiet increases, which is reflected by an increase in the total volume of the activation area.

As shown in Fig. 2, the psychophysiological interaction (PPI) explains the regionally specific neuronal responses in the target regions in terms of the interaction between influences from a source region and task-related

parameters such as babble noise. In this study, the regionally specific neuronal responses are measured in the brain regions shown in Fig. 7 which are the left TE1.1, left BA3a, SMA, right BA4p, right SFG and right putamen, also known as the target regions. The response in these regions is said to be influenced by the interaction between the activity in the STG, the source region, and the task-related parameter which is the babble noise. The left and right STG are chosen to be the source region whose activity interacts with babble noise as well as with the working memory stimulus due to the source region's well known functional role in processing verbal and nonverbal stimuli. Furthermore, STG is also an area in which the primary and secondary auditory areas are located and from where the input (or stimuli) traverses to the other areas in the brain. The activity in STG, therefore on the one hand interacts with babble noise to evoke responses in the target areas in which babble noise is said to modulate the interaction between the STG with other regions connected to it. On the other hand, babble noise also has a direct influence on the target areas as shown in Fig. 2. Thus, it can be said that the responses in the target regions reflect the combined effect of the interaction between the source regions and babble noise and the influence of babble noise itself on target regions.

To measure the influence of the activity in the source region (in this study the left and right STG) on the response of the target regions (as shown in Fig. 7), the PPI data of the target region (as the dependent variable) is plotted against the PPI of the source region (as the independent variable). Figs. 8(a – c) and 9(a – d) illustrate psychophysiologic interaction (PPI) between the two distal brain areas, namely each of the areas in Fig. 7 vs. the left or right STG. A linear increase in the response in all the target regions with the increase in activity in the source region as depicted provides evidence in support of a direct influence of the activity of the source region on the response of the target regions. Linearity is shown in all the plots regardless of whether the task is performed in quiet (dark lines) or in noise (red lines). The slope (b) of these plots reflects the influence the source region exerted over the target region (see Table 3). It can be clearly seen that the slope of the plots of the PPI data acquired during the performance of the working memory task in quiet (that needs less attention) differs from that obtained from the performance of the working memory with babble noise as background (that needs more attention), except for the one shown in Fig. 9(a). The change in the slope of a plot when measured in a different context indicates a psychophysiologic interaction, with babble noise as contextual input modulating the influence of the activity in STG on the response of the target regions. It is very interesting to see that in the left hemisphere, the slope is steeper for the condition that needs more attention (working memory task in noisy background) as compared to the condition requiring less attention (working memory task in quiet). These involve the interactions of left TE1.1 vs. left STG, left BA3a vs. left STG and SMA vs. left STG. In contrast, in the right hemisphere, the slope for the condition requiring less attention (WMQ) is steeper than the slope for the condition requiring more attention (WMN), involving interactions of right SFG vs. right STG, SMA vs. right STG and right putamen vs. right STG. However, the interaction of right BA4p vs. right STG is similar in both WMQ and WMN conditions.

A steeper slope of a PPI plot means that a small change in activity in the STG results in a relatively larger response in the target regions as compared to a plot with less steeper slope. For TE1.1 and BA3a in the left hemisphere including the SMA, the interaction between these regions and STG is said to be mediated by babble noise and their responses to the activity in the left STG is increased in the presence of babble noise. TE1.1, together with TE1.0 and TE1.2 are known to be parts of the primary auditory cortex and are grouped together as TE1 by Morosan et al. (2001). They have tonotopic progressive characteristics such that the high frequencies are represented in TE1.1 while the lower frequencies are in TE1.2. The intermediate frequencies that are important for speech are processed in TE1.0. The TE1.1 region which is not directly involved with phonological working memory task shows an increase in the slope of the PPI plot in noise as compared to in quiet, indicating an increase in the effort in perceiving and processing the sound of the stimulus presented in a noisy background. The same argument is proposed to be valid for the SMA and BA3a. The SMA has the functions of retrieval of learned sequences that are under internal control such as the performance of a sequence of movements from memory. The fact that the SMA is activated even in the performance of phonological working memory task is possibly related to considerations of the diverse activities in which the SMA plays a role, suggesting that existing theories may not fully capture the fundamental functions of this region (Nachev et al. 2008). The significant activation observed in the left BA3a is due to the execution of action and speech when the participant was performing the task in which the participant is required to repeat backward the series of five words presented. The existence of SMA–STG and SMA–BA3a interactions supports the concept of audio-motor interaction (Hickok et al. 2003).

On the opposite hemisphere, the PPI plots of the response of the target regions as a function of the activity of the source region exhibited an increase in the slope of the regression in quiet as compared to in noise. This could be

due to a direct involvement of the three areas, namely the superior frontal gyrus (SFG), putamen and SMA in the execution of the task, in response to the activity in the right STG. The SFG is also known to be located in the dorsolateral prefrontal cortex (DLPFC), classified as BA8 and BA9. This frontal area is involved in management of uncertainty in which increasing uncertainty increases activation. Uncertainty may exist during performance of the task due to the participant's inability to completely capture what is presented. This area is also known to be involved in sustaining attention and working memory (Brzezicka et al. 2011). Similar arguments apply to the putamen in which reinforcement and implicit learning, which are intrinsically part of the task, might explain its significant involvement. Apart from that, the putamen which is known to be interconnected to various pathways will certainly be activated due to its involvement in various regional interactions. As for the right BA4p, the phonological working memory task used required the participant to repeat words, resulting in significant activation in this area. The fact that the function of this area is more inclined towards motor coordination than working memory processing has rendered the PPI with the source region similar whether the task is performed in quiet or in noise. The fact that the SMA has shown a different response in PPI plots for the task presented in noise and in quiet when subjected to the activity in the left and right STG needs further investigation, especially when noise is used as the modulatory input.

5. Conclusion

In conclusion, the phonological working memory task based on the repeat backward test used in this study activates several brain regions that are involved in processing of the task. Similar brain regions are activated regardless of whether the task is performed in noise or in quiet. However, the activated brain regions in the longitudinal range of 70 mm to 20 mm appear to have extended activation when the task is performed in quiet due to the sensitivity of particular brain regions in perceiving the stimulus and later performing the task in quiet. Conjunction analysis reveals task-related areas which are the supplementary motor area (SMA), left TE1.1, left Brodmann Area (BA) 3a, right BA4p, right putamen and right superior frontal gyrus (SFG) which are activated with p values well below 0.05 after controlling for family wise error (FWE). A steeper psychophysiological interaction (PPI) slope for the right hemisphere task-related regions during the quiet condition indicates a larger influence of the source region (right STG) onto the target regions (SMA, right putamen, right SFG), except for BA4p which shows similar PPI slopes for in-quiet and in-noise measurements. For the left hemisphere regions (left TE1.1, left BA3a) and SMA, the interaction is modulated by noise, resulting in the increase in the PPI slope in the noisy condition. It can be said that the babble noise used in this study has little influence on the responses of working memory processes related to areas of input of the source region, resulting in a higher response of those areas in quiet as compared to in noise. This study, even though performed on a single participant has revealed several important fundamental insights for future multiple participant fMRI studies on working memory and is potentially viable to be implemented in a clinical environment. However, due to the different approaches used in studies of working memory which have employed various tasks and methodologies, we found that it is quite difficult for a direct comparison with the results obtained from other studies to be made at this time.

Acknowledgements

The authors would like to thank Sa'don Samian, the MRI Technologist of the Universiti Kebangsaan Malaysia Hospital (HUKM), for the assistance in the scanning and the Department of Radiology, Universiti Kebangsaan Malaysia Hospital for the permission to use the MRI scanner. The authors were also indebted to Dr. Stephan Kiebel and Dr. Carsten Müller from Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany, for valuable discussions on experimental methods and data analyses. Parts of the Methods and Results sections of this article were written at the Max Planck Institute during a research attachment. This work was supported by the UKM-GUP-SK-07-20-205 research grant.

References

- Abd Hamid, A. I., Yusoff, A. N., Mukari, S. Z., & Mohamad, M. (2011). Brain Activation during Addition and Subtraction Tasks In-Noise and In-Quiet. *Malays J Med Sci*, 18(2), 3-15.
- Amaro, E., Jr., & Barker, G. J. (2006). Study design in fMRI: basic principles. *Brain Cogn*, 60(3), 220-232. <http://dx.doi.org/10.1016/j.bandc.2005.11.009>
- Baddeley, A. D. (1996). Exploring the central executive. *Quarterly J Exp Psychol*, 49A, 5-28. <http://dx.doi.org/10.1080/713755608>
- Baddeley, A. D. (2002). Is working memory still working? *European Psychologist*, 7(2), 85-97. <http://dx.doi.org/Doi.10.1021//1016-9040.7.2.85>

- Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In Recent advances in learning and motivation (Bower GA, ed.). Academic Press, New York, p. 47-90.
- Belin, P., Zatorre, R. J., Hoge, R., Evans, A. C., & Pike, B. (1999). Event-related fMRI of the auditory cortex. *Neuroimage*, *10*(4), 417-429. <http://dx.doi.org/10.1006/nimg.1999.0480>
- Bernal, B., & Altman, N. R. (2001). Auditory functional MR imaging. *AJR Am J Roentgenol*, *176*(4), 1009-1015. <http://dx.doi.org/10.2214/ajr.176.4.1761009>
- Brzezicka, A., Sedek, G., Marchewka, A., Gola, M., Jednorog, K., Krolicki, L., & Wrobel, A. (2011). A role for the right prefrontal and bilateral parietal cortex in four-term transitive reasoning: An fMRI study with abstract linear syllogism tasks. *Acta Neurobiol Exp (Wars)*, *71*(4), 479-495.
- Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., & Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage*, *25*(4), 1325-1335. <http://dx.doi.org/10.1016/j.neuroimage.2004.12.034>
- Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997). Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage*, *6*(3), 218-229. <http://dx.doi.org/10.1006/nimg.1997.0291>
- Friston, K. J., Holmes, A., Poline, J. B., Price, C. J., & Frith, C. D. (1996). Detecting activations in PET and fMRI: levels of inference and power. *Neuroimage*, *4*(3Pt1), 223-235. <http://dx.doi.org/10.1006/nimg.1996.0074>
- Friston, K. J., Mechelli, A., Turner, R., & Price, C. J. (2000). Nonlinear responses in fMRI: The Balloon model, Volterra kernels, and other hemodynamics. *Neuroimage*, *12*(4), 466-477. <http://dx.doi.org/10.1006/nimg.2000.0630>
- Gaab, N., Gabrieli, J. D., & Glover, G. H. (2007). Assessing the influence of scanner background noise on auditory processing. II. An fMRI study comparing auditory processing in the absence and presence of recorded scanner noise using a sparse design. *Hum Brain Mapp*, *28*(8), 721-732. <http://dx.doi.org/10.1002/hbm.20299>
- Gaab, N., Gaser, C., Zaehle, T., Jancke, L., & Schlaug, G. (2003). Functional anatomy of pitch memory--an fMRI study with sparse temporal sampling. *Neuroimage*, *19*(4), 1417-1426. [http://dx.doi.org/10.1016/S1053-8119\(03\)00224-6](http://dx.doi.org/10.1016/S1053-8119(03)00224-6)
- Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends Cogn Sci*, *3*(11), 410-419. [http://dx.doi.org/10.1016/S1364-6613\(99\)01388-1](http://dx.doi.org/10.1016/S1364-6613(99)01388-1)
- Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R., Gurney, E. M., & Bowtell, R. W. (1999). "Sparse" temporal sampling in auditory fMRI. *Hum Brain Mapp*, *7*(3), 213-223. [http://dx.doi.org/10.1002/\(SICI\)1097-0193\(1999\)7:3<213::AID-HBM5>3.0.CO;2-N](http://dx.doi.org/10.1002/(SICI)1097-0193(1999)7:3<213::AID-HBM5>3.0.CO;2-N)
- Hall, D. A., Summerfield, A. Q., Goncalves, M. S., Foster, J. R., Palmer, A. R., & Bowtell, R. W. (2000). Time-course of the auditory BOLD response to scanner noise. *Magn Reson Med*, *43*(4), 601-606. [http://dx.doi.org/10.1002/\(SICI\)1522-2594\(200004\)43:4<601::AID-MRM16>3.0.CO;2-R](http://dx.doi.org/10.1002/(SICI)1522-2594(200004)43:4<601::AID-MRM16>3.0.CO;2-R)
- Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T. (2003). Auditory-motor interaction revealed by fMRI: Speech, music, and working memory in area Spt. *J Cogn Neurosci*, *15*(5), 673-682. <http://dx.doi.org/10.1162/089892903322307393>
- Hwang, J. H., Wu, C. W., Chou, P. H., Liu, T. C., & Chen, J. H. (2005). Hemispheric difference in activation patterns of human auditory-associated cortex: An FMRI study. *ORL J Otorhinolaryngol Relat Spec*, *67*(4), 242-246. <http://dx.doi.org/10.1159/000089501>
- Jancke, L., Peters, M., Schlaug, G., Posse, S., Steinmetz, H., & Muller-Gartner, H. (1998). Differential magnetic resonance signal change in human sensorimotor cortex to finger movements of different rate of the dominant and subdominant hand. *Brain Res Cogn Brain Res*, *6*(4), 279-284. [http://dx.doi.org/10.1016/S0926-6410\(98\)00003-2](http://dx.doi.org/10.1016/S0926-6410(98)00003-2)
- Karlsgodt, K. H., Shirinyan, D., van Erp, T. G., Cohen, M. S., & Cannon, T. D. (2005). Hippocampal activations during encoding and retrieval in a verbal working memory paradigm. *Neuroimage*, *25*(4), 1224-1231.
- Langers, D. R., & van Dijk, P. (2012). Mapping the tonotopic organization in human auditory cortex with minimally salient acoustic stimulation. *Cereb Cortex*, *22*(9), 2024-2038. <http://dx.doi.org/10.1093/cercor/bhr282>

- Langers, D. R., Backes, W. H., & van Dijk, P. (2007). Representation of lateralization and tonotopy in primary versus secondary human auditory cortex. *Neuroimage*, *34*(1), 264-273. <http://dx.doi.org/10.1016/j.neuroimage.2006.09.002>
- Lutz, K., Koeneke, S., Wustenberg, T., & Jancke, L. (2005). Asymmetry of cortical activation during maximum and convenient tapping speed. *Neurosci Lett*, *373*(1), 61-66. [http://dx.doi.org/S0304-3940\(04\)01220-010.1016/j.neulet.2004.09.058](http://dx.doi.org/S0304-3940(04)01220-010.1016/j.neulet.2004.09.058)
- Mueller, K., Mildner, T., Fritz, T., Lepsien, J., Schwarzbauer, C., Schroeter, M. L., & Moller, H. E. (2011). Investigating brain response to music: a comparison of different fMRI acquisition schemes. *Neuroimage*, *54*(1), 337-343. <http://dx.doi.org/10.1016/j.neuroimage.2010.08.029>
- Nachev, P., Kennard, C., & Husain, M. (2008). Functional role of the supplementary and pre-supplementary motor areas. *Nature Reviews Neuroscience*, *9*(11), 856-869. <http://dx.doi.org/10.1038/nrn2478>
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97-113. [http://dx.doi.org/10.1016/0028-3932\(71\)90067-4](http://dx.doi.org/10.1016/0028-3932(71)90067-4)
- Specht, K., & Reul, J. (2003). Functional segregation of the temporal lobes into highly differentiated subsystems for auditory perception: An auditory rapid event-related fMRI-task. *Neuroimage*, *20*(4), 1944-1954. <http://dx.doi.org/10.1016/j.neuroimage.2003.07.034>
- Toepper, M., Markowitsch, H. J., Gebhardt, H., Beblo, T., Thomas, C., Gallhofer, B., Driessen, M., & Sammer, G. (2010). Hippocampal involvement in working memory encoding of changing locations: An fMRI study. *Brain Res*, *1354*, 91-99. <http://dx.doi.org/10.1016/j.brainres.2010.07.065>
- Wong, P. C., Jin, J. X., Gunasekera, G. M., Abel, R., Lee, E. R., & Dhar, S. (2009). Aging and cortical mechanisms of speech perception in noise. *Neuropsychologia*, *47*(3), 693-703. <http://dx.doi.org/10.1016/j.neuropsychologia.2008.11.032>
- Wong, P. C., Uppunda, A. K., Parrish, T. B., & Dhar, S. (2008). Cortical mechanisms of speech perception in noise. *J Speech Lang Hear Res*, *51*(4), 1026-1041. [http://dx.doi.org/10.1044/1092-4388\(2008/075\)](http://dx.doi.org/10.1044/1092-4388(2008/075))
- Yusoff, A. N., Mohamad, M., Hamid, K. A., Abd Hamid, A. I., & Mukari, S. Z. M. S. (2011). Acquisition, Analyses and Interpretation of fMRI Data: A Study on the Effective Connectivity in Human Primary Auditory Cortices. *Sains Malaysiana*, *40*(6), 665-678.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

Structural Dynamic Study of Roof Waterproofing Materials

Alim Feizrakhmanovich Kemalov¹, Ruslan Alimovich Kemalov¹, Dinar Zinnurovich Valiev¹ & Ilmira Maratovna Abdrafikova¹

¹Kazan (Volga region) Federal University, 8, Kremlevskaya Street, Kazan, 420008, Russian Federation

Correspondence: Alim Feizrakhmanovich Kemalov, Kazan (Volga region) Federal University, 8, Kremlyovskaya Street, Kazan, 420008, Russian Federation. E-mail: Kemalov@mail.ru

Received: June 27, 2014

Accepted: July 17, 2014

Online Published: August 17, 2014

doi:10.5539/mas.v8n5p115

URL: <http://dx.doi.org/10.5539/mas.v8n5p115>

Abstract

The present research was aimed to develop the scientific applied principles and technologies of composite bituminous materials for civil engineering based on the investigation of the structures of polymer modifier and bitumen-polymer binder (BPB) on its basis with the use of nuclear magnetic resonance (NMR). The method of pulsed NMR was chosen as one of the rapid analysis methods that can be used for the analysis of bitumen-polymer systems, especially when assessing the group chemical composition of residual oil feedstock (ROF), bitumens and composite materials based on them. Using the method of pulsed NMR the regularities of the impact of modifier component composition on the changes of structural-group composition of the original and modified products were specified. Based on the results of research the optimal ratio of bitumen-polymer binder components was investigated, the manufacturability of the process for obtaining of composite bituminous materials for civil engineering with the aim of optimizing the quality of the final products was evaluated. Pulsed NMR - spectroscopy is suggested as input and output quality control of bituminous products. The regularities of redistribution of the phases with different molecular mobility and their relationship with the binder components were investigated. Rapid technique for quantifying the content of polymer in the solvent was developed.

Keywords: pulsed NMR, bitumen-polymer binder, structural-group composition, roof waterproofing materials, thermoplastic resins

1. Introduction

Nuclear magnetic relaxation (NMR) is one of the modern instrumental methods of physical chemistry (Kimmich and Anozzo, 2004; Conte and Alonzo, 2013, Xu et al., 2013, Bayer et al., 2013; Haber-Pohlmeier et al., 2013) and is a fundamental property of nuclear magnetism, characterizing the dynamics of nuclear spins system in the oil bitumen-polymer systems.

Pulsed NMR method increasingly finds its application in rapid analysis of the bitumen-polymer systems, especially when assessing the group chemical composition of residual oil feedstock (ROF), bitumens and composite materials based on them. It should be noted that the study of NMR spectra is based on the resonating nuclei systems that are extremely sensitive to the magnetic environment where the local magnetic fields near the resonating nuclei are dependent on the intra- and intermolecular effects. This determines the value of NMR spectroscopy to investigate the structure and behavior of many-electron (molecular) systems in different fields of science (Washburn and Birdwell, 2013; Nascimento, *et al.*, 2004; Bayer, *et al.*, 2010; Berman, *et al.*, 2013; Charlier, *et al.*, 2013; Kemalov, *et al.*, 2012; Kemalov *et al.*, 2013; Kruk *et al.*, 2012, Twieg, 2013, Frantsuzov, 2010).

Thus, the method of pulsed NMR today attracts not only physicists and chemists, but also experts in the field of technology, which is primarily due to the practical significance of the NMR data, the processes of which depend on many physical and chemical properties of the system and illustrate the excellent dynamic phenomena at a molecular level.

The research of roofing waterproofing materials using NMR relaxometry primarily allows to study the effect of temperature on the structural and dynamic state of the oil disperse system (ODS) at the molecular level with a subsequent change of thermodynamic characteristics of the spin-spin components of the ODS. Thus the structural features of ODS in general can reliably be assumed (Kemalov and Kemalov, 2013).

Waterproofing materials that are related to ODS, are known for a long time and are used for different purposes:

to protect roofs against the penetration of atmospheric precipitates (roofing), for waterproofing in the construction of underwater and above-water parts of dams and bridges (waterproofing). In order to reduce costs and extend the life of roofing materials by virtue of their physical nature and structural features the injection of specially selected polymer modifiers according to their chemical nature and reactivity in the operating conditions in the construction is conducted, resulting in a change in the structure and properties of organic binding materials. Injection of a suitable polymer modifier gives cementing materials more heat and frost resisting, elasticity, increased resistance to fatigue, improves durability (Kemalov and Kemalov, 2013).

One of the most promising polymer materials used in civil construction at the bitumen of different brands modification with is the class of thermoplastic resins (TPR). Due to the fact that the injection of the chosen TPR into the bitumen (Russian standard – “technical terms” -TU 2451-089-05766801-99) in the solid state (with no use of solvent-plasticizer), there is hardening of the polymer-bitumen material, while the softening temperature increases (from 89 to 120 °C) because of the high structuring ability of TPR. This is the reason for the lack of the proper elasticity, frost resisting of roofing material. Therefore, at the modification of this raw material it is advantageous to inject the selected polymer either in the form of a polymer solution (polymer + plasticizer-solvent) or in a way of step-by-step injection of polymer and a plasticizer-solvent. In studies we used the plasticizer that is in a semi-drying oils class (SDOs) as a dissolving agent. Selected SDO (Russian State standard - GOST 8988-59) is characterized by the presence of large amounts of fatty acids and one double bond. According to the type of fatty acids contained in the glycerides, SDO are composed of saturates – 7,3-10 % , components with one double bond – 67,5-75 % and several double bonds – 15-25,2 % (Kemalov and Kemalov, 2013).

2. Method

While developing the scientific applied principles and technologies of composite bituminous materials for civil engineering the structures of polymer modifier and bitumen-polymer binder (BPB) on its basis were investigated by nuclear magnetic resonance (NMR).

This method belongs to the modern instrumental methods of research, which is based on a fundamental property of nuclear magnetism, characterizing the dynamics of nuclear spins in a dispersed system - polymer solutions, bitumen, bitumen-polymer binders and so on. Highly informative parameters of nuclear magnetic relaxation connected to the tested substance properties, the relative simplicity of the experimental determination of these parameters, as well as the reliability of the theoretical interpretation of the data allow to allocate it to the independent physical method of investigation.

For the analysis of polymer solutions of different concentrations using pulsed NMR, the following measurement modes were used: start-up period $T = 3000 \text{ ms} - 12 \text{ s}$, the interval between 90° and 180° pulses $N = 100-2500$, the number of scans $n = 10-30$ (for the spin-spin system) and the start-up period $T = 500 \text{ ms} - 2 \text{ s}$, the interval between 90° and 180° pulses $N = 50-1000$, the number of scans $n = 10-30$ (for the spin-lattice system). Analysis time did not exceed 3 min.

Selecting the temperature range, at which the studies were conducted, was not accidental. In this temperature range of 20-140 °C the major technological operations with oil bitumen are carried out.

The analysis of solutions of TPR in SDO was conducted at a temperature of 140 °C (the temperature of mixing). As seen from the graph (Figure 1) in the polymer solution at a predetermined temperature, the presence of three phases is observed. The phase with the longest time of relaxation will be specified as "a." The phase that has the shortest time of spin-spin relaxation is "c", and the phase having an intermediate value of relaxation time is denoted as "b". The values of spin-spin relaxation time for the lightest phase T_{2a} and the intermediate phase T_{2b} decrease with the increasing of polymer content in the solution. At the same time, the relaxation time of the heaviest phase is almost independent from the content of the TPR in solution. The dependence of the spin-spin relaxation time from the content of the TPR in solution is linear. Therefore, in general, its analytical expression have the form $y=kx+b$, where k – is slope.

3. Analysis and Results

As a result of analysis of experimental data empirical dependence of the spin-spin relaxation time from the content of the TPR in the sample solution was obtained:

$$T_{2a} = -16.996 x + 1503.9 \quad (1)$$

$$T_{2b} = -5.1158 x + 435.74 \quad (2)$$

$$T_{2c} = -0.7248 x + 64.572, \quad (3)$$

where T_{2a} is the spin-spin relaxation time of the “long-time” phase, T_{2b} – the spin-spin relaxation time of the intermediate phase, T_{2c} – the spin-spin relaxation time of “short-time” phase.

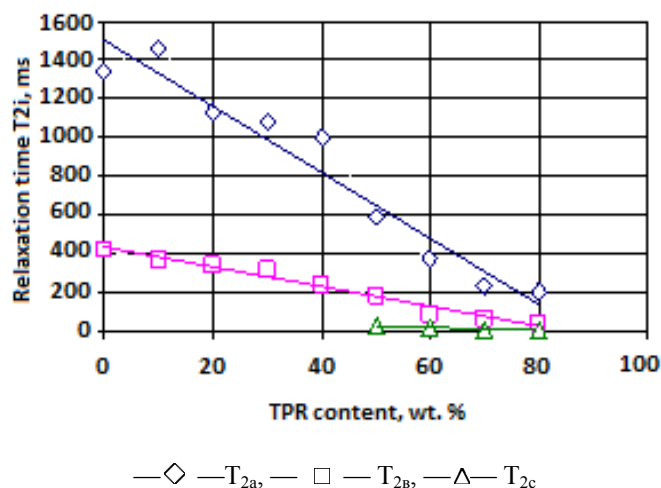


Figure 1. The dependence of spin-spin relaxation T_{2i} from the content of TPR in the solution of SDO at 140 °C

Analysis of the conducted NMR studies experimental data with pure solvent-plasticizer in the entire temperature range from 20 to 140 °C, where the presence of two phases observed, allows suggesting that the presence of the third phase "c" is due to the presence of dissolved TPR. Thus, the phases "a" and "b" characterize the components of the solvent, and the shortest by time phase "c" characterizes the dissolved polymer and some of the solvent that is associated with the polymer. Figure 1 shows that the injection of 10 wt% of TPR into SDO results a decrease of the spin-spin relaxation time values (from 1500 to 1350 and from 425 to 370, respectively) of phases "a" and "b". As mentioned above, the preliminary analysis of the experimental data revealed that the phases "a" and "b" refer to a solvent. Reduction of spin-spin relaxation times of these phases in our view can be explained by the fact that some of the SDO molecules are included into the spatial structure of the polymer, and thus lose some of their mobility. From the nature of the slope (Figure 1), it can be assumed that the polymer has a greater effect on the phase "a", i.e. on the light part of the SDO.

3. Discussion

Decrease of the values of spin-spin relaxation T_{2a} and T_{2b} with the increase of TPR content in SDO also indicates the decrease in the mobility of the system, which confirms the earlier made assumptions about the high structuring ability of TPR.

Figure 2 shows a graph of the proton phases population from the concentration of TPR in SDO.

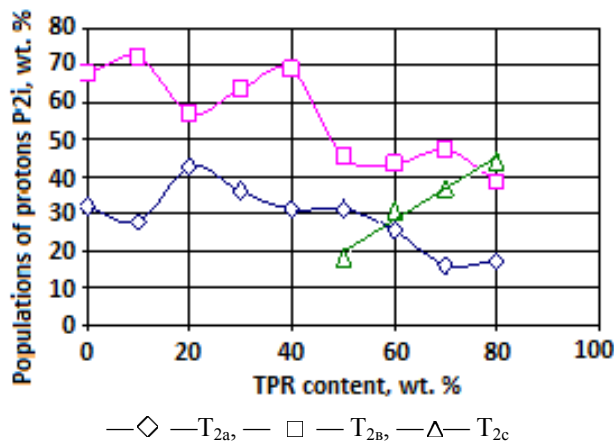


Figure 2. Dependence of the populations of protons in phases P_{2i} from TPR content in the solution of SDO at 140 °C

Diagrams of P_a and P_b dependences have polynomial nature, while P_c dependence is linear. P_a dependence curve has a maximum at the points corresponding to a polymer content of 10 % and 40 % by weight, and the minimum is at the points corresponding to the TPR concentration of 20 % and 50 % by weight. The initial increase in the TPR "long-time" phase (P_{2a}) content from 0 % to 20 % by weight can be explained by the decrease of molecular mobility due to their interaction with the polymer molecules. Thus molecules of this phase verge into a phase with a short relaxation times - phase "b". Probably some of these molecules lose mobility so much that moves into the phase with the shortest time of spin-spin relaxation, i.e. the phase "c". The further small-scale decrease in the phase "a" amount is likely caused by the increased content of phase "b", part of its molecules goes into "c"-phase, if the concentration of TPR is more than 50% by mass. The further reduction of the phase "a" population is probably due to the transition of the molecules of this phase into the phases "b" and "c". As previously mentioned, the dependence of P_{2c} from the content of TPR is linear. As a result of the mathematical analysis of the dependence the empirical expression relating the content of the TPR in the solution and the population of protons in phase "c":

$$P_c = A \cdot C - B \quad (4)$$

where C – the content of the TPR in solution of rapeseed oil, % wt.; P_c - population of protons in phase "c", %; A and B - empirical coefficients: $A = 0.845$, $B = 22.4$

Change in the group composition of bitumen after its modification plays an important role in the study of BPB. NMR method allows determining group composition of bitumen and BPB with an accuracy of 3%, this will allow us to estimate the composition changes, as well as the convergence of the results obtained by other methods (Conte and Alonzo, 2013; Kimmich and Anoardo, 2004, Bayer *et al.*, 2010).

The essence of the definition is as follows. Sample preparation procedure consisted of the analyzed substance to be placed in the tube and fixing it to the sensor unit. To carry out the NMR experiment investigated system (oil residue) is simultaneously placed in a constant magnetic field (MF) and a pairs of pulses of the MF with a variable time interval between pulses. At equilibrium nuclear magnetization is directed along the constant MF. Using a variable magnetic field, the magnetization can be deflected from the constant field direction, thereby moving the nuclear spin system in a controlled non-equilibrium state. Variable magnetic field should be directed perpendicularly to the constant MF, directed perpendicularly to the pulsed magnetic field and opposite to the pulses of pair. Then comes the registration of the magnetization projection of the analyzed bitumen sample on the direction of the constant magnetic field through a fixed time interval after exposure of each pair of magnetic field pulses. In order to increase the informativity of the mode, the initial value of the magnetization projection of the analyzed bitumen sample on the direction of the constant magnetic field is recorded beforehand, whereupon the measuring cycle is performed twice at a sample temperature of 40 °C and 100 °C, respectively (Xu *et al.*, 2013; Kruk *et al.*, 2012; Haber-Pohlmeier *et al.*, 2013). And after each measuring cycle the amplitude of the transverse magnetization of the analyzed sample is determined by extrapolating the linear part of the logarithm of the projection of the sample magnetization and the time interval between the pulses of the magnetic field onto zero time interval between the pulses and then the group content of oils, asphaltenes and resins is determined from the relations: $P_A = A_1 \cdot 100 / A_0$; $P_C = (A_0 - A_2) \cdot 100 / A_0$, $P_B = 100 - P_A - P_C$, where P_A , P_C , P_B – oils, asphaltenes and resins content in the analyzed sample, respectively, mass %; A_0 – the initial value of the projection of the sample magnetization; A_1 and A_2 – the values of amplitude of the transverse magnetization, determined at sample temperature of 40 °C and 100 °C, respectively. Three samples were studied: sample 1 - pure bitumen BN 90/10; sample 2 - BPB (BN 90/10 + 3 % Oil-Polymer resins - OPR, 6% rapeseed oil); sample 3 - BPB (BN 90/10 + 3 % OPR, 7 % rapeseed oil). The components were mixed for 15-20 minutes at 100-120 °C. The results are shown in Table 1.

Table 1. Chemical composition of the BPB

Sample, No.	Oils (saturates)	Resins	Asphaltenes
1	8.14	31.76	60.1
2	18.32	33.65	48.03
3	20.01	33.71	46.28

The optimal composition of plasticizer for BPB is determined (70% wt. OPR + 30% wt. rapeseed oil).

4. Summary

It has shown that the injection of the polymer additive the amount of oil fraction increases, the amount of

asphaltenes decreases, bitumen dilution happens, solvation shell partially immerse into oil fraction. Thus, it was shown that the polymer is interacts to a greater extent with a part of the oil fraction of bitumen (lighter), dissolving therein and thus the polymeric additive increases in percentage of bitumen oil fraction that has been given above using Hildebrandt index, which is by its numerical value practically the same as that of the polymer and of the aromatic maltenes that are included in the oil fraction of the bitumen.

During the comprehensive solution of the scientific applied tasks aimed at developing scientific bases and creating new bitumen-polymer materials recipes and production techniques (GOST 30547-97 "Materials rolled roofing and waterproofing", GOST 30693-2000 "Mastic roofing and waterproofing") based on domestic produced TPR, using a solvent such as plasticizer SDO in the present studies several important and urgent problems of the national economy were solved:

1. Using the method of pulsed NMR the regularities of the impact of modifier component composition on the changes of structural-group composition of the original and modified products were specified;
2. Based on the results of research the optimal ratio of BPB components was investigated, the manufacturability of the process for obtaining of composite bituminous materials for civil engineering with the aim of optimizing the quality of the final products was evaluated.
3. The technological option of the higher developed compositions of roofing waterproofing materials, using pulsed NMR - spectroscopy as input and output quality control of bituminous products was developed.
4. On the basis of pulsed NMR data the regularities of redistribution of the phases P_A , P_B , P_C with different molecular mobility and their relationship with the binder components were investigated. Rapid technique for quantifying the content of polymer in the solvent was developed.

Despite the many advantages (such as non-destructive nature of research), NMR does not provide the data on the chemical composition. In connection with this - in the long term research to develop rapid methods to study the composition of bitumen-polymer systems using NMR spectroscopy in combination with other physical and chemical methods of analysis.

Acknowledgement

The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University

References

- Bayer, J., Jaeger, F., & Schaumann, G. E. (2010). Proton nuclear magnetic resonance (NMR) relaxometry in soil science applications. *The Open Magnetic Resonance Journal*, 3, 15-26. <http://dx.doi.org/1874-7698/10>
- Berman, P., Levi, O., Parmet, Y., Saunders, M., & Wiesman, Z. (2013). Laplace Inversion of Low-Resolution NMR Relaxometry Data Using Sparse Representation. *Concepts in Magnetic Resonance. Part A, Bridging Education and Research*, 42(3), 72-88. <http://dx.doi.org/10.1002/cmr.a>
- Charlier, C., Khan, S. N., Marquardsen, T., Pelupessy, P., Reiss, V., Sakellariou, D., ... & Ferrage, F. (2013). Nanosecond Time Scale Motions in Proteins Revealed by High-Resolution NMR Relaxometry. *Journal of American Chemical Society*, 135, 18665-18672. <http://dx.doi.org/10.1021/ja409820g>.
- Conte, P., & Alonzo, G. (2013). Environmental NMR: Fast-field-cycling Relaxometry. *eMagRes*, 2, 389-398. <http://dx.doi.org/10.1002/9780470034590.emrstm1330>
- Frantsuzov, I. (2010). *An NMR Relaxometry Study of Heteronuclear Effects Upon Proton Transfer in Hydrogen Bonds*. Thesis for the degree of Doctor of Philosophy, the University of Nottingham, the United Kingdom.
- Haber-Pohlmeier, S., Stapf, S., Dusschoten, D., & Pohlmeier, A. (2013). Relaxation in a Natural Soil: Comparison of Relaxometric Imaging, T1-T2 Correlation and Fast-Field Cycling NMR. *The Open Magnetic Resonance Journal*, 3, 57-62. <http://dx.doi.org/1874-7698/10>
- Kemalov, A. F., & Kemalov, R. A. (2013). Structural and Dynamic Studies of Naphtha Crude Residue with Different Chemical Nature. *World Applied Sciences Journal (Special Issue on Techniques and Technologies)*, 22, 16-22. <http://dx.doi.org/10.5829/idosi.wasj.2013.22.tt.22147>
- Kemalov, A. F., Kemalov, R. A., & Valiev, D. Z. (2013). Study of the structure of complex structural units of heavy oil from Zyuzeevskaya field by NMR relaxometry and rheological studies. *Oil Industry*, 2, 63-65. ISSN: 00282448
- Kemalov, R. A., Kemalov, A. F., & Valiev, D. Z. (2012). Thermodynamics of viscous flow activation and structural-dynamic analysis of high-viscosity oil with ultrasonication. *Oil industry*, 12, 100-103.

- Kimmich, R., & Anoardo, E. (2004). Field-cycling NMR relaxometry. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 44, 257-320. <http://dx.doi.org/10.1016/j.pnmrs.2004.03.002>
- Kruk, D., Herrmann, A., & Rössler, E. A. (2012). Field-cycling NMR relaxometry of viscous liquids and polymers. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 63, 33-64. <http://dx.doi.org/10.1016/j.pnmrs.2011.08.001>
- Nascimento, A. M. R., Tavares, M. I. B., & Miguez, E. (2004). Evaluation of Sorva Latex by ^1H NMR Relaxometry. *Annals of magnetic resonance*, 3(3), 82-86. Retrieved from <http://www.auremn.org.br/Annals/2004-vol3-num3/2004-3-82-86.pdf>
- Twieg, M. (2013). Open source NMR relaxometry platform. Unpublished dissertation in partial fulfillment of the requirements for the degree of Master of Science, Case Western Reserve University, Cleveland, USA.
- Washburn, K. E., & Birdwell, J. E. (2013). *Enhanced Laboratory Methods for Shale Analysis using Low Field NMR Relaxometry*. Abstract archive of the Geoconvention-2013: Integration. Retrieved from http://geoconvention.org/archives/2013abstracts/214_GC2013_Enhanced_Laboratory_Methods.pdf
- Xu, F., Leclerc, S., & Canet, D. (2013). NMR relaxometry study of the interaction of water with a nafion membrane under acid, sodium, and potassium forms. Evidence of two types of bound water. *The journal of physical chemistry B*, 117(21), 6534–6540. <http://dx.doi.org/10.1021/jp311062h>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

Automatic Identification of *Ficus deltoidea* Jack (Moraceae) Varieties Based on Leaf

A Fakhri A Nasir¹, M Nordin A Rahman¹, Nashriyah Mat² & A Rasid Mamat¹

¹ Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Tembila Campus, 22200 Besut, Terengganu

² Faculty of Agriculture, Biotechnology and Food Science, Universiti Sultan Zainal Abidin, Tembila Campus, 22200 Besut, Terengganu

Correspondence: M Nordin A Rahman, Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Tembila Campus, 22200 Besut, Terengganu. Tel: 6-096-993-207. E-mail: mohdnabd@unisza.edu.my

Received: June 29, 2014

Accepted: July 4, 2014

Online Published: August 17, 2014

doi:10.5539/mas.v8n5p121

URL: <http://dx.doi.org/10.5539/mas.v8n5p121>

The research is financed by the Ministry of Education Malaysia.

Abstract

Currently, the traditional method used to identify *Ficus deltoidea* Jack (Moraceae) varieties require the plant taxonomists to observe and examine the leaf morphology of herbarium or live specimens. An automated variety identification system would ease the herbs collector to carry out valuable plant identification work. In this paper, a model for *F. deltoidea* varieties identification based on their leaf shape, color and texture was developed. Five different varieties of *F. deltoidea* were used in the proposed work with sixty nine sample data collected for each of varieties. First, the *F. deltoidea* leaves were plucked and the picture of leaves is then taken by a digital scanner in the format of JPEG. For leaf shape, a total of fourteen shape features were extracted based on basic geometric features. The mean of different color channels was calculated in leaf color feature extraction. Furthermore, four texture features based on gray-level co-occurrence matrix was implemented to extract leaf texture properties. By using the leaf structure, a set of three different leaf properties which are leaf shape, color and texture features was extracted. The features weight is then calculated using eigenvalues coefficient in principal component analysis. The best principal components are retained for identification experiments. Lastly, Nearest Neighbor with Euclidean distance was used in variety identification based on three different leaf properties mentioned above. The effectiveness of different leaf features are demonstrated in the identification experiment.

Keywords: *Ficus deltoidea* Jack, plant leaf identification, content-based image retrieval, pattern recognition, image processing

1. Introduction

Herb plants have been used through much of human history as sources of food or food additive, medicine, beauty enhancers used in cosmetic ingredients and fragrances. The use of herbs as medicine has a long history, starting from the Greek civilization in the West and the Arabic, Chinese and Indian civilizations in the East (Ramlan, 2003). Herbs as traditional medicine is still widely practiced today and in modern medicine, it is recognizes as a form of alternative medicine. Many allopathic medicines, which are produced synthetically, also derived from plants such as quinine for malaria, quinidine for heart arrhythmia from *Cinchona* spp and digoxin for heart failure from *Digitalis* spp (Ramlan, 2003).

Ficus deltoidea Jack (Moraceae) or mistletoe fig (common name) or locally known as 'Mas Cotek' among the Malays, is recognized as Malaysian herb plants (Hean et al., 2011). *F. deltoidea* is a native and widely distributed throughout Malaysia, Thailand, Sumatra, Java, Kalimantan, Sulawesi, and Moluccasa (USDA, 2007). In Malay traditional medicine, the dried *F. deltoidea* leaves are marketed as an herbal tea. The decoction of the leaves is believed to improve blood circulation and have aphrodisiac activity and antioxidant and antidiabetic properties (Norhaniza et al., 2007; Sulaiman et al., 2008; Adam et al., 2011). Moreover, other approaches revealed that *F. deltoidea* leaves is a potential alternative medicine to strengthen the uterus for women after birth (Sulaiman et al.,

2008) and effective on enhancement of wound healing (Abdulla et al., 2010).

Figs or *Ficus* plants originated in Asia Minor and can be found throughout Mediterranean, Indian subcontinent, Latin America, Texas, Southern California, until the Far East such as in the Malaysian tropical rain forest (USDA, 2007). It is one of the largest genera of flowering plants with six traditional subgenera that are recognized based on morphology and distribution (Lansky & Paavilainen, 2011). One of the species that included in this genus is *Ficus deltoidea* Jack (Latin name). Seven out of 15 recognized varieties of *F. deltoidea* found in Peninsular Malaysia namely var. *deltoidea*, var. *angustifolia*, var. *trengganuensis*, var. *bilobata*, var. *intermedia*, var. *kunstleri* and var. *motleyana* (Kochummen, 1978; Turner, 1995). *F. deltoidea* is a large shrub or small tree with aerial roots often begins its life as an epiphyte plant. It is also cultivated in various parts of the world as a houseplant or as an ornamental shrub. The name 'Mas Cotek' is given in Peninsular Malaysia because there are fine spots with gold in colour on the surface of each leaves. Generally, the scopes of the study can be specified as modeling an automatic identification for *F. deltoidea* varieties recognition.

2. Problem Statements

Traditionally, herbs have been collected for use from wild areas (Ramlan, 2003). For large scale production, this presents an unfeasible solution due to the lack of reliable supply. Continuous harvesting of wild sources will lead to species endangerment and extinction. Agricultural cropping of herbs as specialty crops can ensure high quality herbal by correct herb identification. Correct herb identification is essential as many herbal species come in many varieties which have similar appearances but have different usages in traditional medicine. For example, *F. deltoidea* var. *deltoidea* and *F. deltoidea* var. *angustifolia* are two of *Ficus deltoidea* Jack (Moraceae) varieties which have more or less similar morphological appearances in the eye of amateur herbs collector. However, according to study made by Nihayah et al. (2012), these two varieties have different mineral content in leaf, whereby *F. deltoidea* var. *deltoidea* contain high amount of magnesium, manganese and potassium compared to *F. deltoidea* var. *angustifolia*. As a conclusion, *F. deltoidea* var. *deltoidea* leaf is more suitable to be served as a tea drink.

Based on the importance of in herb plant to medical practitioners, recent development especially for *Ficus deltoidea* Jack (Moraceae) indicates the need to provide a computer-based identification system in order to get correct identification. Currently, the traditional method used to identify *F. deltoidea* have lots of disadvantages which leads to time consuming, less efficient, troublesome task for nonprofessional and sometimes identification task is too complex. In recent years, the existence of sophisticated technologies such as digital cameras and portable computers has lead to an increasing interest in automating the process of plant species identification. Although there are various types of computer algorithms and databases for plant species identification has been built to speed up taxonomists work, the computer algorithms and databases in herb plant species are very hard to find. In case of *F. deltoidea* plant, there is no image database found in the literature. Plant species identification for herb plant becomes one of the main challenging tasks in computer vision due to lack of proper models or representations. As stated in detail reviews of plant species identification research by Cope et al. (2012), different features are often needed to distinguish different categories of plant. For example, whilst leaf shape may be sufficient to distinguish between some species, other species may have very similar leaf shapes to each other, but have different colored leaves. No single feature, or kind of feature, may be sufficient to separate all the categories, making feature selection a challenging problem (Cope et al., 2012). For this reason, research to get a right model to identify the herbs plant species is essential. This paper objective is to emphasizes the approach of automatic identification model for *Ficus deltoidea* Jack (Moraceae) varieties. Three feature selection approaches which are leaf shape, leaf color and leaf texture are used and explained in the next section.

3. The Model

Many models have been used for plant species identification that most of them follow the same general steps in content-based image retrieval. The process started with image acquisition, followed by preprocessing, segmentation and features extraction. Finally, the identification or commonly known as classification process will look up into a database to find the similarity of the query leaf image. According to Ab Jabal et al. (2013), two important consideration has to be critically made in developing plant species identification which are in features extraction and classification phase. Other phase considered as less important because it is reflected on image acquisition technique. Almost all previous image acquisition approaches are implemented in the same manner. For example Lee & Chen (2006) and Du et al. (2007) used a digital scanner for leaf image capturing and therefore preprocessing and segmentation process done in the same manner.

Leaf shape is a prominent feature that most people use to recognize and classify a plant (Hossain & Amin, 2010). Wu et al. (2007), Hossain & Amin (2010), Kadir et al. (2011) and Zulkifli et al. (2011) had stated that diameter,

physiological length, physiological width, leaf area and perimeter are basic geometry information can be extract from the leaf shape. In addition, leaf color (using mean of RGB color channel), textures (using contrast, correlation and homogeneity) and vein are also considered as important features (Kadir et al., 2011). All these features are useful for recognition of leaf image. For this case study, vein features are not used due to the technique of image acquisition. By using a digital scanner, the leaf vein structures are not clearly seen. A light bank is required to get the clearer vision of leaf vein.

The features extraction process for *F. deltoidea* varietal identification is referred from the works (Lee & Chen, 2006; Wu et al., 2007; Hossain & Amin, 2010; Zulkifli et al., 2011; Fotopoulou et al., 2013; Kadir et al., 2011; Najjar & Zagrouba, 2012). The used of Principal Component Analysis (PCA) and Nearest Neighbor (NN) are referred from (Lee & Chen, 2006; Du et al., 2007; Wu et al., 2007; Kadir et al., 2012). Figure 1 shows the block diagram for the *F. deltoidea* varietal identification model.

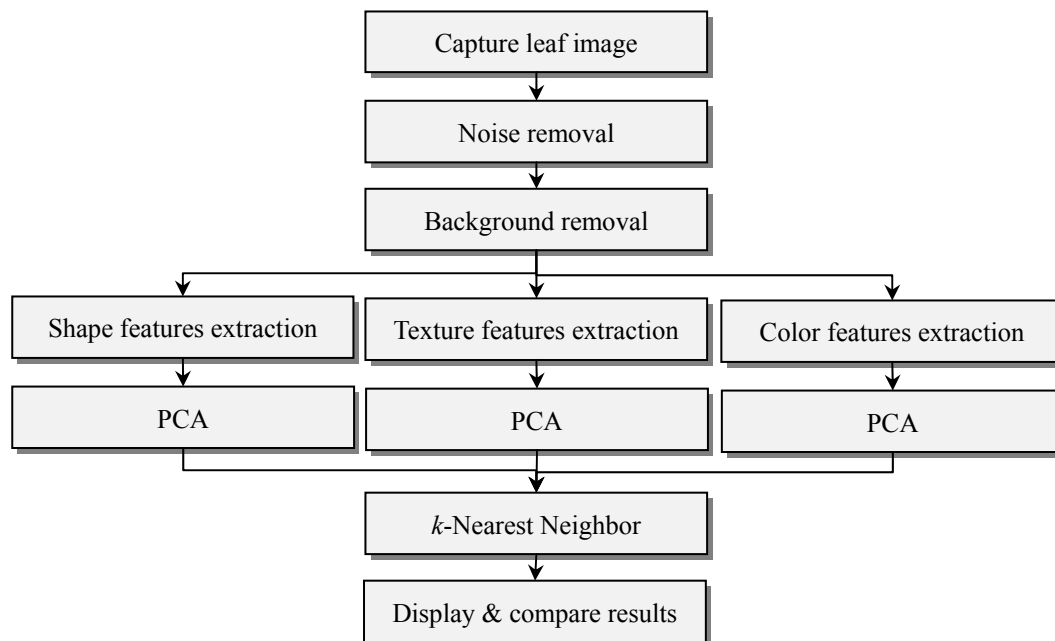


Figure 1. Block diagram of the *F. deltoidea* varietal identification model

3.1 Capture Leaf Image

The current study was carried out on five varieties only excluded for var. *montleya* and var. *intermedia* due to unavailability of these two varieties in the Germplasm Living Collection at the Tembila Campus of Universiti Sultan Zainal Abidin. Since there is no *F. deltoidea* leaf image available elsewhere in literature, the leaf image database used in the following work was collected and built by ourselves. Firstly, the leaves from each plant are plucked. The single leaf was placed on a white A4 size paper, and then the picture of the leaf was taken with a digital scanner. This process generates a standard leaf image in JPEG format with the white background. For this time, a total of 345 *F. deltoidea* leaf images that consisted of 69 leaf images for each of variety have been collected. The representative sample images for the five varieties are shown in Figure 2.

3.2 Noise Removal

Digital images are prone to a variety of types of noise. Noise is the result of errors in the image acquisition process that result in pixel values that do not reflect the true intensities of the real scene. In this case, the image is acquired directly in a digital format. The leaf is inserted to digital scanner may have some dirt that come out from the leaf itself. In order to remove noise median filtering with 3x3 neighborhood is used. Median filtering is similar to an averaging filter, in that each output pixel is set to an average of the pixel values in the neighborhood of the corresponding input pixel.

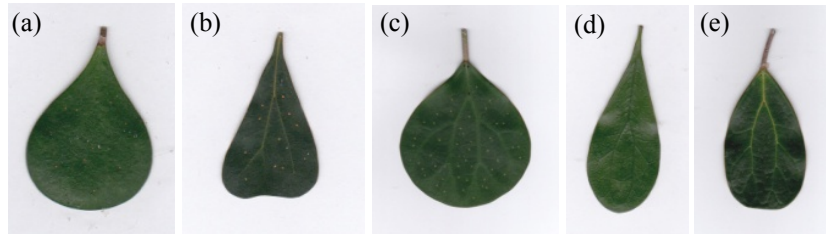


Figure 2. Variety of *F. deltoidea* in the leaf database, (a) var. *deltoidea*, (b) var. *bilobata*, (c) var. *kunstleri*, (d) var. *angustifolia*, and (e) var. *trengganuensis*

3.3 Background Removal

As mentioned before, the captured of *F. deltoidea* leaf image gives an image of leaf with an almost white background with interference by scanner lighting (see Figure 2). In order to remove the white background, segmentation process by using thresholding method based on histogram is used. First, the RGB leaf image is converted to grey-level image by the following equation:

$$\text{Grey} = 0.299 * R + 0.578 * G + 0.114 * B \quad (1)$$

Where; R, G, B correspond to the color of the pixel, respectively. The gray-scale image for plant leaf is shown in Figure 4(b).

The level to convert grayscale into binary was determined according to grayscale histogram. We accumulated the grayscale pixel values for 345 leaves and divided them by 345, the number of leaves. The average histogram to grayscale of 345 leaf images is shown as Figure 3. There are two peaks in the grayscale histogram. The left peak that refers to pixels was consisted of the leaf while the right peak that refers to pixels consisted of the white background (with interference by scanner lighting). The lowest peak between left peak and right peaks is approximately the value of 160 on the average. Therefore, the level of 0.6275 (160/255) is applied. The output image replaced all pixels in the input image with greater than the level by the value of 1 and replaced all other pixels by the value of 0.

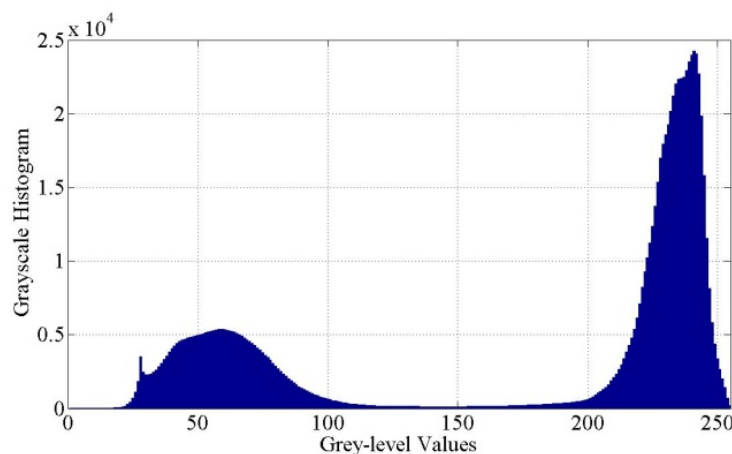


Figure 3. Average grey-level values for 345 leaf images

After image thresholding, some defects might be presented in the leaf images for example some parts of objects might be misclassified as background and some small regions of background might be mistakenly segmented as objects. Consequently, image opening and image closing of mathematical morphology are used for processing image segmented by thresholding. Image opening involves reserving the unsegmented parts of objects using first image dilation, by merging neighboring pixels of an object into the object and the image erosion, by removing boundary pixels of objects. On the contrary, image closing is image erosion followed by image dilation in order to eliminate the unsegmented parts of background. Let E be a Euclidean distance space or an integer grid, A is a binary image in E where B_z is the translation of B by a vector z , and B^s denotes the symmetric of B . The

mathematical definition of dilation for binary images is defined as follows:

$$A \oplus B = \{z \in E | (B^s) \cap A \neq \emptyset\} \quad (2)$$

The erosion of the binary image A is defined by:

$$A \ominus B = \{z \in E | B_z \subseteq A\} \quad (3)$$

The opening of a set (binary image) A by a structuring element B is the dilation followed by the erosion of that set which is:

$$A \cdot B = (A \ominus B) \oplus B \quad (4)$$

The closing of a set (binary image) A by a structuring element B is the erosion followed by the dilation of that set which is:

$$A \cdot B = (A \oplus B) \ominus B \quad (5)$$

Notice that there existed some variances on length of leaf petiole. To keep the precision of leaf shape features extraction, these leaf petioles should be further removed from the obtained binary images. Here, the process of image closing operation to binary images is considered. By performing the closing operation with proper structuring element, the leaf petioles removal can be achieved successfully while preserving the main shape characteristics of leaf objects. The result of leaf petiole removal of binary image is shown in Figure 4(e).

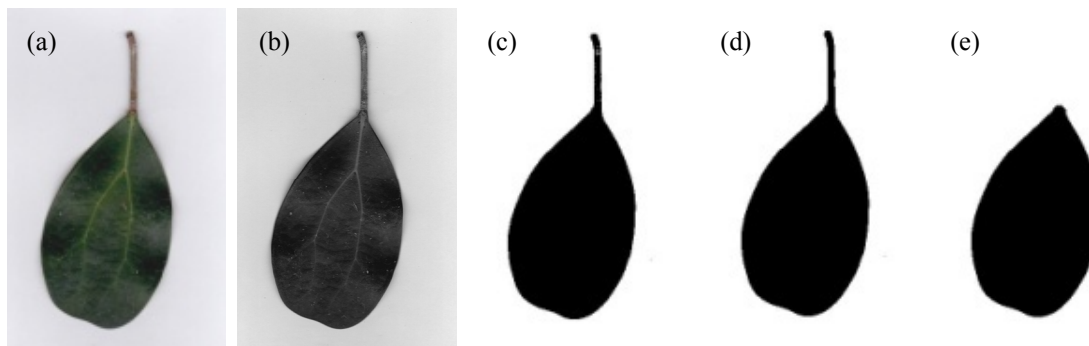


Figure 4. The processed leaf image (example on var. *tregganuensis*), (a) original image, (b) gray image, (c) binary image, (d) smooth binary image, and (e) leaf petiole removal

The leaf in Figure 4(e) is then used for shape features extraction. For color and texture features extraction, the leaf in Figure 4(e) is also used but the black region which is leaf region is converted back to the original RGB color.

3.4 Shape Features Extraction

For efficient shape feature extraction, a *regionprops* function in Matlab is used which calculates the region properties. In this paper, 5 basic region features; area, convex area, perimeter, major axis length, minor axis length and diameter are extracted using this function. Another nine features are then derived from the previous five basic features. In total, 14 digital features are extracted for each of the leaf image.

- 1) Area: The actual number of pixels in the region. It is denoted as A.
- 2) Convex Area: The area of convex hull of the region. The convex area is greater than or equal to area of the region. It is denoted as CA.
- 3) Major Axis Length: Major Axis Length is defined as the length (in pixels) of the major axis of the ellipse that has the same normalized second central moments as the region. It is denoted as MAL.
- 4) Minor Axis Length: Minor Axis Length is defined as the length (in pixels) of the minor axis of the ellipse. It is denoted as MIL.
- 5) Perimeter: The distance around the boundary region. It is calculated by counting the distance between each adjoining pair of pixels around the border of the region. It is denoted as P.

Based on 5 basic features introduced previously, another 9 features are defined for leaf recognition.

- 6) Diameter: Scalar that specifies the diameter of a circle with the same area as a region. It is computed as $4A/\pi$, where A is the leaf area.
- 7) Solidity: This feature is derived from convex hull and defined as A/CA , where A is the area of the leaf and CA is the area of convex hull.
- 8) Roundness: The roundness or circularity ratio is defined as A/P^2 , where A is the leaf area and P is the leaf perimeter.
- 9) Aspect Ratio: The aspect ratio is a ratio between the major axis length and the minor axis length, MAL/MIL .
- 10) Rectangularity: This feature describes the similarity between a leaf and a rectangle. It is defined as $MAL*MIL/A$, where MAL is major axis length, MIL is minor axis length and A is the leaf area.
- 11) Form Factor: It is defined as $4\pi A/P^2$, where A is the leaf area and P is the leaf perimeter. This feature is used to describe the different between a leaf and a circle.
- 12) Eccentricity: This feature is used to differentiate the rounded leaf and the long one. The eccentricity is defined as MIL/MAL , where MIL is the minor axis length of the leaf and MAL is the major axis length of the leaf.
- 13) Perimeter Ratio to Diameter: It is represented as the ratio of the leaf perimeter P and leaf diameter D. This is calculated by P/D .
- 14) Perimeter Ratio of Major Axis Length and Minor Axis Length: This feature is defined as the ratio of leaf perimeter P and the sum of major axis length and minor axis length, $P/(MAL+MIL)$.

3.5 Color Features Extraction

The mean or average for each of color channel value is extracted in this process. Four color channels used in this implementation consists of RGB, HSV, Lab and YIQ. Mean of grayscale value is also used. First, the leaf in Figure 4(e) is converted back to the original RGB color. Then, A colormap functions in Matlab is used to convert the original RGB leaf image into each of color channels mentioned. The mean for each color channel is calculated. For example, in RGB color, mean for Red, Green and Blue values are calculated separately and this process continues to the other color channels. In this way, 13 color features are extracted which are mean value of grayscale, R, G, B, H, S, V, Y, I, Q, L, a and b.

3.6 Texture Features Extraction

A statistical method of examining texture that considers the spatial relationship of pixels is the gray-level co-occurrence matrix (GLCM), also known as the gray-level spatial dependence matrix. The GLCM functions characterize the texture of an image by calculating how often pairs of pixel with specific values and in a specified spatial relationship occur in an image, creating a GLCM, and then extracting statistical measures from this matrix. In Matlab, graycoprops function uses the normalized GLCM to calculate contrast, correlation, energy and homogeneity. These four properties are extracted for texture features. Firstly, the leaf in Figure 4(e) is converted back to the original RGB color. The image is then converted to grey-level image. By using default graycomatrix function, GLCM matrix is then created. Lastly, by using the normalized GLCM, four texture features are finally extracted.

3.7 Principal Component Analysis

Principal Component Analysis is a statistical method to reduce the dimension of variables. It is useful when implemented on a large number of variables with possibly having some redundancy in those variables. In this case, redundancy means that some of the variables are correlated with one another. Because of this redundancy, it should be possible to reduce the observe variables into a smaller number of principal components that will account for most of the variance in the observed variables (O'Rourke et al., 2005).

Technically, a principal component can be defined as a linear combination of optimally-weighted observed variables. The general form for the formula to compute scores on the first component extracted in a principal component analysis is as below:

$$C_1 = b_{11}(X_1) + b_{12}(X_2) + \dots + b_{1p}(X_p) \quad (6)$$

where; C_1 is the subject's score on principal component (the first component extracted), b_{1p} is the regression coefficient (or weight) for observed variable p , as used in creating first principal component, and X_p is the subject score on observed variable p . To calculate the regression coefficient (b_{1p}), there are other four steps must be completed in PCA. 1) Calculate mean for all of the data, 2) standardized the data by subtracting the data with mean, 3) calculate the covariance of the data and 4) calculate the eigenvalues and the eigenvectors based on the

covariance matrix. The coefficient is then obtained from the eigenvalues. Finally, the coefficient, b_{1p} is then placed in above equation.

The number of components extracted is equal to the number of variables being analyzed. In general, the first few components will account for meaningful amounts of variance and the later components will tend to account for only trivial variance. Therefore, the next step is to determine how many meaningful components should be retained for classification. Four criteria may be used in making this decision which is eigenvalue-one criterion, the scree test, the proportion of variance accounted and the interpretability criterion (O'Rourke et al., 2005). In this study, the scree test is considered to make this decision. Figure 5 shows the scree test plot for shape, color and texture features.

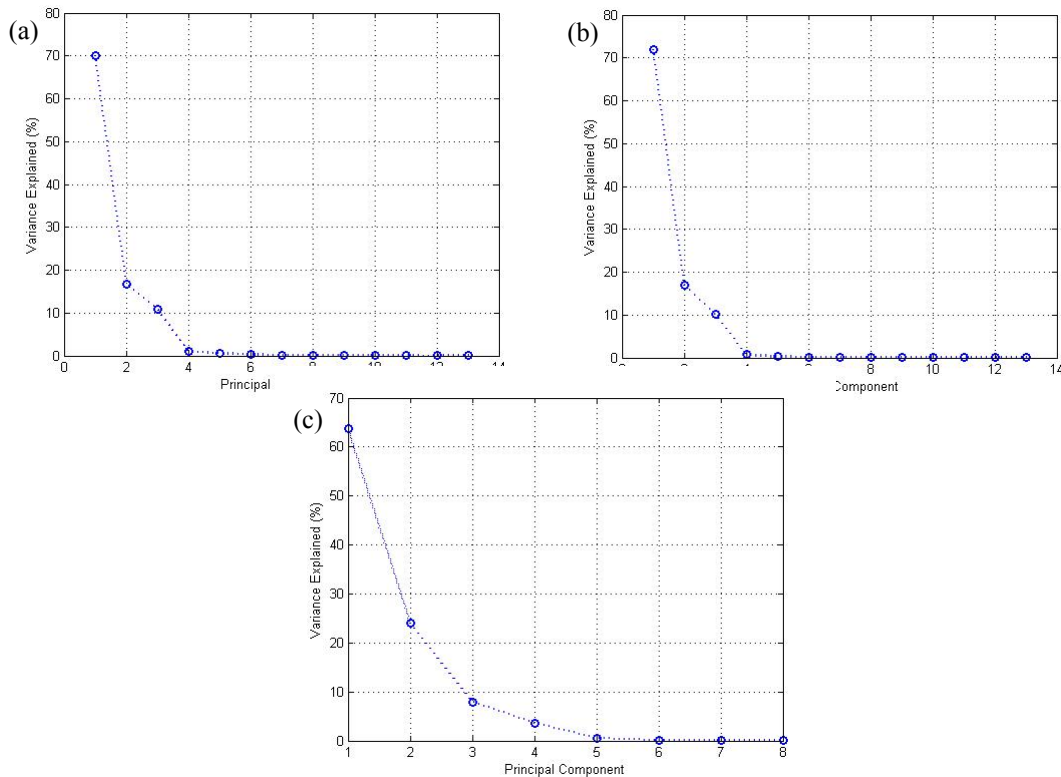


Figure 5. The scree test plot, (a) shape features, (b) color features, and (c) texture features

With the scree test, the eigenvalues (variance explained) associated with each component is plotted. To choose the few components, (Kaiser, 1960) recommend readers to find for a “break” between the components with relatively large eigenvalues and those with small eigenvalues. The components that appear before the break are assumed to be meaningful and are retained for next process. For this case, the “break” is appeared at principal component 3 for all of plots (see Figure 5), therefore only the principal component 1, principal component 2 and principal component 3 should be retained. In this work, principal component 1, principal component 2, principal component 3, single original features and combined original features for each of shape, texture and color are used in the identification experiments.

3.8 Nearest Neighbor

The Nearest Neighbors algorithm is a method for identifying objects based on closest training examples in the feature space. The Nearest Neighbors (NN) classification divides data into a test set and a training set. For each row of the test set, the k nearest (in Euclidean distance) training set objects are found, and the identification is determined by majority vote. In Cartesian coordinates, if $p = (p_1, p_2, \dots, p_n)$ and $q = (q_1, q_2, \dots, q_n)$ are two points in Euclidean n -space, then the distance from p to q , or from q to p is given by:

$$d(p, q) = d(q, p) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \quad (7)$$

4. Experimental Results

In the experiment, the performance comparison of the k -NN classifier with the nearest neighbor (3-NN), (5-NN) and (7-NN) are executed. The odd number is used in order to avoid ties result. When using an even number, for example $k=2$, let say the two closet distance falls into class 1 and class 2. Therefore, the result for this example is ties whereby 50% falls into class 1 and 50% falls in class 2. Thus, even number is not used in this experiment. To each variety of *F. deltoidea*, all leaves image are pickup from database. If one image is selected as a query image, the rest will be testing image samples. In other words, this process namely leave-one-out cross-validation testing. In each of leaf features properties, the average recognition accuracy is calculated for each of test images. Figure 6, Figure 7 and Figure 8 show the recognition accuracy for shape, color and texture features respectively.

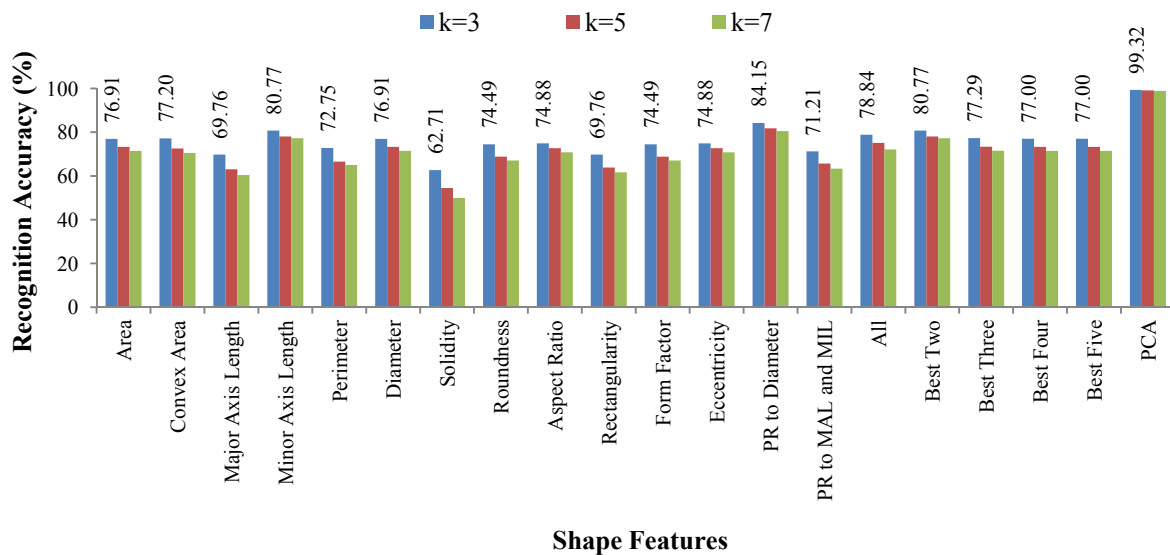


Figure 6. Average recognition accuracy for shape features

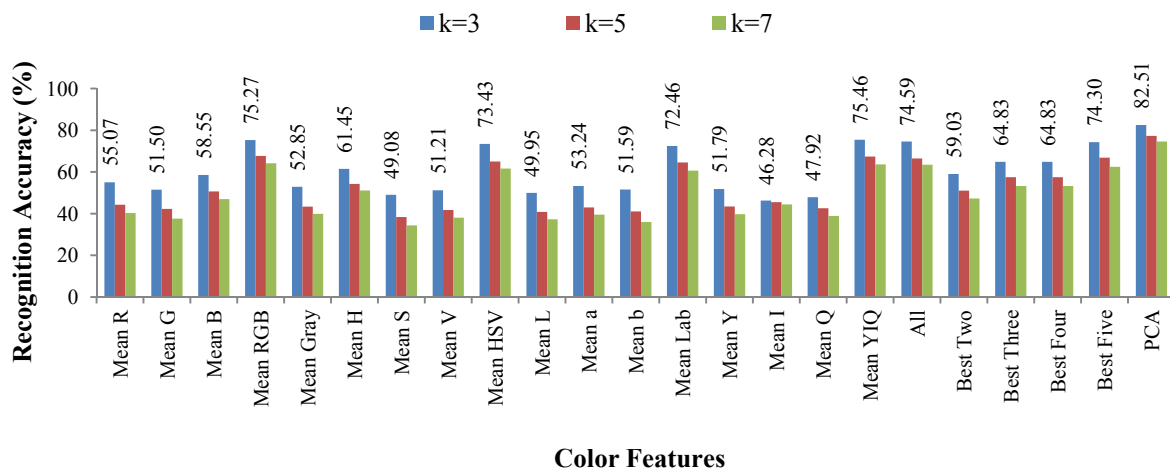


Figure 7. Average recognition accuracy for color features

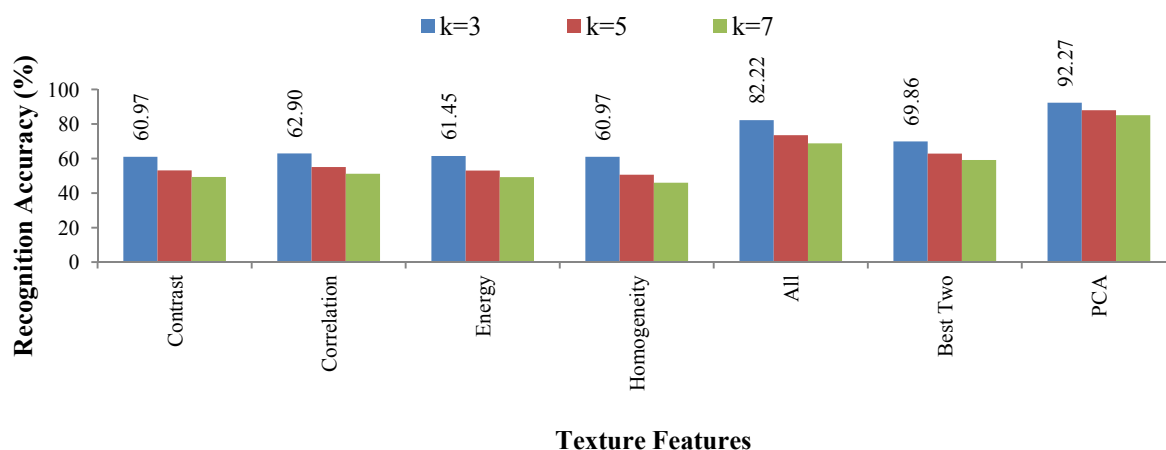


Figure 8. Average recognition accuracy for texture features

Based on Figure 6, Figure 7 and Figure 8, NN with $k=3$ is the best for all features. In shape features, perimeter ratio to diameter (84.15%) is the best feature to distinguish five varieties of *F. deltoidea* followed by minor axis length (80.77%). The recognition accuracy for combination of all shape features, the best two, best three, best four and best five features showed no continuous increment when adding more combined features. The result for PCA is the highest (99.32%) with approaching to 100%. As compared to perimeter ratio to diameter, the use of PCA is able to increase the performance using shape features for about 15.27%.

According to Figure 7, mean H (61.45%) is the best single color features followed by mean B (58.55%) and mean R (55.07%). From this result, combinations of color mean features showed better performance than single feature. For example, the mean for R is 55.07%, G is 51.50% and B is 58.55%. When these three features are combined (mean RGB), the identification accuracy becomes more efficient. As the same to shape features, combination of all color features, the best two, best three, best four and best five features also showed no continuous increment when adding more combined features. The result for PCA is the highest which is 82.15%. As compared to mean YIQ feature, the use of PCA is able to increase the performance using color features for about 8.14%.

The correlation (62.90%) showed the best recognition accuracy for texture analysis. Combination of all texture features is better than using single texture feature. PCA is the best results in this features which boost up 10.87% recognition accuracy compared to the best original features. For clearer view on experimental results, the best recognition accuracy is summarized in Table 1.

Table 1. Summary of experimental results

Feature	The best original	PCA
Shape	(Perimeter ratio to Diameter) 84.15%	99.32%
Color	(mean YIQ) 75.46%	82.51%
Texture	(All) 82.22%	92.27%

5. Conclusion

This paper introduces an automatic *F. deltoidea* varietal identification model based on leaf shape, leaf color and leaf texture which replaced the traditional method that has many disadvantages. The experimental results showed that the use of leaf shape is much better to distinguish five varieties of *F. deltoidea*. In addition, the use of PCA is able to boost up the recognition accuracy. Finally, the use of 3-NN is better for this dataset. In order to generalize the model, the future works will focus on collecting more leaf images from var. *montleya* and var. *intermedia* to complete the seven *F. deltoidea* varieties that found in Malaysia. Finally, the model should be extended to include a graphical user interface to ease user.

Acknowledgments

The authors duly acknowledged the permission of the Faculty of Agriculture, Biotechnology and Food Sciences Universiti Sultan Zainal Abidin (UniSZA), Besut, Malaysia for allowing us to pick and capture leaves of *F. deltoidea* at the Faculty Nursery in the Tembila Campus. Acknowledgement was also given to the nursery staff of the Faculty of Agriculture, Biotechnology and Food Sciences for their continuous support.

References

- Ab Jabal, M.F., Hamid, S., Shuib, S., & Ahmad, I. (2013). Leaf features extraction and recognition approaches to classify plant. *Journal of Computer Science*, 9(10), 1295-1304. <http://dx.doi.org/10.3844/jcssp.2013.1295.1304>
- Abdulla, M. A., Ahmed, K. A. A., Abu-Luhoom, F. M., & Muhanid, M. (2010). Role of *Ficus deltoidea* extract in the enhancement of wound healing in experimental rats. *Biomedical Research*, 21(3), 241-245.
- Adam, Z., Ismail, A., Khamis, S., Mokhtar, M. H. M., & Hamid, M. (2011). Antihyperglycemic activity of *F. deltoidea* ethanolic extract in normal rats. *Sains Malaysiana*, 40(5), 489-495.
- Cope, J. S., Corney, D., Clark, J. Y., Remagnino, P., & Wilkin, P. (2012). Plant species identification using digital morphometrics: A review. *Experts Systems with Applications*, 39(8), 7562-7573. <http://dx.doi.org/10.1016/j.eswa.2012.01.073>
- Du, J. X., Wang, X. F., & Zhang, G. J. (2007). Leaf shape based species recognition. *Applied Mathematics and Computation*, 185(2), 883-893. <http://dx.doi.org/10.1016/j.amc.2006.07.072>
- Fotopoulou, F., Laskaris, N., Economou, G., & Fotopoulos, S. (2013). Advanced leaf image retrieval via Multidimensional Embedding Sequence Similarity (MESS) method. *Pattern Anal. Applic*, 16(3), 381-392. <http://dx.doi.org/10.1007/s10044-011-0254-6>
- Hean, C. O., Rosnaini, M. Z., & Pozi, M. (2011). Traditional knowledge of medicinal plants among the Malay villagers in kampung Mak Kemas, Terengganu, Malaysia. *Studies on Ethno-Medicine*, 5(3), 175-185.
- Hossain, J., & Amin, M. A. (2010). Leaf shape identification based plant biometrics. *Proceedings of the Computer and Information Technology*, 458-463. <http://dx.doi.org/10.1109/ICCITECHN.2010.5723901>
- Kadir, A., Nugroho, L.E., Susanto, A., & Santosa, P. I. (2011). A comparative experiment of several shape methods in recognizing plants. *Int. J. Comput. Sci. Inform. Technol.*, 3(3), 256-263. <http://dx.doi.org/10.5121/ijcsit.2011.3318>
- Kadir, A., Nugroho, L. E., Susanto, A., & Santosa, P. I. (2012). Performance improvement of leaf identification system using principal component analysis. *International Journal of Advanced Science and Technology*, 44, 113-124.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20(1), 141-151. <http://dx.doi.org/10.1177/001316446002000116>
- Kochummen, K. M., & Rusea, G. (2000). *Tree Flora of Sabah and Sarawak* (Vol. 3, pp. 181-334).
- Lansky, E. P., & Paavilainen, H. M. (2011). *Figs: The Genus Ficus*. London New York: CRC Press Boca Raton.
- Lee, C. L., & Chen, S. Y. (2006). Classification of leaf images. *International Journal of Imaging Systems and Technology*, 16(1), 15-23. <http://dx.doi.org/10.1002/ima.20063>
- Najjar, A., & Zagrouba, E. (2012). Flower image segmentation based on color analysis and a supervised evaluation. *Proceedings of the Communications and Information Technology*, 397-401. <http://dx.doi.org/10.1109/ICCITechnol.2012.6285834>
- Nihayah, M., Yong, K. W., & Nur Faizah, A. B. (2012). Determination of mineral content in the *Ficus deltoidea* leaves. *Jurnal Sains Kesihatan Malaysia*, 10(2), 25-29.
- Norhaniza, A., Sin, C. Y., Chee, E. S., Nee, K. I., & Renxin, L. (2007). Blood glucose lowering effect of *Ficus deltoidea* aqueous extract. *Malaysian Journal of Science*, 26(1), 73-78.
- O'Rourke, N., Hatcher, L., & Stepanski, E. J. (2005). *A step-by-step approach to using SAS for univariate & multivariate statistics*. SAS Publishing.
- Ramlan, A. A. (2003). Turning Malaysia into a global herbal producer: A personal perspective. Retrieved from http://www.penerbit.utm.my/syarahana/pdf/09/siri9_teks.pdf
- Rencher, A. C. (2002). *Methods of Multivariate Analysis*. Wiley-Interscience.

- Sulaiman, M. R., Hussain, M. K., Zakaria, Z. A., Somchit, M. N., Moin, S., Mohamad, A. S., & Israf, D. A. (2008). Evaluation of the antinociceptive activity of *Ficus deltoidea* aqueous extract. *Fitoterapia*, 79(7-8), 557-561. <http://dx.doi.org/10.1016/j.fitote.2008.06.005>
- Turner, I. M. (1995). Catalogue of the vascular plants in Malaya. *Garden's Bulletin Singapore* (Vol. 47, pp. 347-757). Singapore Botanic Gardens.
- USDA (2007). *ARS, National Genetic Resources Program, Germplasm Resources Information Network – (GRIN) Database*. Retrieved from <http://ars-grin.gov/cgi-bin/npgs/html/taxon.pl?16826>
- Wu, S. G., Bao, F. S., Xu, E. U., Wang, Y. X., Chang, Y. F., & Xiang, Q. L. (2007). A leaf recognition algorithm for plant classification using probabilistic neural network. *Proceedings of the Signal Processing and Information Technology*, 11-16. <http://dx.doi.org/10.1109/ISSPIT.2007.4458016>
- Zulkifli, Z., Puteh, S., & Mohtar, I. A. (2011). Plant leaf identification using moment invariants and general regression neural network. *Proceedings of the Hybrid Intelligent Systems*, 430-435. <http://dx.doi.org/10.1109/HIS.2011.6122144>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

The Extent of Sunlight Penetration Performance on Traditional Style's Apartment Façade in Putrajaya, Malaysia

Ahmad Sanusi Hassan¹ & Yasser Arab¹

¹ School of Housing, Building and Planning, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia

Correspondence: Ahmad Sanusi Hassan, School of Housing, Building and Planning, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia. Tel: 60-19-506-8260. E-mail: sanusi.usm@gmail.com

Received: July 10, 2014 Accepted: July 15, 2014 Online Published: August 27, 2014

doi:10.5539/mas.v8n5p132

URL: <http://dx.doi.org/10.5539/mas.v8n5p132>

The research is financed by Universiti Sains Malaysia.

Abstract

In this study the analysis is to measure the performance on the extent of sunlight penetration on front apartment facades. The apartments are located in Putrajaya, Malaysia. Putrajaya is the administrative city for the federal government, and it is considered the latest new city in Malaysia with postmodern style buildings which exhibit a range of complex geometric elements blending with colonial, modern and traditional architectural style. The scope of this study is to measure the extent of sunlight penetration on the apartment facades. Two front facades of the apartments were selected in the case studies using computer simulation for the analysis. The selected case studies are sixteen stories apartment located in Precinct 17 and seventeen stories apartment located in Precinct 18. SunTool software is used in the survey to calculate the extent of sunlight penetration. The survey will be conducted at a position when the sun path is perpendicular to the house façade. This analysis will compare results of the computer simulations of these two apartments in relation to the facade design of the architectural elements for shading devices. This study finds that the more the application is the traditional shading elements, the more is the shading performance on the façade walls. However this study also finds that traditional shading elements are designed only to block sunlight from horizontal angle of the sun position. They are not designed to obstruct sunlight from vertical angle of the sun position.

Keywords: sunlight, traditional shading elements, penetration, apartment, Malaysia

1. Introduction

1.1 Problem of Direct Sunlight Penetration

This study highlights problem of direct sunlight penetration to the interior area of apartment building and solar radiation on the apartment facade in tropical countries like Malaysia. Poor design integrated with shading elements which causes poor shading performance is the primary causes of the problem as the hypothesis of this study. This problem creates heat gain to the indoor area which causes low level on human comfort factor to the house occupants. The study on high rise apartment design is important because the building façade has a larger area with a tendency of exposure to direct sunlight compared to the building façade of low rise building (Ismail, 2002). As a result, design with excellent shading elements on the apartment façades is very critical. It is the objective in this study to verify the problem scientifically to dictate this design faults by measuring the extent of sunlight penetration on the apartment façade. Results of this study can be used as a guide for architects and those involved in the building industry to refer as issues to overcome solar radiation in apartment façade design. This reference provides awareness and guidelines for passive design with efficient usages of shading elements to avoid solar radiations in the tropical countries such as Malaysia (Bakhlah & Hassan, 2012).

1.2 Importance of the Problem

The measurement will be calculated using either manual or computer simulation. This study is limited to the apartment design in Putrajaya which is the latest new town built in Malaysia with post modern building styles. It epitomizes the present and future direction of housing industry in Malaysia. One of the popular post modern styles is integration of traditional Malay architectural style. It is the aim of this study to measure effectiveness of the integration of traditional Malay shading elements in post modern apartment which are roof overhang,

recessed wall, attached roof and balcony or corridor. House facade exposed to direct sunlight causes problem of solar radiation. The sun energy will radiate the heat from outside wall transmitted into the interior of the house (Hassan & Ramli, 2010) (Feriadi & Nyuk, 2004). It generates extra heat gains inside the house which causes warm temperature to the indoor area; as a result, it creates uncomfortable thermal condition to the occupants. A report by Centre for Environment, Technology and Development, Malaysia (CETDEM) (2005) on *Malaysian Urban Household Energy Consumption Patterns* stated that electricity consumption on air-conditioning and fan for cooling purpose is the highest with 32.4% of the total electric bill, which indicate problem in indoor heat gains due to unnecessary direct sunlight exposure to the houses in Malaysia. A study on high rise apartment is crucial as apartments in Putrajaya are the most popular house types which represent 73% of the total unit houses (Department of Statistics, Malaysia, 2010).

1.3 Relevant Scholarship

There are many previous related studies on solar radiation in building design. However, there is no scientific study specifically referring to the discussion on traditional Malay shading elements on contemporary apartment design. There are studies by several scholars concerning thermal comfort factors due to problems of indoor heat gains in Malaysia. Early study by Victor Olgyay had argued on the importance of solar control and the integration shading devices in modern building design with reference to the climate of the region, which guides the architects to understand on the passive heating and cooling system in architectural design. One of the recent studies was by Ismail and Idris (2002), and Lim, Ahmad and Ossen (2013) issues on heat gains due to exposure of modern and contemporary high rise building facades to direct sunlight. The other study was by Abdul Rahman (1995) and Omer (2008) on housing design related to thermal comfort with integration of passive design solution to tackle solar radiation.

1.4 Research Design

With reference to the importance of the study in the earlier discussion, this study creates a research design leading to provide answers based on the result of analysis on direct sunlight penetration to indoor area from the apartment façade in Putrajaya, Malaysia. SunTool software will be use in the survey to simulate the results, showing the extent of sunlight penetration values the at the selected sun positions. The software provides reliable data used in a study. It is computer software which provides performance data the same as the data provided by manual calculation in measuring shading performance of the building facade using solar chart superimposed by shadow angle protractor as illustrated in Figure 1. With the results of analysis as performance data, this study can verify the strengths and weaknesses of the traditional shading elements integrated in the apartment façade design. Among the research questions are as follows:

- Are apartments designed with traditional Malay style has a facade with good shading performance?
- To what extent does the performance of traditional shading elements in blocking direct sunlight on the apartment facade?
- What are the types of traditional shading elements which provide excellent shade on the apartment facade?

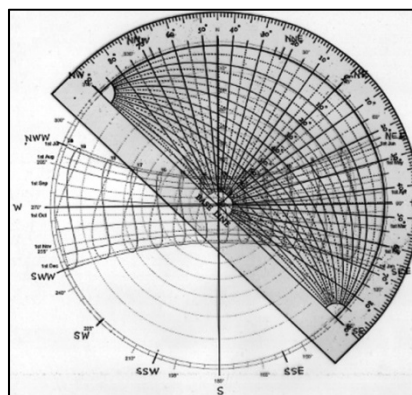


Figure 1. Solar chart superimposed by shadow angle protractor

2. Literature Review

One of the popular post modern style in apartment design in Putrajaya, Malaysia is an integration of traditional

Malay architectural style. According to Hassan (2004), many traditional buildings have a roof system made from two different roof slopes. The upper pitch roof is relatively steep about 45-55 degree compared to the lower pitch roof (also known as 'attached roof'), which has about 25 degrees slope. The two pitches are merged to form two different roofs slopes under one roofing system. This type of roofing system is the most commonly constructed for the core house (*rumah ibu*) normally attached with balcony known as 'anjung'. Based on Khan (1981), the application of pitch roof in building construction in Nusantara Region is a logical choice in hot wet climatic conditions. The steep pitch, permits rapid removal of rain water and creates a high sloping ceiling ideal for inducing air movement, ventilation and escape of hot air, thus bringing in comfort (Hassan & Ku Hassan, 2001).

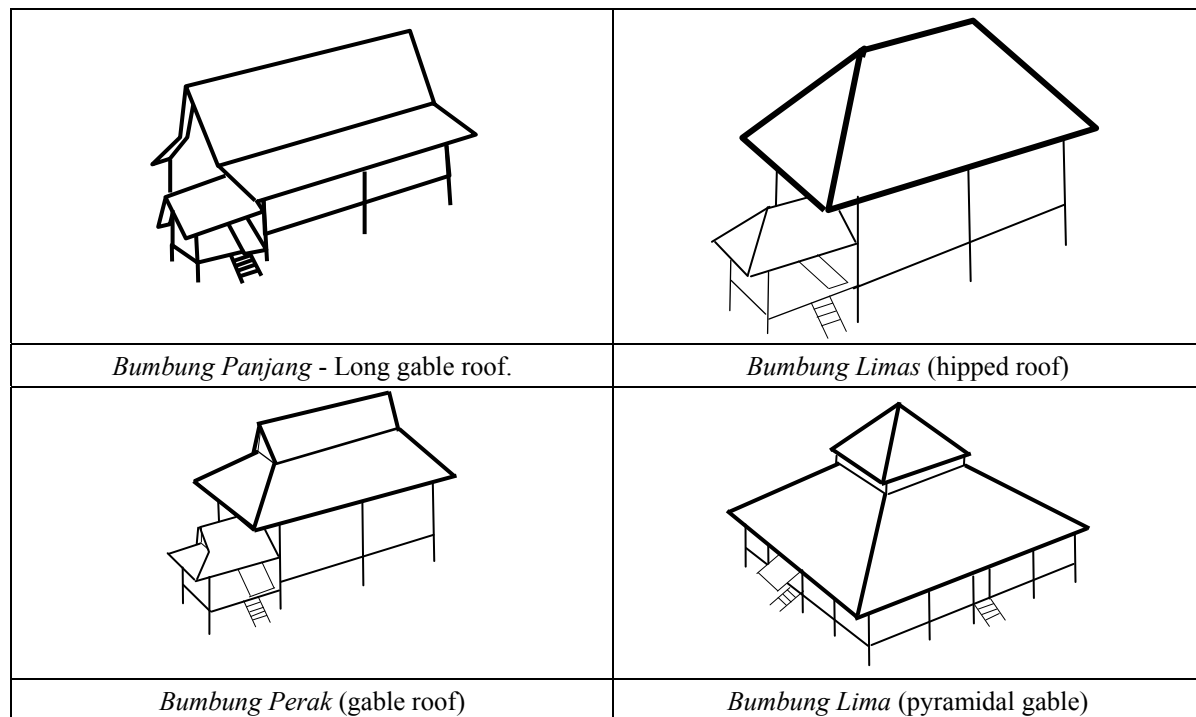


Figure 2. Traditional roof types

Long roof (*Bumbung Panjang*) building is characterised by a long gable roof (Figure 2). Its simple construction accounts for its popularity and is also the most highly developed of the various building types. It has a system which allows extensions and addition of many type and variation. *Long roof building type* is the simplest of the four building forms. It has a simple gable roof, supported by kingposts. The most common roofing material used for the *bumbung panjang* is the *attap*. *Attap* is thatch made from *nipah* and other palm trees found in the local natural vegetation. Limas Roof (*Bumbung Limas*) is characterised by a hipped roof which is believed to be influenced by colonial roof form. The floor plan is almost square in shape and the common types of roofing materials are either clay tiles or asbestos sheet. An urban family as it allows the use of modern furniture which requires higher headroom. Limas roof form is not originated from Nusantara Region but developed through foreign influences such as the colonial Dutch and British building forms during their period in the country.

Perak Roof (*Bumbung Perak*) is believed to be copied from Dutch building forms in the colonial period (Figure 1). The pointed high pitch roof uses a combination of *attap* (*nipah* roof), zinc and other modern roofing materials to allow for the more elaborate roof construction here. The ridge is straight. The centre of the roof is covered with *attap*. This technique ensures that the building does not over heat during the day and has the ventilating properties of the *attap*. Because the edges of the roof are protected by the metal, the roof is last longer. Lima Roof (*Bumbung Lima*) is commonly used for mosque and *Surau* (building not conducted Friday Prayer). It is a square base in plan with recessed wall for corridor (also known as 'serambi') and usually raised several foot above the ground. The roof is a two and three-tiered pyramidal form (pyramid roof with two tiers attached roof system). The tiers of the roof are meant for ventilation purposes in the hot humid climate of Nusantara Region. The top most roofs are lifted high enough to allow the presence of clerestory windows. It is tiered with attached roof and made of more permanent materials like tiles or shingles rather than *attap*. Thus the number of tiers is related to the size of the mosque's plan.

3. Case Study

Two apartment buildings with traditional architectural style were selected as the case studies. This research will compare and analyse the extent of sunlight penetrations and the effectiveness of the existing shading devices on apartment façades with a design of the traditional architectural style. Both apartment locations are in Putrajaya, the administrative city for the federal government of Malaysia. The new town is the latest city built in Malaysia (Hassan, 2005; Hassan, 1999). The first case study (also marked as Traditional 1) is an apartments building consists of sixteen stories located in Precinct 16 (Figure 3). It has Perak roof type. The second case study (Figure 4) (also marked as Traditional 2) is a seventeen storey apartments located nearby artificial lake in Precinct 18. It has long roof type. Since there are many house units in one apartment block, this study has selected facades the most top floor unit for the simulation and analysis.

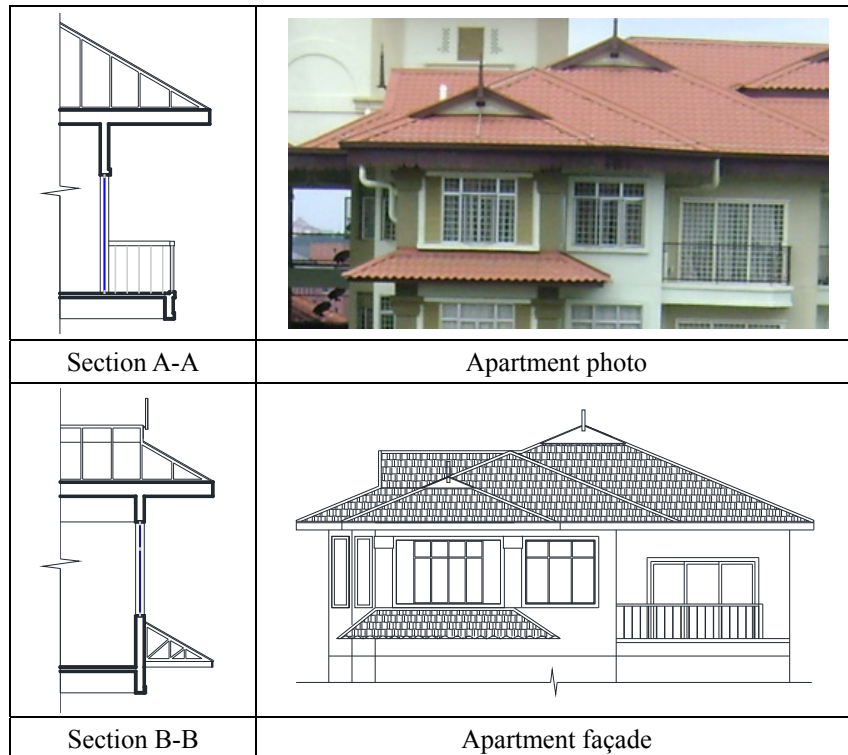
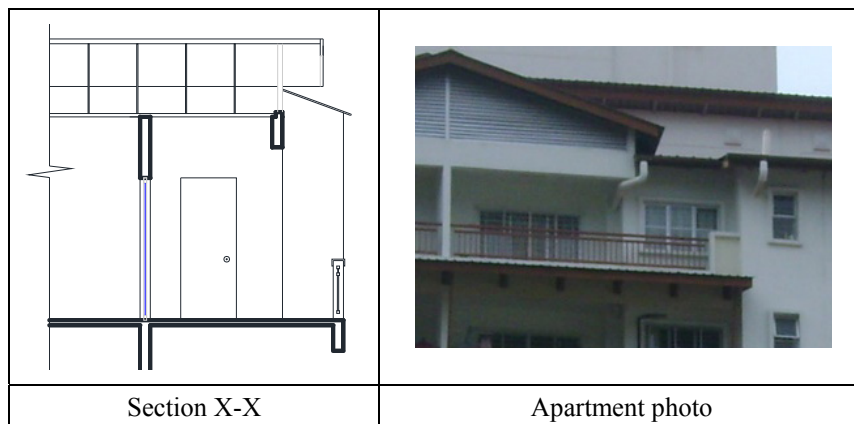


Figure 3. The first case study apartment



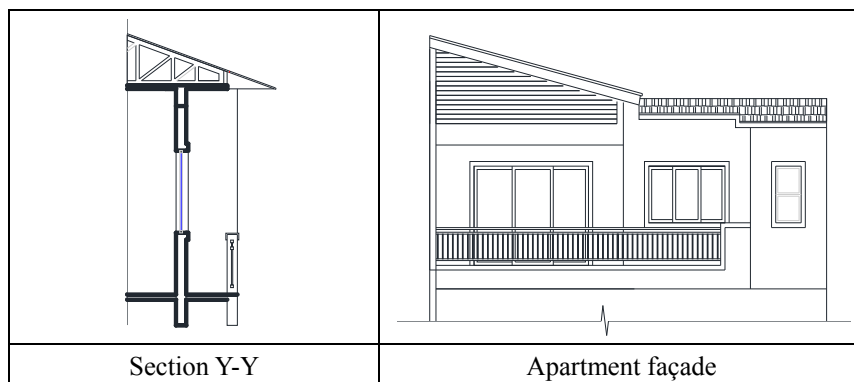


Figure 4. The second case study apartment

4. Methods

The simulation will be done using SunTool software to calculate the extent of sunlight penetration and the sunlight shading. This study aims to get the results of the extent of sunlight penetration when the building facade exposes to the sunlight at the maximum level at a day time; by doing the simulation when the sunlight are perpendicular to the house facade (east facade during morning hours and west facade during evening hours of simulation), then the survey will be able to discuss the efficiency of facade’s shading design (Mazloomi, Hassan, Bagherpour, & Ismail, 2010). In order to get the results at the maximum exposure level, the study will be limited to the changing of the sun path to get the perpendicular of the sunlight to the east (90°) and west (270°) Table 1 and Figure (5), and also the other limitation is that there are at certain times and dates that the sun path’s azimuth is not possible to have perfectly at 90° (Hassan & Arab, 2013; Arab & Hassan, 2012). In these cases, the closest azimuths nearest to 90° will be used when the simulation is made from 8:00 am to 7:00 pm, which are listed in Table 1. It is important for a tropical country like Malaysia where the duration of daytime and night time is almost the same throughout a year.

Table 1. Time, date and azimuth of the sun when the sunlight extent penetration of façade was calculated for cases in Malaysia

Orientation	Time	Date	Azimuth	Orientation	Time	Date	Azimuth
East 90°	7 am	23 March	90°	West 270°	1 pm	16 September	90.5°
	8 am	25 March	90°		2 pm	29 March	89.8°
	9 am	27 March	89.8°		3 pm	18 September	89.8°
	10 am	28 March	90.1°		4 pm	26 March	89.9°
	11 am	29 March	90°		5 pm	24 March	89.9°
	12 pm	29 March	92.2°		6 pm	22 March	89.9°

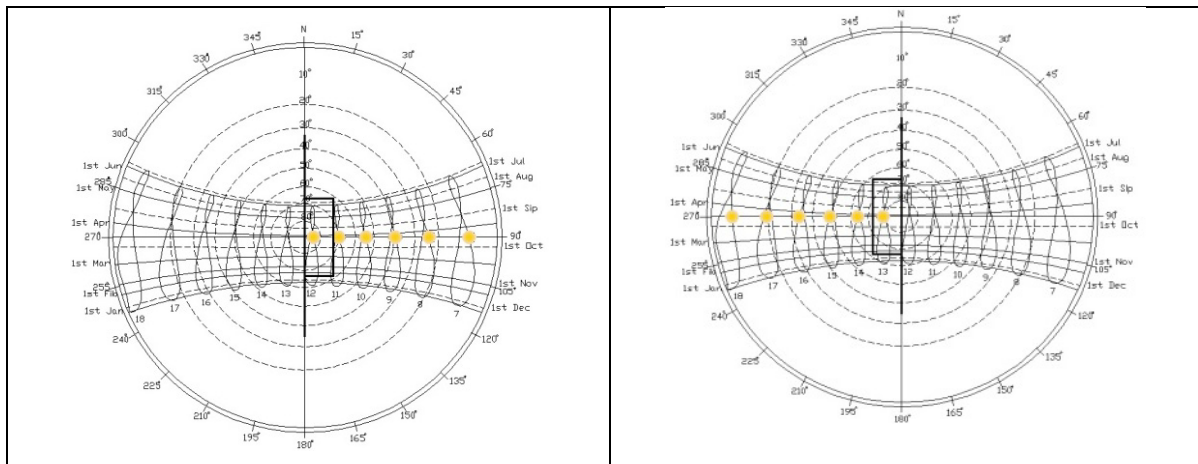


Figure 5. Sun path diagram shows the position of the sun perpendicular to the house facade from 7am to 12pm at orientation of 90° (left) and from 1pm to 6pm at orientation of 270° (right). Source: SunTool Software

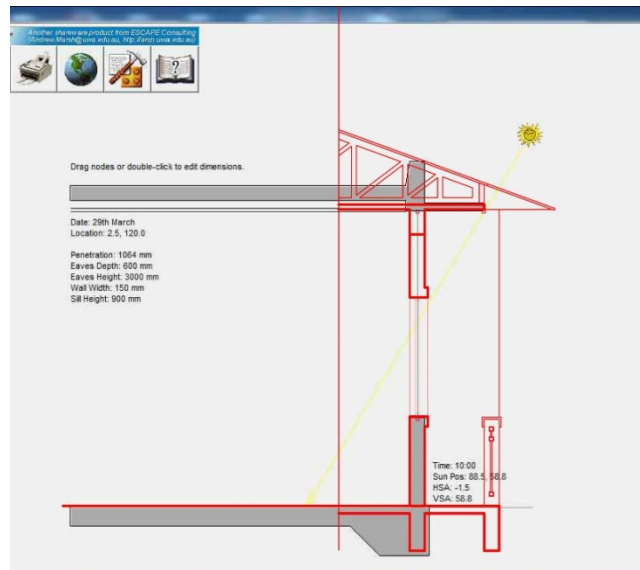


Figure 6. SunTool software (window section)

In order to get the correct position and orientation, all data about the location, time, date and orientation will be keyed into the SunTool software to do the simulation and calculate the percentage of facade’s shading area and the extent of sunlight penetration (Figure 6). And then the facade’s dimensions such as depth of exterior shading device, height, wall’s width and sill height will be keyed in the SunTool software. After drawing the facade and entering all required data, the software will be able to do the simulation and provides the facade shading area and the extent of sunlight penetration inside the house for the analysis. The extent of sunlight penetration inside the room will be measured using the same software (The SunTool) as shown in Figure 7. However, if the overhang shading device is longer than the upper window wall (Figure 8), the second line must be drawn parallel to sun beam of the upper window wall. The second line represents the actual extent of the sunlight penetration. This condition commonly occurs in the early morning and late evening due to low sun angle in the sky. The calculation will be as follow (Hassan & Bakhlah, 2013):

$$\text{Penetration} = \text{Extent of the Sunlight Penetration calculated by the SunTool software (mm)} - \text{Distance between two lines (mm)}$$

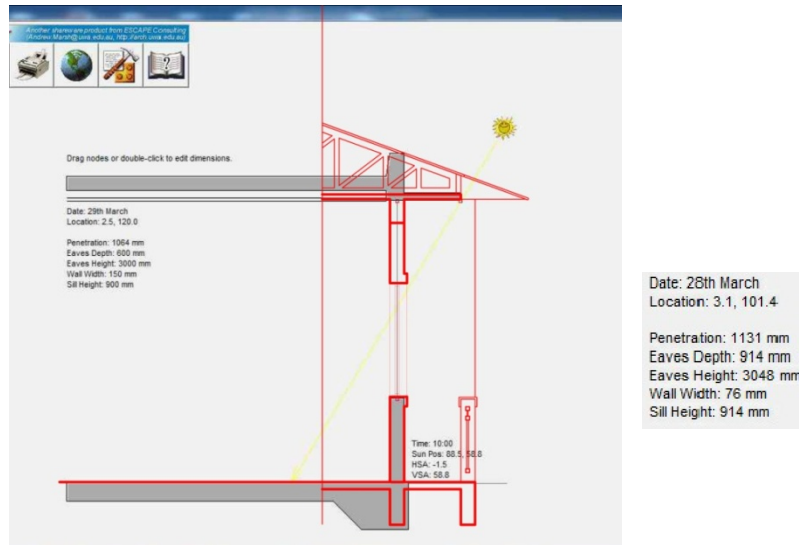


Figure 7. The extent of sunlight penetration given by the SunTool software

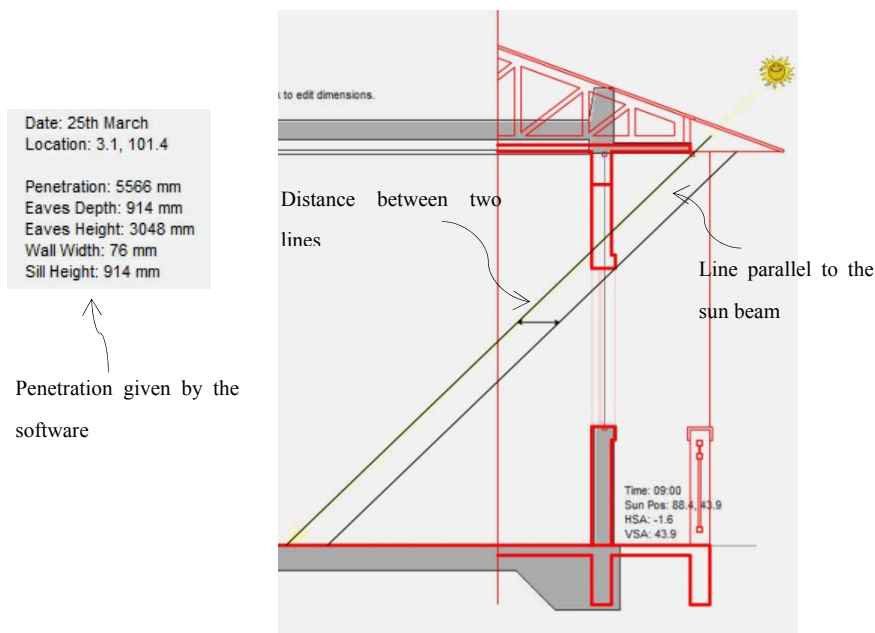
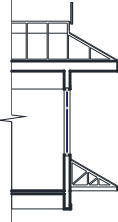
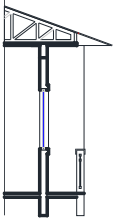
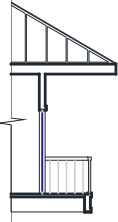
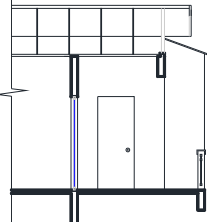


Figure 8. The extent of sunlight penetration calculation when the overhang shading device blocking the sunlight higher than the upper window wall

5. Results of Analysis

This study analysis compares results generated by the computer simulation on the extent of sunlight penetrations of the most top floor unit’s facade of the two traditional style apartments located in Putrajaya, Malaysia. The comparison will be window section between Section A-A from the first case study and Section Y-Y from the second case study. The other comparison is between Section B-B and X-X (door section). Table (2) and Figures (9 and 10) show the comparative data on results of the extent of the sunlight penetration of Window Section A-A and Y-Y and Door Section B-B and X-X of the first and second case study.

Table 2. The extent of sunlight penetration of the first and second case study

Extent of sunlight penetration (mm)				
Time	Window Section		Door Section	
	Traditional 1 Section A-A	Traditional 2 Section Y-Y	Traditional 1 Section B-B	Traditional 2 Section X-X
				
8:00 am	14620	11988	11842	12073
9:00 am	5103	4352	4347	3392
10:00 am	2213	2338	1613	912
11:00 am	738	1298	138	0
12:00 pm	0	328	0	0
1:00 pm	0	0	0	0
2:00 pm	0	0	0	0
3:00 pm	0	660	0	0
4:00 pm	1431	1981	831	36
5:00 pm	2925	3009	2325	1535
6:00 pm	6806	4772	5540	4890
7:00 pm	27919	24280	22774	27121

The first comparison is on window section. The extent of sunlight penetration as shown in Figure 9 for Section A-A was 14620 mm at 8:00 am in the morning (one hour after the sunrise), and then it had decreased gradually from 5103 mm (9:00 am) and 2213 mm (10:00 am) to reach 738 mm at 11:00 am. From 12:00 pm until 3:00 pm the façade had excellent shading performance without sunlight penetration. The overhang roof design was excellent because it was able to block the sunlight penetration into the building during the period from 11:00 am to 3:00 pm where heat level from sunlight was high. However during the evening hours, the extent of sunlight penetration had steady increases from 1431 mm at 4:00 pm and 2925 mm at 5:00 pm to 6806 mm at 6:00 pm. This penetration had caused heat gains to the indoor area of the apartment unit. The maximum sunlight extent penetrated at the last hour of simulation (7:00 pm) with 27919 mm when the sunbeams were almost perpendicular to the west façade.

This analysis finds that the second case study had slightly better shading performance than the first case study on the extent of sunlight penetration during morning hours from 8:00 am to 12:00 pm and late evening hours from 6:00 to 7:00 pm. However during evening hours from 3:00 to 5:00 pm, the results on the extent of sunlight penetration on the apartment facade had slightly poorer performance than the first case study. The sunlight penetration at 8:00 am was 11988 mm and then it had a gradual decline from 4352, 2338 and 1298 at 9:00, 10:00 and 11:00 am respectively to 328 mm at 12:00 pm. The result shows that no sunlight penetration was from 1:00 to 2:00 pm. However, the second case study had a steady increases on the sunlight penetration starting from 660 mm at 3:00 pm, 1981 mm at 4:00 pm and 3009 mm at 5:00 pm to 4772 mm at 6:00 pm. The longest extent was at 7:00 pm with 24280 mm due to the sunbeam almost perpendicular to the house façade during the sunset.

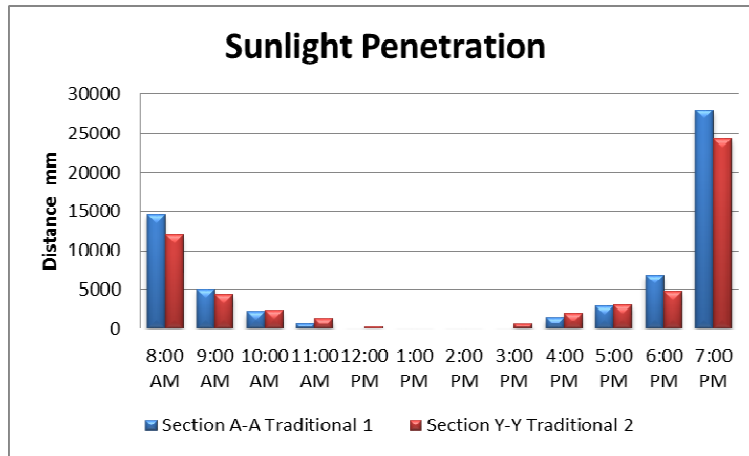


Figure 9. The extent of sunlight penetration of Section A-A and Y-Y

The second comparison on this study is to compare the extent of sunlight penetrations of the door sections at both case studies with reference to results of the simulation on Section B-B and X-X (Figure 10). Table 2 shows that the first case study had slightly poorer results of shading performance than the second case study except at 8:00 am and 7:00 pm. The first case study had 11842 mm extent of the sunlight penetration compared to the second case study with 12073 mm about one hour after the sunrise. From 9:00 to 11:00 am, the extent of sunlight penetration had steady decreased from 4347 mm at 9:00 am and 1613 mm at 10:00 am to 138 mm at 11:00 am. The analysis finds that no sunlight penetration from 12:00 pm to 3:00 pm. However, the sunlight penetration had occurred at 4:00 pm with 831 mm and later had gradual increases from 2325 mm at 5:00 and 5540 mm at 6:00 pm with to the maximum extent of sunlight from the west façade with 22774 mm at 7:00 pm (about sunset time) during the simulation. The second case study had started with 12073 mm at 8:00 am and later had a steady decline to 3392 mm at 9:00 am and 912 mm at 10:00am. There was no sunlight penetration occurred from 11:00 am to 3:00 pm which means that the indoor area were free from sunlight exposure during these 4 hours. At 4:00 pm, the indoor area received a small amount of sunlight penetration with only 36 mm, and later it had the sunlight penetration with gradual incline from 1535 mm at 5:00 pm to 4890 mm at 6:00 pm. Due to then sunbeam at almost perpendicular to house façade at 7:00 pm (sunset), the sunlight had deepest penetration to the indoor area with 27121 mm.

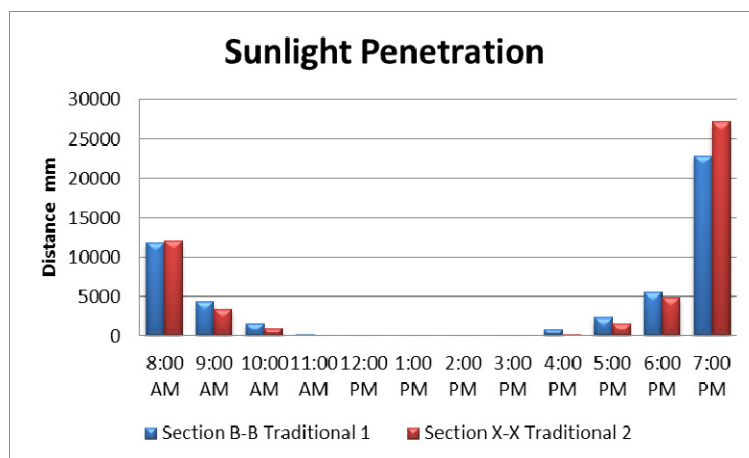


Figure 10. The extent of sunlight penetration of Section B-B and X-X

6. Discussion

The findings from the results of analysis are discussed as follows:

1. This study finds that the early morning hours show sunlight penetration between 10000 mm and 15000 mm

- in both of window and door sections at both case studies. However, these extents of sunlight penetration are not critical as they occurred about one hour after sunrise time and during early morning.
2. Sunlight penetration to the indoor area from sunrise time from to 10:00 am is not acute because it is during early morning hours which are necessary for the human skin exposure to sunlight for gaining Vitamin D formation on the skin tissue (Landry & Breton, 2009).
 3. The façade had satisfactory shading performance from 10:00 to 11:00 am. The longest extent of sunlight penetration ranging from 138 to 1298 mm.
 4. Both case studies have excellent shading performance from 12:00 am to 3:00 pm as most of the indoor areas did not have sunlight penetration. If they had, the sunlight penetration was very minimal.
 5. All facades have poor shading performance from 4:00 to 5:00 pm. The range of the extent of sunlight penetration was from 831 mm at 4:00 pm except Door Section X-X in the Second Case Study, to 3009 mm at 5:00 pm. Having exposed to sunlight penetration during these hours is intolerable in passive design approach as it will generate a high heat gains from the harsh and warm sunlight, therefore creating solar radiation.
 6. All the case studies had long extent of sunlight penetration at 6:00 pm. However the heat from sunlight is not as severe as the heat gained from 11:00 am to 5:00 pm.
 7. Having sunlight penetration at 7:00 pm is not critical as it is sunset hour.
 8. All the case studies applies traditional roof overhang concept in the design. In general, the second case study had better shading performance than the first case study. The recessed wall design in the case study 2 has given additional shading elements in blocking the sunlight. Recessed wall is another traditional shading element. In addition, the Door Sections had better shading performance than Window Sections. The Door Sections have integrated balcony or corridor element which blocks the sunlight.
 9. Window Section in the first case study had the poorest performance. It only relies on its attached roof/roof overhang for the shading elements. Door Section in the second case study had the best performance with integration of roof overhang, balcony and recessed wall in the façade design.

The study finds that Door Section in the second case study has the best shading performance followed by Door Section in the first case study, Window Section in the second case study and Window Section in the first case study. Recessed wall, balcony, attached roof and roof overhang are a typical shading elements the traditional architectural style. They are very effective in preventing sunlight from getting penetration inside the house. This study finds that the more the application is the traditional shading elements, the more is the shading performance on the façade walls. However, traditional shading elements are not effective to block the sun angle especially from 3:00 to 6:00 pm which creating heat gains due to sunlight penetration to the indoor area for the both case studies. The reason is traditional shading elements are designed only to block sunlight from horizontal angle of the sun position. They are not designed to obstruct sunlight from vertical angle of the sun position. New shading devices for vertical sun angle must be integrated in the building façade. Maximum extent of sunlight penetration occurred during the sunrise (7:00-8:00 pm) and sunset (6:30-7:30 pm) hour because of the low angle of the sun position in the sky during these times (Landry & Breton, 2009). However, the heat gain in indoor area due to solar radiation is tolerable because it generated by pleasant and light sunlight.

Acknowledgments

The authors would like to express appreciation for the financial support under Research University Grant by Universiti Sains Malaysia.

References

- Abdul Rahman, A. M. (1995). Housing design in relation to environmental comfort. *Build. Res. Inform.*, 23, 49-54. <http://dx.doi.org/10.1080/09613219508727423>
- Arab, Y., & Hassan, A. S. (2012). Daylighting analysis of pedentive dome's mosque design during summer solstice with case studies in Istanbul, Turkey. *International Transaction Journal of Engineering, Management, & Applied Sciences & Technologies*, 3(2), 167-183.
- Bakhlah, M. S., & Hassan, A. S. (2012). The study of air temperature when the sun path direction to Ka'abah: with a case study of Al-Malik Khalid Mosque, Malaysia. *International Transaction Journal of Engineering, Management & Applied Sciences & Technologies*, 3(2), 185-202.
- CETDEM. (2005). *Malaysian Urban Household Energy Consumption Patterns*. Centre for Environment, Technology and Development, Malaysia. Petaling Jaya: CETDEM.
- Department of Statistics Malaysia. (2010). *Characteristics of Living Quarters 2010*. Putrajaya: Department of

- Statistics Malaysia.
- Feriadi, H., & Nyuk, H. W. (2004). Thermal Comfort for Naturally Ventilated Apartments in Indonesia. *Energy and Building*, 36(7), 614-626. <http://dx.doi.org/10.1016/j.enbuild.2004.01.011>
- Hassan, A. S. (2004). *Issues in Sustainable Development of Architecture in Malaysia*. Penang: Universiti Sains Malaysia Press.
- Hassan, A. S. (1999). Putra Jaya: The Direction of Malaysian New Town. *Proceedings of the 5th International Congress of Asian Planning Schools Association (APSA)* (pp. 165-175). Seoul: Seoul National University.
- Hassan, A. S. (2005). *Konsep Rekabentuk Bandar di Semenanjung Malaysia: Kuala Lumpur dan Bandar-Bandar di Sekitarnya*. Penang: Universiti Sains Malaysia Press.
- Hassan, A. S., & Arab, Y. (2013). The essence of design with light: Single pedentive dome mosque in Turkey and Bosnia Herzegovina during winter solstice. In S. Omer, & A. S. Hassan, *From Anatolia to Bosnia: Perspectives on Pedentive Dome Mosque Architecture*. Penang: Universiti Sains Malaysia Press.
- Hassan, A. S., & Bakhlah, M. S. O. (2013). Shading Analysis on Front Facade of Modern Terraced House Type in Petaling Jaya, Malaysia. *Procedia Social and Behavioural Sciences*, 91, 13-27. <http://dx.doi.org/10.1016/j.sbspro.2013.08.396>
- Hassan, A. S., & K. A. Ku Hassan. (2001). *Konsep Perumahan Tradisional Berkelompok dan Berdensiti Tinggi di Pantai Barat Semenanjung Malaysia (High Density Traditional Clustered Housing Concept at Western Coastal Area in Peninsular Malaysia)*. Penang: Universiti Sains Malaysia Press.
- Hassan, A. S., & Nawawi, M. S. A. (2014). Malay Architectural Heritage on Timber Construction Technique of the Traditional Kampung Laut Old Mosque, Malaysia. *Asian Social Sciences*, 10(8), 230-240. <http://dx.doi.org/10.5539/ass.v10n8p230>
- Hassan, A. S., & Ramli, M. (2010). Natural ventilation of indoor air temperature: A case study of the Traditional Malay House in Penang. *American Journal of Engineering and Applied Sciences*, 3(3), 521-528. <http://dx.doi.org/10.3844/ajeassp.2010.521.528>
- Ismail, A. M. (2002). Issues in tropical architecture: tall buildings, wind and ventilation. *Journal of Housing, Building and Planning*, 9, 63-79.
- Ismail, A. M., & Idris, M. F. (2002). Issues in tropical architecture: High-rise buildings and wind driven ventilation. *Proceeding of The 2nd Civil Engineering National Seminar*, Universiti Sains Malaysia.
- Khan, H. (1981). *MIMAR 2: Architecture in Development*. Singapore: Concept Media Ltd.
- Landry, M., & Breton, P. (2009). *Daylight simulation in Autodesk 3ds Max Design 2009 - Advanced Concepts*. San Rafael: Autodesk Inc.
- Landry, M., & Breton, P. (2009). Daylight simulation in Autodesk 3ds Max Design 2009-advanced concepts. *Autodesk Inc*.
- Lim, Y. W., Ahmad, M. H., & Ossen, D. R. (2013). Internal shading for efficient tropical daylighting in malaysian contemporary high-rise open plan office. *Indoor and Built Environment*, 22(6), 932-951. <http://dx.doi.org/10.1016/j.buildenv.2012.04.015>
- Mazloomi, M., Hassan, A. S., Bagherpour, P. N., & Ismail, M. R. (2010). Influence of Geometry and Orientation on Flank Insolation of Streets in an Arid Climate City. *American Journal of Engineering and Applied Sciences*, 3(3), 540-544. <http://dx.doi.org/10.3844/ajeassp.2010.540.544>
- Olgay, V., & Olgay, A. (1957). *Solar Control and Shading Devices*, New Jersey: Princeton University Press.
- Omer, A. M. (2008). Energy, environment and sustainable development. *Renew. Sustain. Energy Rev.*, 12(9), 2265-2300. <http://dx.doi.org/10.1016/j.rser.2007.05.001>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

Parameters Estimation of LFM CW Signals Based on Periodic Fractional Fourier Transform

Ying Liu¹, Dianren Chen¹ & Lei Chen¹

¹ College of Electronic Information and Engineering, Changchun University of Science and Technology Changchun, China

Correspondence: Lei Chen, Changchun University of Science and Technology, Changchun, China. Tel: 86-138-4304-7402. E-mail: chenlei511@126.com

Received: July 13, 2014 Accepted: July 23, 2014 Online Published: August 28, 2014

doi:10.5539/mas.v8n5p143

URL: <http://dx.doi.org/10.5539/mas.v8n5p143>

Abstract

When using fractional Fourier transform (FrFT) to detect and estimate linear frequency-modulated continuous wave radar signals, two problems will appear: multiple peaks occur in FRFT image and the output SNR at the true parameter values does not increase when the observation time is longer than the signals period. A multiple period LFM CW signals parameters estimation method based on period FRFT (PFRFT) is studied in this paper. The PFRFT formula of multiple period LFM CW signals is given. The relationship among PFRFT output SNR, observation time and sample signals SNR is analyzed. The estimation accuracy formula of PFRFT is derived. At last, numerical simulation shows the effectiveness of the algorithm and PFRFT is superior to FRFT for estimating a multiple periods LFM CW signals.

Keywords: LFM CW, FRFT, PFRFT, Parameter estimation, SNR

1. Introduction

Linear frequency-modulated continuous wave (LFMCW) signal is used in various fields, including radar, sonar and electronic monitoring. Frequency analysis method for the detection of non-stationary signals are commonly used, such as: Wigner-Ville distribution (WVD), Short-time Fourier transform (STFT), fractional Fourier transform (FrFT) and so on [1]. WVD can be effectively used for a single period LFM CW signal parameter estimation, however, in the case of multi-periods signals, cross-term would interfere estimation of signals center frequency and modulation frequency; Smooth Pseudo Wigner-Ville Distribution (SPWVD) and Time-frequency distribution series (TFDS) can effectively suppress the influence of cross terms while reduce the frequency resolution of the estimated parameters [2]. As the FrFT effectively overcomes the two problems above, the algorithm and related derivative algorithm have been widely studied and applied to the field of LFM CW signals detection and parameter estimation. P.R. White and others in-depth researched and analysis of the use of the FrFT performance parameters of chirp signals detection and estimation in the literature [3]. Qu Qiang and others studied the chirp signals detection and parameter estimation based on adaptive FrFT in the literature [4]. C. Clemente and others use FrFT algorithm to study doppler and CS imaging algorithm. Liu Jiancheng and others detailed analyzed the fractional Fourier domain SNR of LFM signals under Gaussian white noise background in the literature [5]. The Researches mentioned above do not give a method to improve the detection and parameter estimation performance of FrFT in the noise environment.

It is presented a method of estimation and detection LFM CW signals parameters based on PFRFT algorithm. When estimation and detection the parameters of a single period LFM CW signal, this algorithm and FRFT have the same advantages, but to estimate the parameters of the multi-periods LFM CW signals, this algorithm is better.

2. PFRFT of LFM CW Signals

2.1 LFM CW Signals Mathematical Model

The LFM CW signals shown in Figure 1, the mathematical expression of the m-th period of the LFM CW signals [7] as follows:

$$s_m(t) = A_m \cdot e^{j(2\pi f_m t + \pi g_m \text{mod}(t, T_z)^2)} \quad (1)$$

Where A_m is the signals amplitude; f_m is the signals initial frequency; T_z is the period; g_m is FM rate, $g_m = B_m / T_z$; B_m is the FM bandwidth; j is the imaginary unit; t is time variable, $m=1 \dots M$, M is the number of the signals' period. The parameter of every LFMCS signals' period is same. It is used the symbol A, f and g to express the amplitude, starting frequency and frequency modulation rate of the signals in this paper.

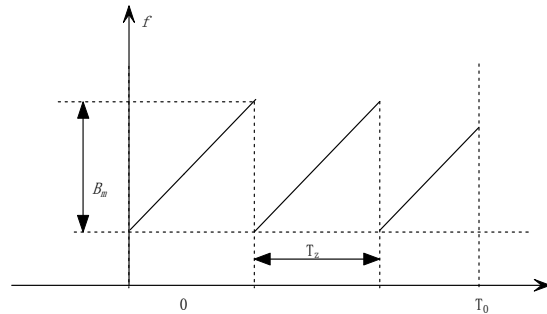


Fig. 1. LFMCS Signals

2.2 FrFT

There have a variety of definitions of FrFT due to the different application background. Although the physical interpretations of these definitions are not the same, they are equivalent in mathematics. One definition of arbitrary can be derived by other definitions. The FrFT of signal $s(t)$ can be expressed as follows^[8]:

$$FRFT_s(\alpha, u) = \int_{-\infty}^{+\infty} K_\alpha(u, t) s(t) dt \tag{2}$$

$$K_\alpha(u, t) = \begin{cases} \sqrt{\frac{1-j \cot \alpha}{2\pi}} e^{j\frac{t^2+u^2}{2} \cot \alpha - jut \csc \alpha}, & \alpha \neq n\pi \\ \delta(t-u), & \alpha = 2n\pi \\ \delta(t+u), & \alpha = (2n+1)\pi \end{cases} \tag{3}$$

Where $\cot(\cdot)$ means the cotangent, $\csc(\cdot)$ means the cosecant.

Two different situations due to the LFMCS signals' observation time is discussed as below:

(1) When the signals period number $M=1$, observation time $T \leq T_z$, The signal can be expressed as $s_1(t) = A \cdot e^{j(2\pi f \cdot t + \pi g \cdot t^2)}$, its FrFT can be expressed as:

$$\begin{aligned} FRFT_{s_1}(\alpha, u) &= \int_{-\infty}^{+\infty} K_\alpha(u, t) s_1(t) dt \\ &= \sqrt{\frac{1-j \cot \alpha}{2\pi}} e^{j\frac{u^2}{2} \cot \alpha} \int_0^{T_0} A \cdot e^{j(2\pi f - u \csc \alpha)t} \cdot e^{j(\pi g + \frac{\cot \alpha}{2})t^2} dt \end{aligned} \tag{4}$$

$FRFT_{s_1}(\alpha, u)$ achieves peak point at $2\pi f - u_0 \cdot \csc \alpha_0 = 0$, $\pi g + \frac{\cot \alpha_0}{2} = 0$, then the estimation result is

$$f_0 = \frac{u_0 \cdot \csc \alpha_0}{2\pi}, \quad g_0 = -\frac{\cot \alpha_0}{2\pi}.$$

(2) When the signals period number $M > 1$, observation time $T = M \times T_z$, The signal can be expressed as:

$s_2(t) = A \cdot e^{j(2\pi f t + \pi g \text{mod}(t, T_z)^2)} = \sum_{i=0}^{M-1} s_1(t + i \cdot T_z)$. Based on the Characteristics of FrFT^[7], the FrFT of the signal can be expressed as:

$$\begin{aligned}
 FrFT_{s_2}(\alpha, u) &= \int_{-\infty}^{+\infty} K_{\alpha}(u, t) s_3(t) dt = \int_{-\infty}^{+\infty} K_{\alpha}(u, t) \sum_{i=0}^{M-1} s_1(t - i \cdot T_z) dt \\
 &= \sum_{i=0}^{M-1} \int_{-\infty}^{+\infty} K_{\alpha}(u, t) s_1(t - i \cdot T_z) dt \\
 &= \sum_{i=0}^{M-1} FRFT_{s_1}(\alpha, u - i \cdot T_z \cos \alpha) e^{j \frac{(i \cdot T_z)^2}{2} \sin \alpha \cos \alpha - u \cdot i \cdot T_z \cdot \sin \alpha} \\
 &= \sum_{i=0}^{M-1} \sqrt{\frac{1 - j \cot \alpha}{2\pi}} e^{j \frac{(i \cdot T_z)^2}{2} \cot \alpha} \cdot e^{j \frac{(i \cdot T_z)^2}{2} \sin \alpha \cos \alpha - u \cdot i \cdot T_z \cdot \sin \alpha} \\
 &\quad \cdot \int_0^{T_0} A \cdot e^{j(2\pi f - (u - i \cdot T_z \cos \alpha) \cdot \csc \alpha) \cdot t} \cdot e^{j(\pi g + \frac{\cot \alpha}{2}) \cdot t^2} dt
 \end{aligned}
 \tag{5}$$

By formula (5), it can be seen that the $FrFT_{s_2}(\alpha, u)$ achieves peak points at $2\pi f - (u - i \cdot T_z \cos \alpha) \cdot \csc \alpha = 0$,

$\pi g + \frac{\cot \alpha}{2} = 0$. so $u = \frac{2\pi f_0 u_0}{\csc \alpha_0} + i T_z \cdot \cos \alpha$, $\alpha_0 = -\cot^{-1} 2\pi g_0$. In this case the FM rate estimation is correct.

Due to $i=0,1,\dots, M-1$, u has M values, FRFT results will appear M peaks, so the starting frequency can not be accurately estimated.

Using MATLAB to simulate the above situation, the radar echo parameters are shown in the following table. The simulation results shown in figure (2):

Table 1. Simulation parameters

Signals parameters	(1)	(2)
Start frequency	30Hz	30Hz
Frequency modulation rate	8 Hz/s	8 Hz/s
period	2.56s	2.56s
The sampling frequency	200Hz	200Hz
Number of cycles	1	3
Signals to noise ratio	-5 dB	-5dB

In figure (2), (a) (b) (c) means the projection on α -axis, μ -axis and two-dimensional map of signals FRFT when $M=1$. By calculation, the results of simulation and practical results are the same. (d) (e) (f) means the projection on α -axis, μ -axis and two-dimensional map of signals FRFT when $M=3$. Graph (a) (d) peak coordinates are consistent, that is to say, signals modulation frequency estimation is correct, Fig. (e) (f) have 3 peak values respectively, that will interfere estimation of initial frequency. At the same time, comparing $M=1$ and $M=3$, the output signals to noise ratio does not significantly improve while increasing the sample signals observation time. It is considered that FRFT not optimal multi-periods LFM CW signals parameter estimation algorithm.

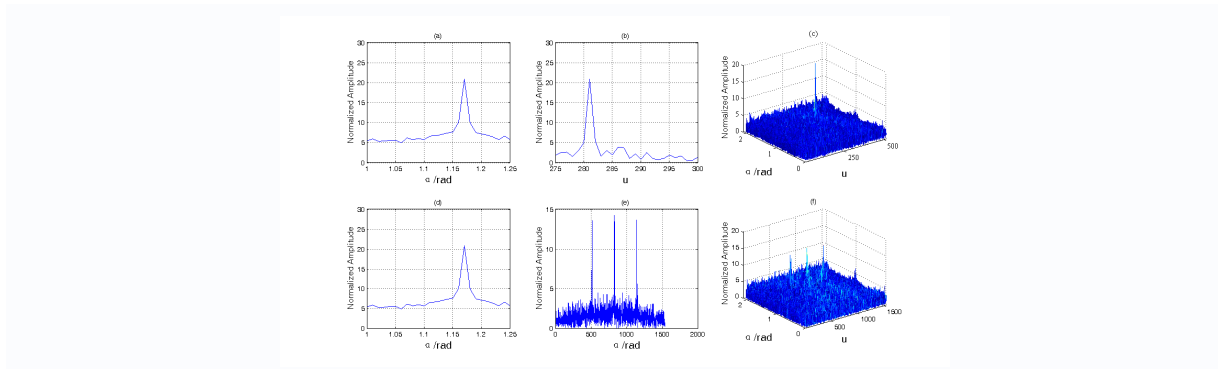


Figure 2. The sample signals FRFT when $M=1$ and $M=3$

2.3 FRFT output SNR

Since the FrFT on the α shaft $(0, 2\pi)$ symmetry^[8], in order to facilitate analysis, we take $0 < \alpha < \pi/2$. Assuming Gaussian white noise conditions, the observe signals is $x(t) = s(t) + n(t)$, $0 < t < T_0$, T_0 means signals observation time, $n(t)$ is Gaussian white noise whose mean is 0 and variance is σ^2 . Literature [5] proposed that the square of the peak on two-dimensional FrFT domain is signals power, and the square of the peak of the noise as the noise power. The FRFT output SNR of the LFM CW signals can be expressed as^[8]:

$$SNR_{FRFT}(u_0, \alpha_0) = \frac{A^2 T_0^2}{2A^2 T_0 \sigma^2 + \sigma^4} = \frac{SNR_{in}^2 \cdot T_0^2}{2T_0 \cdot SNR_{in} + 1} \quad (6)$$

Where^[9]: $SNR_{in} = A^2 / \sigma^2$.

Combining formula (6) and figure 2, simulated the FRFT output SNR with the parameters in Table 1, the result is shown in Figure 3. The results shows that when the observation time $T_0 > T_z$, the SNR in the peak does not increase, while $T_0 < T_z$, the output SNR increases with the observation time increase. It is considered that FRFT is not optimal parameter estimation algorithm when the numbers of the signals' period $M > 1$.

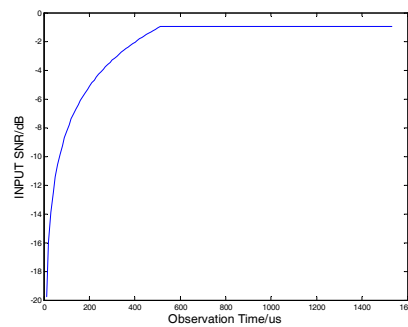


Figure 3. Relationship between FRFT output SNR and observation time

2.4 Periodic FrFT

Through the above analysis, it finds that LFM CW signals estimate by traditional FRFT, when $M=1$, FRFT can effectively estimate the parameters, and the SNR in the peak increases with the observation time increasing. When $M > 1$, FRFT result appears M peaks that interfere starting frequency estimation, although it could effectively estimate the FM rate of the signals, and when the observation time is longer than T_z , the SNR no

longer increase, the max value is $\frac{SNR_{in}^2 \cdot T_z^2}{2T_z \cdot SNR_{in} + 1}$. That means FRFT is not optimal during multiple periods of

LFMCW signals parameter estimation. Therefore this paper presents a Multi-cycle LFM CW parameter estimation method base on period FRFT algorithm, definitions LFM CW signals parameter field $\Omega = (f, g, \tilde{T}_z)$, where f, g, \tilde{T}_z means the starting frequency, modulation frequency and period respectively, when f, g is in the parameter estimation results use α, u representing, signals parameter field $\Omega = (f, g, \tilde{T}_z)$, definition of period FrFT is as follows:

$$PFRFT_s(\Omega') = \int_{-\infty}^{+\infty} F_\alpha(u, t, \tilde{T}_z) s(t) dt \quad (7)$$

Where

$$F_{\alpha}(u, t, \tilde{T}_z) = \begin{cases} \sqrt{\frac{1-j \cot \alpha}{2\pi}} e^{j \frac{(\text{mod}(t, \tilde{T}_z))^2 + u^2}{2} \cot \alpha - ju t - csc \alpha} & , \alpha \neq n\pi \\ \delta(t-u), & \alpha = 2m\pi \\ \delta(t+u), & \alpha = (2n+1)\pi \end{cases} \quad (8)$$

Where $\text{mod}(\bullet)$ means modulo operation, with formula (1) obtains

$$\begin{aligned} PFRFT_{s_1}(\Omega') &= \int_{-\infty}^{+\infty} F_{\alpha}(u, t, \tilde{T}_z) s_1(t) dt \\ &= \sqrt{\frac{1-j \cot \alpha}{2\pi}} e^{j \frac{u^2}{2} \cot \alpha} \int_0^{T_0} A \cdot e^{j(2\pi f - u - csc \alpha)t} \cdot e^{j\pi g \cdot \text{mod}(t, \tilde{T}_z)^2 + \frac{\cot \alpha}{2} \text{mod}(t, \tilde{T}_z)^2} dt \end{aligned} \quad (9)$$

The physical meaning of the formula (9), searches the true period with the multi-period LFM CW signals and do FrFT after fold-weighted with the true period, then search the peak value in the three-dimensional parameter field and get the coordinates of the peak, at last, calculate the exact value of the signals' parameters.

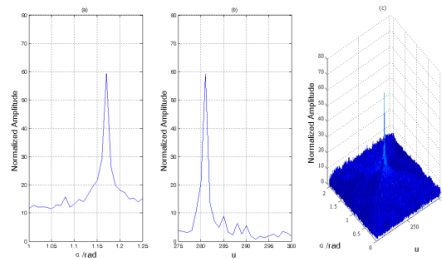


Figure 4. M=3 LFM CW PFRFT results

To simulate the performance of PFrFT with the parameters in Table 1, the result as shown in Figure 4. in the case of $M = 3$, there is a peak in the two-dimensional domain consisting of α, u , PFRFT output of f, g parameter estimates are correct by calculation, and the peak amplitude is 60 which 3 times of the peak amplitude in figure 2, and SNR in the peaks is significantly higher.

Figure 5 shows the value curve of the peak when searching the value of period \tilde{T}_z . Maximum peak is achieved in the period $\tilde{T}_z = 2.56s$, it is consistent with the actual value

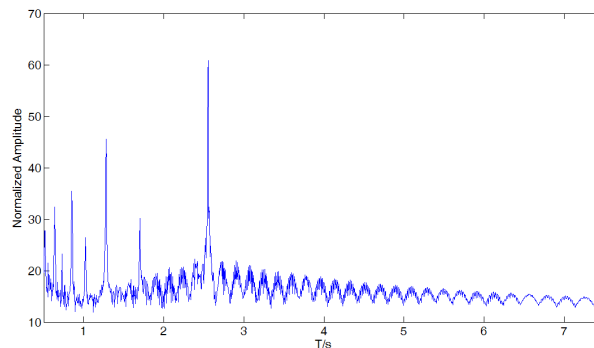


Figure 5. PFRFT output amplitude of the peak point in different period T_z

3. PFRFT Performance Analysis

When analyzing PFRFT output SNR, it lets $|PFRFT_x(\Omega)|^2$ as observing statistic. Assuming Gaussian white noise environment, the observed signals is $x(t) = s(t) + n(t)$ $t \leq T_0$ SNR is SNR_m , observation time is expressed as T_0 , $n(t)$ is Gaussian white noise whose mean is 0 and variance is σ_n^2 . Defined:

$$\begin{aligned} P_x(\Omega) &= |PFRFT_x(\Omega)|^2 = \left| \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F_\alpha(\Omega) s_1(t_1) F_\alpha^*(\Omega) s_1^*(t_2) dt_1 dt_2 \right| \\ &= |PFRFT_s(\Omega)|^2 + |PFRFT_n(\Omega)|^2 + 2\text{Re}[PFRFT_s(\Omega) \cdot PFRFT_n^*(\Omega)] \end{aligned} \quad (10)$$

Where $\text{Re}(\cdot)$ is realistic calculation, $*$ is conjugate, $n(t)$ is Gaussian white noise, so the above equation can be written as:

$$\begin{aligned} P_x(\Omega) &= |PFRFT_x(\Omega)|^2 \\ &= |PFRFT_s(\Omega)|^2 + 2\text{Re}[PFRFT_s(\Omega) \cdot PFRFT_n^*(\Omega)] \end{aligned} \quad (11)$$

3.1 Output SNR

According to Redefinition of SNR in the literature[5], PFrFT output SNR can be expressed as:

$$SNR_{PFRFT} = \frac{|PFRFT_s(\Omega'_0)|^4}{\text{var}(P_x(\Omega'_0))} \quad (12)$$

Where Ω'_0 is the parameter of peak, $\text{var}(\bullet)$ is variance calculation.

According to formula (7), $|PFRFT_x(\Omega)|^2$ at the peak of PFRFT could be expressed as:

$$|PFRFT_s(\Omega'_0)|^2 = A^2 T_0^2 \frac{\text{csc} \alpha_0}{2\pi} \quad (13)$$

According to formula (13):

$$\begin{aligned} E(P_x(\Omega'_0)) &= E(|PFRFT_s(\Omega)|^2 + 2\text{Re}[PFRFT_s(\Omega) \cdot PFRFT_n^*(\Omega)]) \\ &= A^2 T_0^2 \frac{\text{csc} \alpha_0}{2\pi} + \frac{T_0 \sigma_n^2 \text{csc} \alpha}{2\pi} \end{aligned} \quad (14)$$

$$\begin{aligned} E(|P_x(\Omega'_0)|^2) &= E(|PFRFT_s(\Omega)|^4 + 4\text{Re}[PFRFT_s(\Omega) \cdot PFRFT_n(\Omega)] \cdot |PFRFT_s(\Omega)|^2 \\ &\quad + 4\text{Re}^2[PFRFT_s(\Omega) \cdot PFRFT_n(\Omega)]) \end{aligned} \quad (15)$$

Then

$$\begin{aligned} \text{var}(|PFRFT_s(\Omega'_0)|^2) &= E(|PFRFT_x(\Omega'_0)|^4) - E^2(|PFRFT_s(\Omega'_0)|^2) \\ &= \frac{T_0^2 (1 + \text{csc}^2 \alpha_0)}{4\pi^2} [2A^2 T_0 \sigma_n^2 + \sigma_n^4] \end{aligned} \quad (16)$$

So

$$SNR_{PFRFT} = \frac{(A^2 T_0^2 \frac{\text{csc} \alpha_0}{2\pi})^2}{\frac{T_0^2 (1 + \text{csc}^2 \alpha_0)}{4\pi^2} [2A^2 T_0 \sigma_n^2 + \sigma_n^4]} = \frac{SNR_m^2 \cdot T_0^2}{2T_0 \cdot SNR_m + 1} \quad (17)$$

Comparing the formula (17) and formula (6), PFrFT and FrFT of LFM CW signals have the same output SNR expression. But the value range of T_0 is not the same, T_0 in the formula (6) is not more than T_z and T_0 in the formula (17) means sample observation time. It shows that PFrFT and FrFT of the signals have the same output SNR when the sample observation time is less than or equal the period of the signal. But with increasing sample observation time, FrFT output SNR does not increase, while PFrFT output SNR continues increasing, PFrFT takes full advantage of all the information of sample signals. As shown in Figure (6).

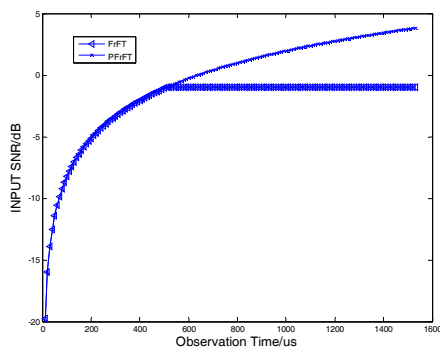


Figure 6. Compare PFRFT、FRFT output SNR

3.2 Output SNR Analysis

First-order perturbation method is used to analysis the PFrFT output SNR as follows. according to formula (11), $\delta P(\Omega)$ is interference term affecting the signals estimate, then

$$P_x(\Omega) = P_s(\Omega) + \delta P_s(\Omega) \tag{18}$$

Where

$$\begin{aligned} \delta P_s(\Omega) &= 2 \operatorname{Re}[PFRFT_s(\Omega) \cdot PFRFT_n^*(\Omega)] \\ &= 2 \operatorname{Re}[\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F_\alpha(\Omega) s_1(t_1) F_\alpha^*(\Omega) n^*(t_2) dt_1 dt_2] \\ &= 2 \operatorname{Re}[\frac{1-j \cot \alpha}{2\pi} A \cdot e^{ju^2 \cot \alpha} \int_0^{T_0} \int_0^{T_0} e^{j(2\pi f - u \csc \alpha)t} \cdot e^{j\pi g \cdot \operatorname{mod}(t_1, T_z)^2 + \frac{\cot \alpha}{2} \operatorname{mod}(t_1, T_z)^2} n^*(t_2) \cdot e^{-j \frac{(\operatorname{mod}(t_2, T_z))^2}{2} \cot \alpha + ju \cdot t_2 \cdot \csc \alpha} dt_1 dt_2] \end{aligned} \tag{19}$$

By the formula (18), due to the introduction of the noise signals, the peak coordinate of $P_x(\Omega)$ moves from the original place $\Omega_0 = (f_0, g_0, T_z)$ to $\Omega_0 + \delta \Omega_0 = (f_0 + \delta f_0, g_0 + \delta g_0, T_0 + \delta T_z)$, first-order partial derivatives of $P_x(\Omega)$ is 0, with $\Omega_0 + \delta \Omega_0$, there is

$$\frac{\partial}{\partial f} [(P_s(\Omega) + \delta P_s(\Omega))]_{f_0 + \delta f, g_0 + \delta g} = 0 \tag{20}$$

$$\frac{\partial}{\partial g} [(P_s(\Omega) + \delta P_s(\Omega))]_{f_0 + \delta f, g_0 + \delta g} = 0 \tag{21}$$

With the Taylor formula at $(P_s(\Omega) + \delta P_s(\Omega))$, the formula (19) (20) can be written as:

$$\frac{\partial}{\partial f} [P_s(\Omega)]_{f_0, g_0} + \frac{\partial}{\partial f} [\delta P_s(\Omega)]_{f_0, g_0} + \frac{\partial^2}{\partial f^2} [P_s(\Omega)]_{f_0, g_0} \cdot \delta f + \frac{\partial^2}{\partial g \partial f} [P_s(\Omega)]_{f_0, g_0} \delta g \cong 0 \tag{22}$$

$$\frac{\partial}{\partial g} [P_s(\Omega)]_{f_0, g_0} + \frac{\partial}{\partial g} [\delta P_s(\Omega)]_{f_0, g_0} + \frac{\partial^2}{\partial g^2} [P_s(\Omega)]_{f_0, g_0} \cdot \delta g + \frac{\partial^2}{\partial g \partial f} [P_s(\Omega)]_{f_0, g_0} \delta f \cong 0 \tag{23}$$

Assuming

$$B = \frac{\partial^2}{\partial f^2} [P_s(\Omega)]_{f_0, g_0} = \frac{A^2 T_0^2 \csc \alpha_0}{12\pi} \tag{24}$$

$$C = \frac{\partial^2}{\partial g^2} [P_s(\Omega)]_{f_0, g_0} = \frac{A^2 T_0^6 \csc \alpha_0}{720\pi} \quad (25)$$

$$u = \frac{\partial}{\partial f} [\delta P_s(\Omega)]_{f_0, g_0} \quad (26)$$

$$v = \frac{\partial}{\partial g} [\delta P_s(\Omega)]_{f_0, g_0} \quad (27)$$

$P_s(\Omega)$ gets extremism when $f = f_0$, $g = g_0$, then

$$\frac{\partial}{\partial f} [P_s(\Omega)]_{f_0, g_0} = 0 \quad (28)$$

$$\frac{\partial}{\partial g} [P_s(\Omega)]_{f_0, g_0} = 0 \quad (29)$$

$$\frac{\partial^2}{\partial f \partial g} [P_s(\Omega)]_{f_0, g_0} = 0 \quad (30)$$

Therefore, formula (20) (21) can be expressed as:

$$\begin{bmatrix} B & 0 \\ 0 & C \end{bmatrix} \cdot \begin{bmatrix} \delta f \\ \delta g \end{bmatrix} = \begin{bmatrix} -u \\ -v \end{bmatrix} \quad (31)$$

The estimation error of start frequency and MF rate can be expressed as:

$$\delta f = -\frac{u}{B} \quad (32)$$

$$\delta g = -\frac{v}{C} \quad (33)$$

Formula (24) (25) show that B , C are constants, $E(u) = 0$, $E(v) = 0$, so according to formula (32) (33), obtains

$$E(\delta f) = E\left(-\frac{u}{B}\right) = 0 \quad (34)$$

$$E(\delta g) = E\left(-\frac{v}{C}\right) = 0 \quad (35)$$

Therefore, the PFrFT multi-periods LFMCW signals parameters estimation is unbiased estimation.

$$\delta^2 g = \frac{E[|u|^2]}{B^2} = \frac{6\sigma_0^2}{A^3 T_0^3} + \frac{6\sigma_0^4}{A^4 T_0^4} \quad (36)$$

$$\delta^2 f = \frac{E[|u|^2]}{B^2} = \frac{360\sigma_0^2}{A^3 T_0^5} + \frac{360\sigma_0^4}{A^4 T_0^6} \quad (37)$$

As can be seen in formula (36) (37), the estimating mean square error of start frequency and FM rate is decreases with observation time increasing, so, it is useful to improve the signals parameter estimation accuracy by increasing the observation time. It has been shown to be equivalent to the generalized likelihood ratio test (GLRT) and maximum likelihood estimator (MLE) in detection and estimation of LFMCW signals

4. Conclusion

This paper analyzes the advantages and disadvantages about LFMCW signals parameter estimation by FRFT, and proposes a multi-periods LFMCW signals parameter estimation algorithm basing PFrFT algorithm, the conclusions as follows:

(1) When using FRFT to estimate multi-period LFM CW signals parameters, there are a plurality of peaks appearing in converting figure, and the SNR do not increase when the observation time is more than one period. It means that FRFT is not optimal multi-period LFM CW signals parameter estimation algorithm.

(2) PFRFT output SNR is $\frac{SNR_m^2 \cdot T_0^2}{2T_0 \cdot SNR_m + 1}$, When the sample SNR is large enough, The PFRFT output SNR at

the true parameter values varies approximately linearly with the signal observation time

(3) In the noise environment, estimating the multi-period LFM CW signals parameters with PFRFT is unbiased estimates.

(4) It is solved the cross-term problems of WVD, the resolution problems of SPWVD and TFDS, inadequate use of the sample data of FrFT.

References

- Chen, P., Hou, C. H., & Liang, Y. H. (2007). Research on Discrete FrFT-Based Detection Algorithm for LFM Echo of Underwater Moving Target. *ACTA Armamentarii*, 7(28), 834-838.
- Deng, B., Tao R., Ping, D. F., & Ma, L. (2009). Moving-Target-Detection Algorithm with Compensation for Doppler Migration Based on FRFT. *ACTA Armamentarii*, 10(30), 1304-1308.
- Geroleo, F. G. (2012). Brandt-Pearce M. Detection and estimation of LFM CW radar signals. *IEEE Transactions on Aerospace and Electronic Systems*, 48(1), 405-418. <http://dx.doi.org/10.1109/TAES.2012.6129644>
- Geroleo, F., G., & Brandt, P. M. (2010). Detection and estimation of multi-Pulse LFM CW radar signals [C]//Radars Conference. Washington, DC: IEEE, 1009-1013.
- Liu, F., Sun, D. P., & Huang, Y. (2009). Fast parameter-estimation of LFM signals based on improved combined WVD and randomized hough transform. *ACTA Armamentarii*, 30(12), 1642-1646.
- Liu, F., Xu, H., Tao, R., & Wang, Y. (2012). Research on resolution among multi-component LFM signals in the fractional Fourier domain. *Science China*, 2(42), 136-148.
- Liu, J. C., Liu, Z., & Wang, X. S. (2007). SNR Analysis of LFM signals with gaussian white noise in Fractional Fourier Transform domain. *Journal of Electronics & Information Technology*, 10(29), 2338-2340.
- Millioz, F., & Davies, M. E. (2011). Detection and segmentation of FMCW radar signals based on the chirplet transform[C]//2011 IEEE International Conference on Acoustics, Speech and Signals Processing. Prague, CS: IEEE, :1765-1768.
- Peter O'Shea. A. (2004). Fast Algorithm for Estimating the Parameters of a Quadratic FM Signals. *Ieee Transactions on Signals Processing*, 2(52), 385-393. <http://dx.doi.org/10.1109/TSP.2003.821097>
- Qu, Q., & Jin, M. L. (2009). Adaptive FrFT Based Chirp Signals Detection and Parameter Estimation. *Journal of Electronics & Information Technology*, 12(31), 2937-2940.
- Wang H. Y., Qiu, T. S., & Chen, Z. (2008). Nonstationary Random Signals Analysis and Processing. Beijing: National Defense Industry Press, 311-313.
- Wang, R., & MA, Y. (2014). DOA Estimation of Wideband Linear Frequency Modulated Pulse Signals Based on FrFT. *ACTA Armamentarii*, 2(35), 207-213.
- Wen, J. Y., Zhang, H., & Wang, Y. (2012). Parameters estimation algorithm of LFM pulse compression radar signals. *Transactions of Beijing Institute of Technology*, 32(7), 746-750.
- White, P. R., & Locke, J. (2012). Performance of methods based on the FrFT for the detection of linear frequency modulated signals. *IET Signals Processing*, 6(5), 478-483. <http://dx.doi.org/10.1049/iet-spr.2011.0189>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

Logic of Interval Uncertainty

Evgeny A. Kuzmin¹

¹ Chair “Corporate Economics”, Ural State University of Economics, Ekaterinburg, Russia

Correspondence: Evgeny Kuzmin, Chair “Corporate Economics”, Ural State University of Economics, Str. March 8, 62, Ekaterinburg, 620144, Russia. Tel: 7-906-800-0070. E-mail: KuzminEA@gmail.com

Received: July 13, 2014 Accepted: July 16, 2014 Online Published: August 28, 2014

doi:10.5539/mas.v8n5p152

URL: <http://dx.doi.org/10.5539/mas.v8n5p152>

Abstract

The scientific category of uncertainty refers to that group of terms, an interpretation of which is not unambiguous and exact. In non-eliminability of the category soft content barrier there is an objective transition to the interval uncertainty. This research is an attempt to solve the issue of estimating the interval uncertainty based on methods of a logical analysis and a comparison. The approach presented by the paper is opposed to known methods of a mechanical selection of values following a given function. In the course of the research, there has been introduced a concept of the “tension uncertainty” for scientific use. Overall results obtained from the research allow calculating values of the interval uncertainty and assess their quality. The scientific competency of methods is achieved in theoretically tested solutions.

Keywords: interval uncertainty, fuzziness in economics, tension uncertainty, information entropy, uncertainty errors

1. Introduction

The intervality is an inherent property of parametric systems that function in a stochastic space. It is difficult to imagine that the economic or any other system possesses exact knowledge regarding a dynamics nature of its basic characteristics. Most often, this knowledge or, rather, ideas of distributed probabilities are approximate. In other words, they have a fuzzy nature. An indubitable fact is that “standard cases” for social systems are not typical. On the contrary, *uncertainty conditions are typical*. This conclusion does not require any proof and is present as a refrain in multiple academic papers, among which, for example, research by Parygin (1978), Leshkevich (1994), Afanasyev (1975) etc. are worth mentioning.

A quintessence of epistemological generalization is an idea saying that the inform ware for a managerial solution might look as a range from “the near empty to the almost comprehensive one” according to Afanasyev (1975). It seems that they are “almost” exact knowledge and understanding of a case at hand that do not allow the equally exact uncertainty to appear, the range of which is also large-scale starting from “almost unambiguous to an almost infinite number of alternatives” (Afanasyev, 1975). At the same time, regarding both well-known point entropy (an idealized standard), and its interval image, a starting point for scientific formalization and methodological classification of the uncertainty theory is a visualized assessment of their observability in the economic space. This research has been put in contrast to known approaches of mechanically selected uncertainty values on the interval. This research was designed to identify the uncertainty interval in a logical analysis of the function, take into account antinomy properties of entropy and consistency requirements.

2. Literature Review

Multiple academic papers have addressed complex systems in terms of the uncertainty. Scientists have paid their attention to solving various fundamental and applied tasks, where the entropy intervality is taken for granted, i.e. basic conditions, under which there is a progressing imaginative or real experiment. Thus, Kreinovich et al. (2011), Xiang (2007), Nguyen et al. (2008) have used terms of the interval uncertainty to refine estimates for statistical characteristics. Their Russian colleagues have made similar attempts. Panyukov and Latipova (2008) in interval estimates of the uncertainty will explain an equilibrium in the J. von Neumann’s model of an expanding economics. Thus, numerous aspects of social and economic interaction strongly emphasize an interval nature of these phenomena development, which is quite natural.

A separate corpus of scientific papers deals with issues of network optimization for complex systems. The interval uncertainty therein is inter alia a consequence of incomplete measurements, such as in the paper by Shen

et al. (2011), where with the interval-matrix approach there is one of the earliest attempts to describe the uncertainty of an internal state of the network; or as a result of the uncertainty (or incompleteness) that relates to a constraint force between nodes of the system. Deepening a range of problems, Shang (2011) comes closer to an answer to the similar question, but for scale-free networks yet. A main emphasis is made on exploring sustainability influenced by the "gray information", i.e. the information comprehensive in its scope of objects (nodes within the system), but inexact. It is this inexactness that produces the interval of fiducial perception. A certain interest under this direction is generated by research in interval-dependent and interval-independent complex dynamical networks. There will be name worthy papers by Li and Yue (2010), as well as Li, Wong and Tang (2012), etc. Both publications do not only point out to an influence of the subordination factor between nodes of the network, directly preconditioning dependences, but also the time factor. Transitions from one mode to another with a certain probability produce mixed time delays aggravating the system uncertainty.

The paper by Fortin, Zieliński, Dubois and Fargier (2010) has recently summarized the issue of the interval uncertainty for the theory of complex networks. This research revises the issue of PERT-scheduling with the incomplete information, seemingly well-known following the Kelley's critical path method (1961). Oddly enough, but vagueness in time distribution to execute tasks, a moment to start the ones and finish the others have for a long period of time remained outside the research attention in terms of the interval uncertainty. There is a particular emphasize on the fact that actually distribution of probabilities allows simulating variability of repetitive tasks, but not the uncertainty due to a lack of the information (e.g., Dubois, Prade and Smets (1996), Ferson and Ginzburg (1996), et al.). Only as an approximate solution to the mentioned range of problems we consider approaches by Chanas and Kamburowski (1981), Hapke, Jaskiewicz and Słowiński (1994), Loostma (1997), McCahon and Lee (1988), as well as Prade (1979) and Rommelfanger (1994).

However, the uncertainty cannot and must not only act as a condition, under which there are economic or other processes in progress. With insufficiently elaborated issues of calculating the interval uncertainty in methods of a logical choice, their methodological support is a fundamental scientific challenge. A solution to this challenge is in line with a proposed viable alternative, where initial terms of the uncertainty have not been set by parameters of a case, but become an independent object to be scientifically explored.

The first step to compile in an estimate way the interval uncertainty is a probability. It is an available *range of probabilities* for each alternative to be implemented in their assembly that leads to the entropy of a dimension value to appear. On the other hand, with apparent simplicity of the approach, implying an everyday substitution of probability values, boundaries of the interval uncertainty lie in a different plane. A cause for this lies in a property of antinomy.

An antinomy property of the uncertainty, whereby the entropy of a single instance may be located in a negative area (for more details see Kuzmin (2012)), is based on a displacement and asymmetry of the uncertainty function. The antinomy nature explains all the complexity of logical calculus for the interval uncertainty, when given boundaries of the certainty in the totality of values contain the internal range of the interval uncertainty. It is a search for it that this research is focused on. As a result, the above-mentioned antinomy specifics makes three common cases that call for matching: a full match and overlap, divided into the matching right and left parts. As a result, two combined area appear: the first one is at a distance from the lowest limit up to $1/e$ (basis of the natural logarithm); the second, on the contrary, is at a distance from $1/e$ up to the top limit of the probability (see *Figure 1*). An upper extremum of the entropy (calculated per alternative) corresponds to the permanent constant value of the probability, causing both non-linearity and function displacement at the same time. Thus, *a major scientific challenge shapes itself, when the available interval of the probability that characterizes case fuzziness, within its boundaries does not correspond to the uncertainty interval.*

Another, by no means unimportant, issue is the content of probability intervals. Surely, the range might be both continuous and discontinuous, and dotted. Recognising the interval as a continuous one is an assumption for this paper. In this regard, the view of Kyburg (1989) and Neapolitan (1996) is amazing. They share an opinion that probability intervals can always be reduced to point wise values. However, to our mind, this is not always the case. Depending on requirements of refinement and measurement accuracy as such, point wise values cannot be applied. There are natural barriers preventing decomposition of a probability line into point wise factors. Papers by Kyburg (1989) and Neapolitan (1996) do not mention this, but each probability only appears and lives within the framework of an adopted coordinate system, where a numerical order of the probability has a crucial value. As a result, one might reasonably state that the point look cannot be a worthy replacement for the linear probability (unbroken), even if it had an interval limit. A disclaimer expressed by some scientists, such as Sternin and Shepelev (2012), saying that "...an interval estimate contains a continuum of possible point wise implementations" only partially resolves existing contradictions in views.

Halpern and Fagin (1992) at the same time pay attention to the theory of probabilities, which explains a mechanism to join the interval. Later even Yager and Kreinovich (1999) suggested and gave grounds for a formula of statistical estimates for the so-called secondary interval. It is indicative that to get such estimates one needs to know the distributive law, to identify which within a fuzzy set “with a multi-valued scale of truth” (Borisov *et al.*, 1989) is at least difficult. From Parsons’s (2001) point of view, it proves again an ambiguous nature of the uncertainty, for which no specific characteristic and quite accurate values are typical. Somewhat earlier, Ang and Tang (1975) came to such conclusion, when they reasonably assumed that “the information can be often expressed in terms of the lower and upper variables of the limit”. This implies that the probability like the information possesses constraints, going beyond which can be a trouble. One might say that *thereby the uncertainty should remain unknown to some extent.*

Returning to Parsons's ideas (2001), it is worth paying attention to his statement of the interval interpretation, "the presentation interval, whether this is a probability or not, is intuitively compelling in making the inexactness clear." An undisputed nature of the statement should be read with a limitation that clarity appears in understanding the essence of arising divergences. At the same time, Parsons's considers the interval uncertainty a result of the fuzzy probability manifested itself. Therefore, major efforts by researchers address the issue of "how to make assumptions of the partial information that have not been clearly defined." Thus, making corrections to the interval uncertainty or in other words reducing it to the point uncertainty to some extent, some scholars see in processing the information itself. And according to Parsons's statements, “these systems can cope with a situation... even in the absence of necessary and partially defined probabilities...” It seems that such the way would not be without results, though manipulations with the interval uncertainty to optimize management decisions taken could be more efficient. After all, generating new information to narrow down a gap of the uncertainty is linked, on the one hand, to arising ambiguity of new data, and on the other – to the complexity of artificial refinement of the state. A compromise when dealing with the source information is achieved in a revision of probability intervals. This conclusion made by Loui (1986) is shared by Parsons's and consistent with the authors' hypotheses regarding the interval normalization.

Considerations on a heuristic evaluation of probability distribution are separately worth looking at. Problems concerned in the work by Beliaev and Makarov (1974), are solved using a principle of the maximum entropy as a common method for these purposes, which, nevertheless, with all internal consistency of the approach, does not avoid a number of disadvantages. Dichotomy in a search for favourable and hence *inter alia* labelling unfavourable distribution is counterbalanced using the confidence interval, when a case and a condition are specific enough, but retain their development uncertainty. At the first glance, such a conclusion sounds like a kind of paradox. However, specificity lays in a structure and probabilities of alternatives, which have predefined values. The *confidence interval mostly influenced by the idiosyncratic “errors of uncertainty” (in terms of the alpha and beta errors) seems to be* a kind of prologue to studies in distribution fuzziness. The introduction to the research methodology refines meaningful aspects differentiated by types of the uncertainty errors. To the author’s mind, one needs to distinguish between alpha-, beta- and gamma-errors of the uncertainty (so-called standard errors), that set out possible errors in perception and, mainly, in its quantity assessment. We should clarify some of their characteristics. An essence of the alpha- error depends on making a set of alternatives recognised as basic. Subjective deviation from multiple possible outcomes from certain scenarios (hypotheses) shows the additional uncertainty, a value of which is calculated by relating two states - actual and conditionally ideal. The Beta-error of the uncertainty goes to a quality side of alternatives and implies fuzziness in defining likelihood or frequency for each outcome. A nature and an extent of different errors’ participation in assessing a numerical value of the uncertainty are presented with more details in other research, e.g. Avdiysky and Kuzmin (2013).

3. Methods

3.1 Reducing Probabilities

Stochastic dynamics within complex systems in fuzzy distribution sets certain limits, i.e. requirements to intervals of the alternatives’ probability. The single instance, being internally consistent by formal unity criteria must be harmoniously integrated into an overall ensemble. This suggests a need in a cross-validation of alternatives’ probabilities, designated within the research as the convergence requirement or convergence condition. Objectively available limitations related to the information and the uncertainty also lead to this. Starting from a paradigm objective, offered by Lindley (1987) (not without assumptions according to Elishakoff and Ohsaki (2010)), an idea of drafting a “single instance” with acceptable alternatives, that is such a hypothetical event, which has its unitary probability in its possible implementation to assess and calculate the uncertainty. A range of intervals in this case is adjusted by excluding those probabilities that within the ensemble

of alternatives go beyond pragmatic affordability.

The above-mentioned conventionally ideal case, regarding to the alpha-error calculation, is also suitable to explain the proposed condition, "... based on the idealized interval adjustment, there is a provided tool to expand coverage to include its true [frontiers - *scholium*]" (Spall, 2002). Its essence lies in a principle check of extreme instance probabilities – those of the upper and lower thresholds. Scientific grounds for a methodological *imperative of probabilities convergence (or reducing)* is based on the information coverage. Completeness of alternatives should be a hypothetical case, which occurrence probability is equal to one. Verification of ranges assumes both their targeted expansion, and contraction. Herewith, *a process of such a transformation can go either from the upper or lower limit of the probability interval.*

Hereby, there are actually only a few given initial parameters set: a number of alternatives making up the ensemble and the probability matching them, though either the maximum or the minimum one. Only to confirm this rule, there are exceptions, when initial probability intervals meet a "reducing" requirement. By the way, Elishakoff and Ohsaki (2010) ground a similar condition in the form of optimization and anti-optimization models, that stand for options of best and the worst scenario approaches. Drawing parallels between conclusions and ideas by Elishakoff and Ohsaki (2010), one can quite reliably refer to the optimization model as an approach of a noncompliant attitude towards the uncertainty. An opposite picture concerns the anti-optimization model, where there is a tolerant perception of the entropy. In both cases, models operate with probabilistic values, which, even with expanding or contracting intervals require reducing. The logic of probability intervals adjustments is presented below.

Taking a_1 and a_2 for the lower and upper bounds of the p_i probability range, respectively for the i -th alternative within the hypotheses ensemble, one can represent a mathematical probability convergence (or reducing) record, that looks like as follows:

Terms		Adjustment	
Data	Restrictions	At the lower bound	At the upper bound
$p_1(a_1; a_2)$	$a_1 \neq a_2$		$a_1 = a_2 - \left(1 - \frac{n - \sum_{i=1}^n a_2}{n - 1} \right)$
$p_2(a_1; a_2)$	$a_1 < a_2$		
$p_3(a_1; a_2)$	$a_{1,2} \neq 0$		or
$p_{\dots}(a_1; a_2)$	$a_{1,2} \neq 1$	$a_2 = a_1 + (1 - \sum_{i=1}^n a_1)$	
$p_{n-1}(a_1; a_2)$	$\sum_{i=1}^n a_1 < 1, \sum_{i=1}^n a_2 > 1$		$a_1 = a_2 - \left(\frac{\sum_{i=1}^n a_2 - 1}{n - 1} \right)$
$p_n(a_1; a_2)$	$0 < a_{1,2} < 1$		

It is amazing that for all cases of the probability interval reducing there is a coherence equation identified by the research. Its statement says, *the sum of probabilities for the lower bound is equal to a converted value of the sum of probabilities of the upper bound, which depends on a number of alternatives –*

$$\sum_{i=1}^n a_1 = \frac{\sum_{i=1}^n a_2 - 1}{n - 1}. \tag{2}$$

The abovementioned assumes that initially set probability ranges might be mutually non-coherent. Reducing probabilities solves this problem. A contraction process results in making the coherent probabilistic logic. But methodological difficulties do not end at this point. Here one should remind that gaps in hypotheses distribution do not allow getting reliable estimates, i.e. the uncertainty would be a priori incomplete without taking into account the additional complexity of the choice. Guo and Tanaka (2007), with their ideas concerning mostly the economic profit, make a fair conclusion that "an optimal order is obtained based on a partially ordered array of interval expectations" (Guo and Tanaka, 2010).

3.2 Search for “Single Instance” Criteria. Combination Uncertainty

Having obtained representative intervals for probabilities per each alternative included in the uncertainty area, we have a scientific challenge of a different kind, a solution to which involves a search for equivalent or close options on a contour of the entropy function.

Apodictic ideas of optimizing the uncertainty situation suggests that with quantity similarity or equality of values in intervals of different alternatives there is a possibility of cases combination, where a level of the interval uncertainty remains unchanged. As a result, one can reasonably say of heterogeneity within the interval entropy. On the one hand, this results from the function specifics. On the other hand, this results from an objective need to “gain to fit” the probability, as relevance of the interval entropy estimates is provided by obtaining the single instance, which in its turn cannot be only composed of the lower or the upper boundaries of the range of alternative probabilities.

A single-directed search for a range of the interval uncertainty (by reducing the probability) makes us only a bit closer to a founded hypothetical scenario. A conceptual approach to an assessment of the interval uncertainty, proposed by the paper, assumes a two-level comparison: the first level influences private alternatives, while the second one aims at probabilities harmonization, when the single event is made with upper and lower bounds of the uncertainty. An outcome of assessment calculations is the dual uncertainty that in the first case is referred to as the uncertainty of combinations (intermediate value), while in the second case as the actual interval uncertainty.

One can trace overlapping “surplus” or “shortage” in possibilities up to a single constant calculating the uncertainty of combinations. A starting point for this is a collatable analysis of alternatives. The technique allows looking at those alternatives, which within the reduced probability of intervals (in threshold values) have already had some maximal or minimal uncertainty, but have not still met a requirement of the single instance. Hereinafter, a search for the interval uncertainty will be focused on a particular combination of probabilities. At the same time, a structure of hypotheses, according to the technique, is differentiated into four groups depending on a correspondence between the upper and lower threshold of probabilities per alternative and the given conditions (3). It is worth saying again that the total moment probability for the uncertainty of combinations might be either over or lower than one (it can seldom have values equal to one).

$$H_{\text{extra comb}}^{\max} = H(A) + H(B) + H(C) + H(D)$$

$$H_{\text{extra comb}}^{\max} = \begin{cases} a_1 < \frac{1}{e}, a_2 \geq \frac{1}{e} : H(A) = m \left(-\frac{1}{e} \right) \log_b \left(\frac{1}{e} \right) \\ a_1 < \frac{1}{e}, a_2 < \frac{1}{e} : H(B) = \sum_{k=1}^k -a_2 \log_b (a_2) \\ a_1 > \frac{1}{e}, a_2 > \frac{1}{e} : H(C) = \sum_{l=1}^l -a_1 \log_b (a_1) \\ a_1 = \frac{1}{e}, a_2 > \frac{1}{e} : H(D) = t \left(-\frac{1}{e} \right) \log_b \left(\frac{1}{e} \right) \end{cases} \quad (3)$$

where $H_{\text{extra comb}}^{\max}$ is the maximum uncertainty of combinations; $H(A), H(B), H(C), H(D)$ stands for component grouped parts of the uncertainty; a_1, a_2 stands for the lower and upper bounds of the range of probabilities of the i -th alternative; m, k, l, t is a number of matches under the terms of calculation.

A situation is somewhat different with finding the *minimum uncertainty of combinations*. With the arched function, the minimum of the i -th alternative entropy can be located both on the right part (increasing area), and on the left (decreasing area) part (see Figure 1). Herewith, it is worth paying attention to displacement of the function itself. It is necessary to make a few remarks concerning this.

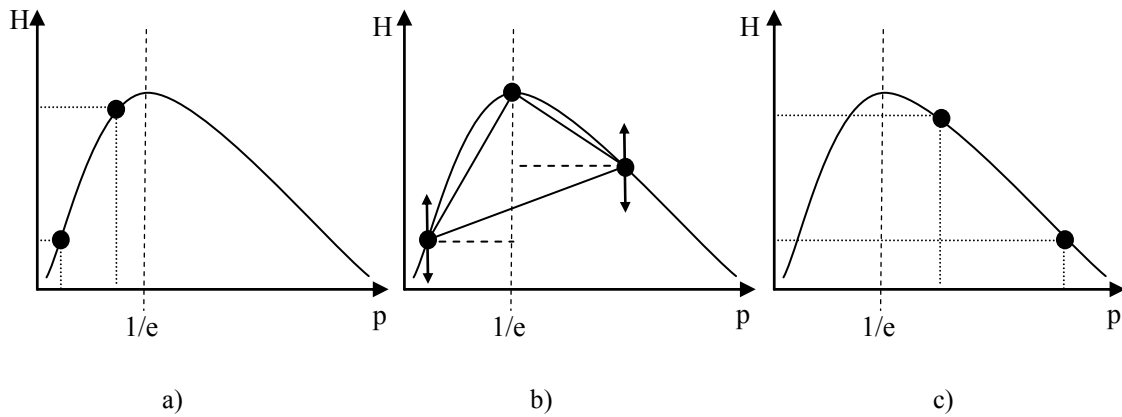


Figure 1. Common cases and fields of matched probabilities when assessing the interval uncertainty in decomposition of private alternatives: a) match on the right side of probabilities distribution; the minimum and maximum uncertainties fully correspond with boundaries of the probability interval; b) complete match; identifying the probability is associated with an issue of uncertainty comparison, when the maximum value is known attributable to 1/e, whereas the minimum is unknown, which locates either to the right, or to the left of the central axis; c) match in the left part of the probability distribution; the maximum and minimum uncertainties are inversely proportional to relations between probability boundaries.

A main thesis is that different values of the probability can make the identical uncertainty. The approach, proposed by the author, proves that firstly with the identical uncertainty deviations of probabilities from a central axis are non-equal. Secondly, these probabilities make the dual minimum of the uncertainty, but in calculations, only one of them is taken into account, i.e. that one that to the more extent makes a set of alternatives closer to the single instance. Thirdly, deviation on the right side (β) in a function polygon is always over than deviations on its left part (α). Hence, formalized terms for the observable minimum of the uncertainty are complemented with a special factor ($k_{\alpha,\beta}$), which accentuates a move from a case of the entropy identity (equality) in case of a numerical non-equality of probabilities for considered alternatives.

Based on conclusions presented, a process to obtain the *minimum uncertainty of combinations* looks like the following:

$$H_{\text{extra comb}}^{\min} = H(A) + H(B) + H(C) + H(D) + H(C \vee D) \tag{4}$$

$$H_{\text{extra comb}}^{\min} = \left[\begin{array}{l} a_1 \geq \frac{1}{e}, a_2 > \frac{1}{e}, a_1 < a_2 : H(A) = \sum_{m=1}^{m'} -a_2 \log_b(a_2) \\ a_1 < \frac{1}{e}, a_2 \leq \frac{1}{e}, a_1 < a_2 : H(B) = \sum_{k=1}^{k'} -a_1 \log_b(a_1) \\ a_1 < \frac{1}{e}, a_2 \geq \frac{1}{e}, (a_2 - \frac{1}{e}) > (\frac{1}{e} - a_1 k_{\alpha,\beta}) : H(C) = \sum_{l=1}^{l'} -a_2 \log_b(a_2) \\ a_1 < \frac{1}{e}, a_2 \geq \frac{1}{e}, (a_2 - \frac{1}{e}) < (\frac{1}{e} - a_1 k_{\alpha,\beta}) : H(D) = \sum_{t=1}^{t'} -a_1 \log_b(a_1) \\ a_1 < \frac{1}{e}, a_2 > \frac{1}{e}, (a_2 - \frac{1}{e}) = (\frac{1}{e} - a_1 k_{\alpha,\beta}) : H(C \vee D) = \sum_{w=1}^{w'} (-a_1 \vee -a_2 \log_b(a_1 \vee a_2)) \end{array} \right]$$

where $H_{\text{extra comb}}^{\min}$ is the minimum uncertainty of combinations; m', k', l', t' is a number of matches under the terms of calculation.

3.3 Interval Uncertainty

A scientific search for the interval uncertainty goes along with a verification of previously obtained moment probabilities for the hypothetical event, which appear in terms of calculation concerning the maximum (3) and

the minimum (4) uncertainty of combinations. Probabilities make a matrix, which is a basis for a secondary comparison. In particular, for the maximum, an equiprobable instance becomes such a comparison criterion ($p=1/n$).

3.3.1 Maximum of the Interval Uncertainty

The maximum interval uncertainty, as in other calculation options, assumes making an acceptable scenario. To the author's mind, the process is largely connected to unification of manifested alternatives (equal probability). Herewith, as it is known, a numerical limit for the uncertainty is reached. All arithmetic of the method is divided into a number of stages, while a result of their implementation is an approximate estimate. A disclaimer on an approximate nature of calculations is due to the nonlinear dynamics of the uncertainty, allowing errors that is actually inherent to the standard error area.

1) The first step to obtain the actual uncertainty includes gradually converging probabilities. At this stage, the maximum entropy antinomy ($p=1/e$) opposes to the equal probability maximum. By comparing the previously obtained matrix for the uncertainty of combinations and the decrease delta, it is possible adjust to some extent the additive sum and take the situation closer to the single equiprobable instance. Indicative parameters, i.e. the decrease delta and the growth delta, are measured with the module. Actually, each of them shows possibilities of a change to the probability (per alternative), depending on set intervals that correspond to the principle of convergence, towards the single event. An essential assumption is an idea of elimination, saying that to reduce to the single event one needs to change the given probability (within intervals) for only one alternative (such calculation is made for all the alternatives).

$$P_{\text{real (interim)}}^{\max} = \left[\begin{array}{l} (P_i^{\max\text{comb}} - \Delta p_i \downarrow) \leq \frac{1}{n} : \frac{1}{n} \\ (P_i^{\max\text{comb}} - \Delta p_i \downarrow) > \frac{1}{n} : P_i^{\max\text{comb}} - \Delta p_i \downarrow \end{array} \right] \quad (5)$$

where $P_{\text{real (interim)}}^{\max}$ is an intermediate probability matrix for the maximum actual uncertainty in interval measurement; $P_i^{\max\text{comb}}$ is the probability of the alternative, corresponding to the maximum of the uncertainty of combinations; $\Delta p_i \downarrow$ is the probability decrease delta.

2) Mostly, the additive sum of the smoothed-out probability according to (5) to the full extent have not correlated proposed requirements. However, at the same time, new logic regularities appear that determine subsequent conversions. An obvious consequence is such a situation, in which there is a need for adjustment. With the alternative probability exceeding $1/n$ value (such difference is accumulated), a reduction value appears that makes an excess probabilistic load. Its smoothing-out happens at the next stage, where, firstly, those alternatives should be clarified, where probabilities of their occurrence are over than $1/n$. Secondly, one needs to use the upper and lower bounds of the set range to evaluate a potential of possible variations in values. For this, we introduce a number of variable and conditional parameters, intended to simplify perception of the calculations logic. All comparative conversions and calculations occur at the second methodological step.

A sequence of calculations can be presented as the following formalization:

$$y_1^{\max} = \left[\begin{array}{l} P_{\text{real (interim } 1)_i}^{\max} - \frac{1}{n} = 0 : a_1 \\ P_{\text{real (interim } 1)_i}^{\max} - \frac{1}{n} \neq 0 : P_{\text{real (interim } 1)_i}^{\max} \end{array} \right], \quad y_2^{\max} = \left[\begin{array}{l} P_{\text{real (interim } 1)_i}^{\max} - \frac{1}{n} = 0 : a_2 \\ P_{\text{real (interim } 1)_i}^{\max} - \frac{1}{n} \neq 0 : P_{\text{real (interim } 1)_i}^{\max} \end{array} \right]. \quad (6)$$

Then with similar conditions of variables y_1^{\max} and y_2^{\max} one needs to identify similar values and differentiate in pairs differences between them (matrix r_1^{\max} and matrix r_2^{\max}).

$$r_1^{\max} = \left[\begin{array}{l} y_1^{\max} = y_2^{\max} : y_2^{\max} \\ y_1^{\max} \neq y_2^{\max} : 0 \end{array} \right], \quad r_2^{\max} = \left[\begin{array}{l} y_1^{\max} = y_2^{\max} : 0 \\ y_1^{\max} \neq y_2^{\max} : y_2^{\max} \end{array} \right]. \quad (7)$$

The next variable (U^{\max}) is an indicator that tracks those alternatives, the probability of which will change in contrast to a number of hypotheses made earlier on the intermediate matrix (5). The author's approach involves a method to average remaining hypotheses, as only in this case the uncertainty is maximum. Herewith, so-called

unchanging alternatives become such not so much because of assessment preferences of the uncertainty of combination, as of being a result of exact compliance with interval analysis requirements.

$$U^{\max} = \begin{cases} r_2^{\max} = 0 : r_2^{\max} \\ r_2^{\max} \neq 0 : r_2^{\max} - \frac{\sum_{i=1}^n r_1^{\max} + \sum_{i=1}^n r_2^{\max} - 1}{\sum_{i=1}^n r_2^{\max} (r_2^{\max} > 0)} \end{cases} \quad (8)$$

3) The resulting matrix (8) completes the second step of intermediate comparative estimates and leads to the final stage, where there will be obtained a quantitative evaluation of the approximate maximum uncertainty for the interval. More precisely, it leads to determination of the moment probability, with which such the uncertainty may occur.

Thus, the matrix of the actual probability (the upper bound of the interval) makes such an instance, which was before referred to as a single, that is, in fact, the main requirement to comprehensive estimates for the uncertainty.

$$P_{\text{real}}^{\max} = \begin{cases} \sum_{i=1}^n (P_{\text{real}(\text{interim})_i}^{\max} - \frac{1}{n}) = 0 : \frac{1}{n} \\ \sum_{i=1}^n (P_{\text{real}(\text{interim})_i}^{\max} - \frac{1}{n}) \neq 0 : r_1^{\max} + U^{\max} \end{cases} \quad (9)$$

$$\text{wherein } \sum_{i=1}^n P_{\text{real}_i}^{\max} = 1.$$

At the same time, the explained approach with all advantages and its logical structure operates with a prerequisite that cannot avoid criticism. We mean averaging alternatives: even if variable probabilities of alternatives are only in one of the function areas, it is not possible to argue that their values placed at equal distances have uncertainty levels placed at the same distances between them.

However, strong arguments have remained in favour of the averaging idea (Note 1). The error in calculations caused by the function linearization is not as weighty to reject such an assumption. Previously marked features of the entropy function show that significance of the numerical error in probability averaging increases as we approach the extremum, or more precisely, a limit of the uncertainty. It is known that prevailing one or more alternatives, in which a frequency of manifestations is not identical, leads to occurrence of the case certainty. Averaging, on the contrary, makes the uncertainty and makes the choice more complex. It is quite *possible that to maximize the uncertainty, some alternatives not yet fixed within the scenario should have the probability equal to 1/n, whereas other alternatives will act as a compensating factor with the different probability* for a holistic combination of options to make the single unstance. Based on these and other reasons put forward in the course of the research, the methodical approach has kept an assumption of probability averaging understanding that an estimate of the maximum interval uncertainty is only approximate and different by a value of a corresponding error.

3.3.2 Minimum of the Interval Uncertainty

Making a pause at a point of giving grounds for a calculation of the maximum uncertainty intervally interpreted, one should address calculations for the range lower bound. Approaches to their estimate differ due to antinomy properties of the function. It is worth saying again that the minimum of the uncertainty is ambiguous and under specific accident circumstances, it is dual. Thus, the approach to the assessment assumes well-founded research in several basic conditions that allow eliminating duality contradictions.

It seems that the easiest condition for a differentiated comparison is convergence of probabilities. Based on a need to construct the single instance, criteria for the convergence condition are grouped within *three options* at least. The first implies that the additive sum of probabilities of alternatives included in the minimum uncertainty of combinations is equal to one. Herewith, further conversions and calculations are no longer required; the minimum actual uncertainty has been already achieved. The second and the third options are based on the opposite saying that the sum of hypotheses probabilities is inconsistent with the single instance, and hence the probability is either over, or lower than one. One needs to note that only one option reflects an actual situation with the uncertainty. To observe logic in the method presentation, it is useful to present mathematical formalization and explanation for each of them.

1) According to the abovementioned, refinement of the first option is only required because of a holistic nature of the method. Here, the sum of probabilities with the minimum uncertainty of combinations (4) have been equal to one that allows founded assuming on obtaining the interval entropy minimum herewith. A verification of the instance is rather formal.

$$\sum_{i=1}^n p_i(H_{\text{extra comb}}^{\min}) = 1 \mid \Rightarrow p_i(H_{\text{extra comb}}^{\min}) = p_i(H_{\text{real}}^{\min}). \tag{10}$$

A different picture emerges in a case when the additive sum of probabilities according to (4) is not equal to one. There is an objective need in value convergence and hence in covering its surplus as in the third option or a shortage as in the second option. It is worth referring to each option.

2) A search for the minimum of the interval entropy in the second option goes towards matching the uncertainty of critical levels (we mean multiple values of probabilities, i.e. of the lowest and upper boundaries of the interval, as well as of the variable 1/n and the constant 1/e) and the uncertainty of combinations for different hypotheses. A result of this is a notation of those alternatives, where the probability becomes as a kind of a fixed one. Its value gets one of values of a critical level, mostly of the lowest bound of the probabilities range. Remaining hypotheses (alternative) highlight a need in frequency “finish gathering” up to the single instance. A solution to this task lies in an area of identifying the so-called first-match (Note 2).

Formalizing the approach for the second option, as in other examples given above, involves calculating intermediate variables to fix various logical search requirements. According to the abovementioned, such an indicator includes the first match criteria. However, it follows reconfiguration of probability critical values. A mathematical expression for this procedure applied for this type of calculations is as follows:

$$P_{\text{lower limit}}^{\min} = \left[\sum_{i=1}^n p_{\text{extra comb}_i}^{\min} < 1 : p_{\text{extra comb}_i}^{\min} \right], \quad P_{\text{upper limit}}^{\min} = \left[\sum_{i=1}^n p_{\text{extra comb}_i}^{\min} < 1 : a_2 \right]. \tag{11}$$

New thresholds for ranges combine in themselves a value for the resulting probability for the minimum uncertainty of combinations and the upper probability of the specified interval. A need in such a transformation is caused by an objective need to cover a frequency deficit. Returning to the first match criterion, it will be useful to clarify the indicator figure (q^{\min}), based on which the criterion is defined:

$$q^{\min} = \left[\begin{array}{l} (p_{\text{lower limit}}^{\min} \log_b p_{\text{lower limit}}^{\min} - p_{\text{upper limit}}^{\min} \log_b p_{\text{upper limit}}^{\min}) > 0 : p_{\text{lower limit}}^{\min} \log_b p_{\text{lower limit}}^{\min} - p_{\text{upper limit}}^{\min} \log_b p_{\text{upper limit}}^{\min} \\ (p_{\text{lower limit}}^{\min} \log_b p_{\text{lower limit}}^{\min} - p_{\text{upper limit}}^{\min} \log_b p_{\text{upper limit}}^{\min}) \leq 0 : \text{error} \end{array} \right]. \tag{12}$$

The presented matrix (12) is an estimate of a difference between uncertainties according to (11). A search for the first match is performed with the minimum value among obtained values. It is this alternative, for which there will be acceptable an adjustment to the probability. Herewith, other hypotheses have remained in the same, unchanged form. Thus, a final structure of the alternatives probability to calculate the minimum of the interval certainty is calculated observing the following way:

$$p(H_{\text{real}}^{\min}) = \left[\begin{array}{l} q_i^{\min} = \min^1 \{q^{\min}\} \mid < q_i^{\min} >_1 : p_{\text{lower limit}}^{\min} + (p_{\text{upper limit}}^{\min} - a_1) \\ q_i^{\min} \neq \min^1 \{q^{\min}\} : p_{\text{lower limit}}^{\min} \end{array} \right], \tag{13}$$

$$\text{wherein } \sum_{i=1}^n p_i(H_{\text{real}}^{\min}) = 1,$$

where $\min^1 \{q^{\min}\}$ is the first order minimum within q^{\min} array of values.

3) The third option to estimate the minimum uncertainty comes from the fact that the additive sum of probabilities for hypothetical accident alternatives, on the contrary, is over one. This has resulted in an uncompromising search for the minimum entropy. However, the underlined difference does not underplay importance of using the method of the “first match”. In both cases, they identify such an alternative, which is less in its uncertainty inflection. In other words, such the hypothesis is remarkably consistent in its uncertainty dynamics. To say more, in its low minimum entropy spread within the given range. Because the uncertainty of

combinations has been an extreme manifestation of the minimum or maximum uncertainty, but has not yet been the single instance, this method seems to be justified and scientifically grounded.

To smooth categoricity in choosing the momentum probability value, possible because of equal ranges between alternatives, the third option assumes using *not only the method of the “first match”, but also the “the last match”* (Note 3). Moreover, the method of the “first match” becomes a more rigid form as of the only match. Each method demarks possibly repeated or identical ranges of hypotheses probabilities. Assuming that alternatives are mutually distinct, the first match method, in fact, shows a single hypothesis, for which there will be appropriate a change to the probability with minimal consequences. A reverse situation occurs when some or all of alternatives are similar. A need occurs to reduce probabilistic tension by influencing only some of these alternatives - apart from the first until the last match. However, with absent identically similar hypotheses there is an exception to rules. Obviously, the set of alternatives is not homogenous; it can include hypotheses with the minimum uncertainty in both the right and the left part of the function. In this, a need occurs to apply these two methods in combination. Their incorporation marks a new synthetic method, i.e. the method of *“inter congruence”*. Adjustments within it only goes for those alternatives that follow according to the minimum criterion specified with index q^{\min} (Note 4). This occurs until the moment probability sum for the scenario is equal to one. A limit number of such points depends on a number of numeric values $q^{\min} (W_{q^{\min}})$, which are reduced by an alternative. Decrease in a number of selected hypotheses depends on maintaining the probability balance. Its essence is in leaving a “free alternative”, which takes probabilistic values based on compared uncertainties. Such alternative is recognized as the last of the ranked ($Rank_i$) set of values for the q^{\min} indicator (ascending from 1 to n). It is important that ranking goes for all the hypotheses without no repetitions in the rank value. Based on the assumptions made above, the initial probability of boundary values in the third option also faces changes. Since the additive sum of probabilities is within a zone of excess tension, the lowers limit replaces the top one: we consider a possibility for optimal reducing in probabilities. Then, we obtain limits as follows:

$$p_{\text{lower limit}}^{\min} = \left[\sum_{i=1}^n p_{\text{extra comb}_i}^{\min} > 1 : p_{\text{extra comb}_i}^{\min} \right], \quad p_{\text{upper limit}}^{\min} = \left[\sum_{i=1}^n p_{\text{extra comb}_i}^{\min} > 1 : a_1 \right]. \quad (14)$$

Herewith, one should distinguish probabilities (14) and boundary values obtained as expressed in the equation in the second option (11). Their identical notation has been kept due to an exclusive situation, when a calculation is only possible under one of options.

The following equation establishes the probability-decrease delta (s) per alternative regarding the previously established hypotheses within the minimum uncertainty of combinations.

Method	Formalization
	$S_1 = \left[\begin{array}{l} q_i^{\min} = \min^1 \{q^{\min}\} < q_i^{\min} >_1 : p_{\text{lower limit}}^{\min} - a_1 \\ q_i^{\min} \neq \min^1 \{q^{\min}\} \vee q_i^{\min} = \min^{m>1} \{q^{\min}\} : 0 \end{array} \right] \quad (15)$
The First Match (single match by the first minimum)	<p>Happens only when the following equation is implemented:</p> $F_{S_1} = \sum_{i=1}^n p_{\text{lower limit}_i}^{\min} - \sum_{i=1}^n S_{1_i} = 1$ <p>where $< q_i^{\min} >_1$ is the first-order cyclic group, denotes a term, under which the equation is only implemented once until the first match; \min^1 is the first minimum of the array.</p>
The Last Match (multiple matches, except the first one by	$S_2 = \left[\begin{array}{l} q_i^{\min} = \min^{m>1} \{q^{\min}\} : p_{\text{lower limit}}^{\min} - a_1 \\ q_i^{\min} \neq \min^{m>1} \{q^{\min}\} \vee q_i^{\min} = \min^1 \{q^{\min}\} < q_i^{\min} >_1 : 0 \end{array} \right]$ <p>Only occurs when the following equation is implemented:</p>

Method	Formalization
the minimum difference)	$F_{s_2} = \sum_{i=1}^n p_{\text{lower limit}_i}^{\min} - \sum_{i=1}^n s_{2_i} = 1$
The Inter Matches (matches by the first, second etc. minimum - till the last included)	$S_3 = \left[\begin{array}{l} q_i^{\min} = \min^{m < n} \{q^{\min}\} \vee \text{Rank}_i(q_i^{\min}) < W_{q^{\min}} : p_{\text{lower limit}}^{\min} - a_1 \\ q_i^{\min} = \min^{m = n} \{q^{\min}\} : 0 \end{array} \right]$ <p>Only occurs only when the following equation is implemented:</p> $F_{s_3} = \sum_{i=1}^n p_{\text{lower limit}_i}^{\min} - \sum_{i=1}^n s_{3_i} = 1$

Each method seems to generate its proper solution. Moreover, as it was stated above, the choice is only made regarding one of the methods. Thus, the actual probability for the minimum uncertainty within the interval can be obtained as follows:

$$p(H_{\text{real}}^{\min}) = \left[\begin{array}{l} F_{s_1} = 1 \mid \langle \{F\} \rangle_1 : p_{\text{lower limit}_i}^{\min} - s_{1_i} \\ F_{s_2} = 1 \mid \langle \{F\} \rangle_1 : p_{\text{lower limit}_i}^{\min} - s_{2_i} \\ F_{s_3} = 1 \mid \langle \{F\} \rangle_1 : p_{\text{lower limit}_i}^{\min} - s_{3_i} \end{array} \right] \tag{16}$$

wherein $\sum_{i=1}^n p_i(H_{\text{real}}^{\min}) = 1$.

An outcome from the approach presented by the author is a set of two probability matrices, i.e. for the minimum (16) and for the maximum (9). Complexity and multiplicity of a logical search for the interval uncertainty depend on specifics in obtaining extrema for the function of the information entropy. To some extent, the challenge lies in hidden valuation approaches of the interval uncertainty. On the one hand, it has been taken for granted, i.e. environmental conditions under which an economic subject exists. However, on the other hand, it is sometimes not enough to be aware of fuzziness in the entropy distribution. Assumptions of certain disclosure of intervals as such are additional information actually not available to the subject. With absent properly elaborated approaches to evaluate the interval uncertainty (as opposed to approaches of mechanically selected values), such scientific objectives have been rent from the real life and oversimplified. We have mentioned above of a danger from simplification when obtaining the uncertainty and this reinforces relevance of this research.

On this occasion, it would be appropriate to refer to considerations made by Shary (2012), who in a definition for the interval uncertainty says of its important role for science. He comes to a fair conclusion that “the interval analysis and its specific methods have ... the highest significance in issues, where the uncertainty and ambiguity arise from the outset and are an integral part of task setting process”.

4. Discussion

The interval analysis of the uncertainty in presented estimation methods puts a theoretical and methodological issue of applied usage of results. Contradictions, on the one hand, concern an objectively available natural uncertainty limit for the equiprobable outcome. On the other hand, the range of the interval uncertainty, as the wider the range of the minimum and maximum entropy is, the more the case approaches a status of one with a totally unknown outcome. This really causes displaced characteristics of the uncertainty, both with initial fuzziness of the system parameters, and their ultimate specificity. The obtained results for the interval estimation might relatively not highly efficient being applied in practice. The main cause for this is the mentioned uncertainty, but not related to a description of a choice situation as such, but to a confirmation of the “quality” of uncertainty in this choice. The mentioned methodological challenge was referred to as a challenge of uncertainty convergence.

4.1 Methodological Issue of Uncertainty Convergence

Referring with more details to the issue, we need to explain its content. For the interval analysis, the uncertainty lies between the minimum and maximum limits. It can be conventionally called a baseline case. Nevertheless, boundaries of the uncertainty besides their calculated values have quite natural restrictions. The minimum uncertainty cannot obviously be below zero, while the maximum cannot be over the uncertainties for the equiprobable outcome. If a restraining barrier for the minimum uncertainty is neglected for a while, its maximum limit causes the convergence challenge. Indeed, the greater the probability range is, the greater the uncertainty range is. *A further expanding probability only aggravates the case, where the lowest boundary of the interval has its shift.* A logical consequence from this is a fact that the interval uncertainty cannot describe to a proper degree all the complexity and ambiguity of an expected choice.

At a turn of limit values for the minimum and maximum interval estimates, the uncertainty overlooks conceptual vagueness related to the uncertainty itself. Indeed, the uncertainty then does not allow making a reliable picture of a case. Its values fall in natural constraints, where there are descriptions for all possible options for a case to be resolved, even hardly probable. This emphasizes imperfection and limitations in theoretical and methodological solutions to the interval uncertainty. A logical conclusion suggests itself saying that *its assessments, on the one hand, are only fair for certain ranges of probabilities for the ensemble of alternatives, thus confirming a conclusion of redundancy and pessimism in an array of ensured responses (according to Shary (2007). On the other hand, a new type of the uncertainty appear (like a technical parameter).* Semantic formalization of such a conclusion can be expressed in the “uncertainty under uncertainty”. It is there, where their convergence challenge focus on.

In either event, known approaches to solve this issue can be reduced to intervals convergence. It is worth saying that about at the same time many researchers addressed these issues. Thus, Ben-Nairn and Prade (2008) made an efficient attempt to solve a task of trust towards a particular economic agent with the generalized interval, “which necessarily describes the past behaviour [alongside existing behaviour patterns in the present time - *scholium*]”. But at the same time, a difficulty arises, i.e. whether such an interval is sufficiently compressed. There is another solution to the issue in the paper by Levin (2004), where he argues his stand of over fuzzy sets. However, the convergence issue cannot be solved by a simple combination of intervals or separation from them such a component, which locates at their intersection point.

4.2 Tenversion Uncertainty

Based on the abovementioned, we can say of a substitutive role of the additional uncertainty with respect to the interval one. To ease the reading, to the author’s mind, its new type is appropriate to be referred to using a concept of the *tenversion uncertainty* (Note 5). It assumes describing how difficult it is to obtain the specific uncertainty among the interval range. At the same time, converging upper and lower bounds of the alternative probability reduces the *tenversion uncertainty*. As far as you get closer to the specific instance, it reduces to zero. For a clear description it would be appropriate to refer to the uncertainty function for the interval (see Figure 2 and 3).

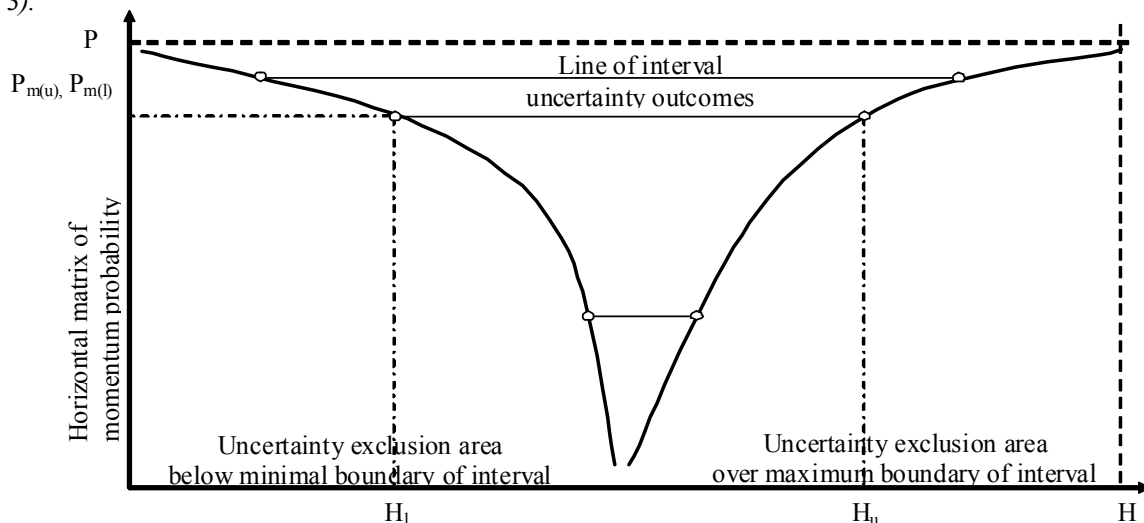


Figure 2. Function of the interval uncertainty for inhomogeneous alternatives (where H is entropy, P is possibility, Pm is momentum possibility, l-low, u-upper).

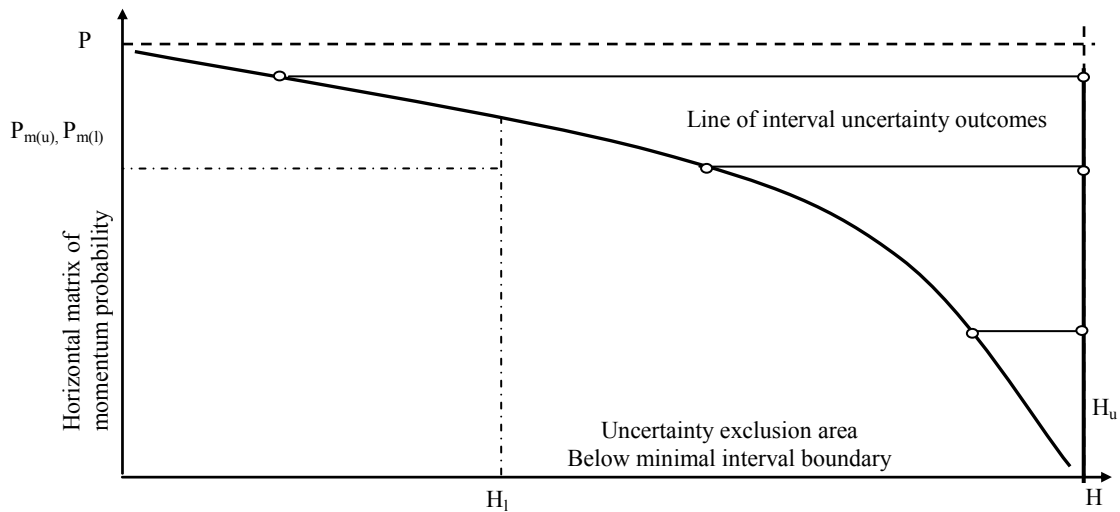


Figure 3. Function of interval uncertainty for homogeneous alternatives (where H is entropy, P is possibility, Pm is momentum possibility, l-low, u-upper).

Being in a conjunction with the interval, the tenversion uncertainty has its quite reasonable limit equal to the maximum of the instance uncertainty, described in the interval analysis, except that a structure of alternatives gets its identically similar probability disregarding not specified range restrictions. Scientific consistency for this assumption is based on the so-called extreme case when the frequency range of source alternatives causes the uncertainty contained in natural constraints.

There have been no similar solutions to the issue of evaluating the tenversion uncertainty in known academic papers. Therefore, the author proposes an approach to calculate it. This suggests that such a task is put for the first time. The calculation method involves a removal of a dummy component from its structure. Its role is to collect “exact probabilities”, i.e. those probabilities, which cannot be below the set case. All the hypotheses recently developed, must be similar at the same time. Hence, the probability remaining from the single instance (minus the “exact” one) is distributed between them. A calculation can be represented as follows:

$$H_t(I) = -\sum_{j=1}^n c_j \times \log_b c_j \tag{17}$$

$$c_j = \frac{a_2 - a_1}{n} = \frac{1 - \sum_{i=1}^n a_i}{n} \text{ where } \sum_{j=1}^n c_j < 1,$$

where $H_t(I)$ is the tenversion uncertainty and c_j is the equiprobable alternative under the tenversion uncertainty.

It may seem that a term when alternatives for the tenversion uncertainty do not constitute the single instance is erroneous. However, it seems be not the case. The dummy component complements the single instance, whereas the uncertainty depending on the lowest bound of the probability has been included in the interval uncertainty, even in a form of some wide range. As a result, there has remained a probability pool blurring clear boundaries of choice complexity and hence the uncertainty at time of such choice.

5. Conclusion

Summarizing the findings one can say for sure that the interval uncertainty is not only a characteristic form of the unknown, inexact and ambiguous in economic, but also in any other sphere of the social life. The environment is fuzzy and stochastic (probabilistic, - in research by Kim, Ovseyevich and Reshetnyak (1993), in a contrast to non-stochastic aspect - in papers by Elishako, Cai and Starnes (1994)), where solving certain administrative tasks has been reduced to optimization and anti-optimization techniques. It facilitates identifying aspects of the uncertainty previously unknown to scientists. The presented in the research new type of the uncertainty, called tenversion, in this context, is an illustrative example. The presented approach to logical evaluation of the interval uncertainty shows its balanced consistency in methods, their continuity in learning the

experience gained in uncertainty studies. Many fundamental conclusions such as on probability averaging, accentuate available essential contradictions between set and conventionally established stereotypes. Further research would help making the uncertainty theory smoother in recently found specifics and essential manifestations.

Acknowledgement

The Russian Scientific Foundation (RSF), project No. 14-18-00574 “The Information and Analytical System Anti-Crisis: Region Diagnosis, Threat Assessment and Scenario Forecasting to Preserve and Strengthen Economic Security and Raise Welfare in Russia” supported this research.

References

- Afanasyev, V. G. (1975). *Social Information and Social Management*. [Sotsial'naiia informatsiia i upravlenie obshchestvom]. Moscow, Russia: Politizdat.
- Ang, A. H. S., & Tang, W. H. (1975). *Probabilistic Concepts in Engineering Planning and Design*. New York: John Wiley.
- Avdiysky, V. I., & Kuzmin, E. A. (2013). Criterion conditions to manifest stability in economics: indicative parameter of uncertainty. [Kriterial'nye usloviia proiavlennii ustoiichivosti v ekonomike: indikativnyi parametr neopredelennosti]. *Issues of Risk Analysis*, 6(10), 10-22. Retrieved June 14, 2014, from <http://elibrary.ru/item.asp?id=21338344>
- Belyaev, L. S., & Makarov, A. A. (1974). *Uncertainty Factor in Making Optimal Decisions in Large Energy Systems*, Vol. 2. [Faktor neopredelennosti pri priniatii optimal'nykh reshenii v bol'shikh sistemakh energetiki]. Novosibirsk, Russia: Siberian Energy Institute.
- Ben-Nairn, J., & Prade, H. (2008). Evaluating Trustworthiness from Past Performances: Interval-Based Approaches. In S. Greco, & T. Lukasiewicz (Eds.), *Scalable Uncertainty Management: Lecture Notes in Computer Science* (Vol. 5291, pp. 33-46). Springer. http://dx.doi.org/10.1007/978-3-540-87993-0_5
- Borisov, A. N., Alekseev, A. V., & Merkureva, G. V., et al. (1989). *Processing Fuzzy Information in Decision Making Systems*. [Obrabotka nechetkoi informatsii v sistemakh priniatii reshenii]. Moscow, Russia: Radio and Communication.
- Chanas, S., & Kamburowski, J. (1981). The use of fuzzy variables in PERT. *Fuzzy Set Systems*, 5, 1-19. [http://dx.doi.org/10.1016/0165-0114\(81\)90029-4](http://dx.doi.org/10.1016/0165-0114(81)90029-4)
- Dubois, D., Prade, H., & Smets, P. (1996). Representing partial ignorance. *IEEE Transactions on Systems, Man and Cybernetics*, 26, 361-377. <http://dx.doi.org/10.1109/3468.487961>
- Elishako, I., Cai, G. Q., & Starnes, J. H., Jr, (1994). Non-linear buckling of a column with initial imperfection via stochastic and non-stochastic convex model. *International Journal of Non-Linear Mechanics*, 29(1), 71-82. [http://dx.doi.org/10.1016/0020-7462\(94\)90053-1](http://dx.doi.org/10.1016/0020-7462(94)90053-1)
- Elishakoff, I., & Ohsaki, M., (2010). *Optimization and Anti-optimization of Structures under Uncertainty*. London, United Kingdom: World Scientific.
- Fagin, R., & Halpern, J. Y. (1991). A New Approach to Updating Beliefs, Uncertainty. In Bonissone, et al. (Eds.), *Artificial Intelligence 6* (pp. 347-374). Elsevier Science Publishers.
- Ferson, S., & Ginzburg, L. (1996). Different methods are needed to propagate ignorance and variability. *Reliability Engineering and Systems Safety*, 54, 133-144. [http://dx.doi.org/10.1016/S0951-8320\(96\)00071-3](http://dx.doi.org/10.1016/S0951-8320(96)00071-3)
- Fortin, J., Zieliński, P., Dubois, D., & Fargier, H., (2010). Criticality analysis of activity networks under interval uncertainty. *Journal of Scheduling*, 6 (13), 609-627. <http://dx.doi.org/10.1007/s10951-010-0163-3>
- Guo, P., & Tanaka, H., (2007). *Interval regression analysis and its application*. Ipm30: Interval and Imprecise Data Analysis. Proceedings of the 56th Session, Bulletin of the International Statistical Institute, Lisboa, Portugal.
- Guo, P., & Tanaka, H., (2010). On Probabilities Interval. In Huynh, V. N., Nakamori, Y., Lawry, J., & Inuiguchi, M. (Eds.), *Integrated Uncertainty Management and Applications*. Springer. http://dx.doi.org/10.1007/978-3-642-11960-6_15
- Halpern, J., & Fagin, R. (1992). Two views of belief: Belief as generalized probability and belief as evidence. *Artificial Intelligence*, 54, 275-317. Retrieved June 14, 2014, from <http://researcher.watson.ibm.com/researcher/files/us-fagin/ai92.pdf>

- Hapke, M., Jaszkievicz, A., & Słowiński, R. (1994). Fuzzy project scheduling system for software development. *Fuzzy Sets and Systems*, 67, 101-107. [http://dx.doi.org/10.1016/0165-0114\(94\)90211-9](http://dx.doi.org/10.1016/0165-0114(94)90211-9)
- Kelley, J. E. (1961). Critical path planning and scheduling-mathematical basis. *Operations Research*, 9, 296-320. <http://dx.doi.org/10.1287/opre.9.3.296>
- Kim, Y. V., Ovseyevich, A. I., & Reshetnyak, Y. N. (1993). Comparison of stochastic and guaranteed approaches to the estimation of the state of dynamic systems. *Journal of Computer and System Sciences International*, 31(6), 56-64. Retrieved June 14, 2014, from <http://elibrary.ru/item.asp?id=12733744>
- Kreinovich, V., Wu, B., & Xiang, G. (2012). *Computing Statistics under Interval and Fuzzy Uncertainty: Applications to Computer Science and Engineering*. Berlin, Germany: Springer. <http://dx.doi.org/10.1007/978-3-642-24905-1>
- Kuzmin, E. A. (2012). Uncertainty and Certainty in Management of Organizational and Economic Systems. [Neopredelennost' i opredelennost' v upravlenii organizatsionno-ekonomicheskimi sistemami]. Ekaterinburg, Institute of Economics, Ural Branch of the RAS.
- Kyburg, H. E., Jr. (1989). Higher Order Probabilities. In Kanai, L. N., Levitt, T. S., & Lemmer, J. F. (Eds.), *Uncertainty in Artificial Intelligence 3* (pp. 15-22). Elsevier Science Publishers, Amsterdam.
- Leshkevich, T. (1994). *Uncertainty in World and World of Uncertainty: Philosophical Reflections on Order and Chaos*. [Neopredelennost' v mire i mir neopredelennosti: filosofskie razmyshleniia o poriadke i khaose]. Rostov-on-Don, Russia: Rostov University Press.
- Levin, V. I. (2011). Simulating optimization tasks under interval uncertainty conditions. [Modelirovanie zadach optimizatsii v usloviakh interval'noi neopredelennosti]. *News of the V.G. Belinsky Penza State Pedagogical University*, 26, 589-595. Retrieved June 14, 2014, from <http://cyberleninka.ru/article/n/modelirovanie-zadach-optimizatsii-v-usloviyah-intervalnoy-neopredelennosti>
- Levin, V. I. (2004). Interval logic and over fuzzy sets. Theory and application. [Interval'naia logika i sverkhnechetkie mnozhestva. Teoriia i primeneniie]. *Bulletin of the Tambov State Technical University*, 4-1(10), 924-929. Retrieved June 14, 2014, from http://www.vestnik.tstu.ru/rus/t_10/pdf/10_4_004.pdf
- Li, H., & Yue, D. (2010). Synchronization of Markovian jumping stochastic complex networks with distributed time delays and probabilistic interval discrete time-varying delays. *Journal of Physics A: Mathematical and Theoretical*, 43(10), 105101. <http://dx.doi.org/10.1088/1751-8113/43/10/105101>
- Li, H., Wong, W. K., & Tang, Y. (2012). Global Synchronization Stability for Stochastic Complex Dynamical Networks with Probabilistic Interval Time-Varying Delays. *Journal of Optimization Theory and Applications*, 2(152), 496-516. <http://dx.doi.org/10.1007/s10957-011-9917-0>
- Lindley, D. V. (1987). The probability approach to the treatment of uncertainty in artificial intelligence and expert systems. *Statistical Science*, 2, 17-24. <http://dx.doi.org/10.1214/ss/1177013427>
- Loostma, F. A. (1997). *Fuzzy logic for planning and decision-making*. Dordrecht: Kluwer Academic.
- Loui, R. (1986). Interval-based decisions for reasoning systems. In L. N. Kanal, & J. E. Lemmer (Eds.), *Uncertainty in Artificial Intelligence* (pp. 459-472). Elsevier Science Publishers, Amsterdam, Netherlands. <http://dx.doi.org/10.1016/B978-0-444-70058-2.50039-5>
- McCahon, C. S., & Lee, E. S. (1988). Project network analysis with fuzzy activity times. *Computers and Mathematics with Applications*, 15, 829-838. [http://dx.doi.org/10.1016/0898-1221\(88\)90120-4](http://dx.doi.org/10.1016/0898-1221(88)90120-4)
- Neapolitan, R. E. (1996). Is higher-order uncertainty needed? *IEEE Transactions on Systems, Man and Cybernetics*, 26, 294-302. <http://dx.doi.org/10.1109/3468.487955>
- Nguyen, H. T., Kosheleva O., Kreinovich, V., & Person, S. (2008). Trade-off between Sample Size and Accuracy: Case of Dynamic Measurements under interval uncertainty. In Huynh, V. N., Nakamori, Y., & Ono, H. et al. (Eds.), *Interval / Probabilistic Uncertainty and Non-classical Logics*. Springer. http://dx.doi.org/10.1007/978-3-540-77664-2_5
- Panyukov, A. V., & Latipova, A. T. (2008). Evaluation of Equilibrium in von Neumann's Model with Interval Uncertainty of Source Data. [Otsenka ravnovesiia v modeli Neimana pri interval'noi neopredelennosti iskhodnykh dannyykh]. *Equilibrium Models in Economics and Energetics. Proceedings of the All-Russian Conference and Section of Mathematical Economics at the 14th Baikal International School-Workshop*. Irkutsk, Russia: Energy Systems Institute of Siberian Branch of the Russian Academy of Sciences.

- Parsons, S., (2001). *Qualitative Methods for Reasoning under Uncertainty*. United States: MIT Press.
- Parygin, B. D. (1978). *Technological Revolution and Personality*. [Nauchno-tehnicheskaja revoliutsiia i lichnost']. Moscow, Russia: Politizdat.
- Prade, H. (1979). Using fuzzy sets theory in a scheduling problem: a case study. *Fuzzy Sets and Systems*, 2, 153-165. [http://dx.doi.org/10.1016/0165-0114\(79\)90022-8](http://dx.doi.org/10.1016/0165-0114(79)90022-8)
- Rommelfanger, H. (1994). Network analysis and information flow in fuzzy environment. *Fuzzy Sets and Systems*, 67, 119-128. [http://dx.doi.org/10.1016/0165-0114\(94\)90212-7](http://dx.doi.org/10.1016/0165-0114(94)90212-7)
- Shang, Y. (2011). Robustness of scale-free networks under attack with tunable grey information. *EPL (Europhysics Letters)*, 95(2), 28005. <http://dx.doi.org/10.1209/0295-5075/95/28005>
- Shary, S. P. (2007). Interval analysis or Monte Carlo methods? [Interval'nyi analiz ili metody Monte-Karlo?]. *Computational Technologies*, 1(12), 103-115.
- Shary, S. P. (2012). *Finite Measured Interval Analysis*. [Konechnomernyi interval'nyi analiz]. Novosibirsk, Russia: Institute of Computational technologies of SB RAS; XYZ Press.
- Shen, B., Wang, Z., Ding, D., & Shu, H., (2013). H_∞ state estimation for complex networks with uncertain inner coupling and incomplete measurements. *IEEE Trans Neural Netw Learn Syst.* 24(12), 2027-2037. <http://dx.doi.org/10.1109/TNNLS.2013.2271357>
- Spall, J. C. (2002). Uncertainty Bounds in Parameter Estimation with Limited Data. In Dror, M., L'Ecuyer, P., & Szidarovszky, F. (Eds.), *Modeling Uncertainty: an Examination of Stochastic Theory, Methods, and Applications* (pp. 685-710). Springer. http://dx.doi.org/10.1007/0-306-48102-2_27
- Sternin, M., & Shepelyov, G., (2012). Evaluating interval alternatives: uncertainties and preferences. [Otsenka interval'nykh al'ternativ: neopredelennosti i predpochtenii]. *Information Models and Analyses*, 1, 357-369. Retrieved June 14, 2014, from <http://www.foibg.com/ijima/vol01/ijima01-4.pdf>
- Xiang, G. (2007). *Fast Algorithms for Computing Statistics under Interval Uncertainty, with Applications to Computer Science and to Electrical and Computer Engineering*. United States: The University of Texas at El Paso. Computer Science; ProQuest.
- Yager, R. R., & Kreinovich, V. (1999). Decision making under interval probabilities. *International Journal of Approximate Reasoning Qualitative*, 22, 195-215. [http://dx.doi.org/10.1016/S0888-613X\(99\) 00028-6](http://dx.doi.org/10.1016/S0888-613X(99) 00028-6)

Notes

Note 1. It is quite reasonable that a view of averaging incorrectness assumes an opposition between the uncertainty nature and framing of that alternative among many similar, which would make a picture of its dominating over others, as it is such alternative, which comes out of a total ensemble of hypotheses. Averaging only applies to “free” alternatives; therefore, remaining may differ from the others. Then the uncertainty will be relatively less, as a case of choice is clearer. However, it is difficult to say definitely that the averaging method cannot be used. On the contrary, an example of the equiprobable outcome shows that the method is reasonable to maximize the uncertainty. A mixed option implies separating alternatives, their differentiation into fixed and mobile (free) hypothesis in a probabilistic manifestation. Hence, a significant assumption occurs, in which a set of alternatives is only partially homogeneous; therefore, fixed alternatives appear as dominating. Similar ideas validating the averaging one can trace in the paper by Levin (2011), who says of existing optimization approach based on average values of the system parameters, enclosed within the interval value.

Note 2. Saying the “first match”, the author assumes identifying such an alternative, for which for the first time there will be met additional requirements regardless whether there are similar or exactly one-to-one hypothesis of analogous dimensions in intervals of probabilities. The first match is intended to reveal the alternative that should degrade an overall layout of the uncertainty in favour of probability reducing by, on the one hand, comparing uncertainties of the lowest and upper thresholds, and, on the other hand, by calculating a difference between them, that will be a criterion of such a hypothesis found. We should clarify that the “first match” has been accepted as one of possible methods. It is not excluded and it is even allowed to choose any other serial match of a minimum difference of boundary uncertainties between alternatives, if any. It is important for such a match of two or more to be exclusive.

Note 3. Saying the method of “the last match” the author refers to logical and arithmetic operations with those alternatives that: firstly, are identically one-to-one and similar to each other; secondly, they cover all the

matching alternatives except the first one. It is worth saying that with synectic similar of alternatives, a choice of the first one is based on identification simplicity, while it is not excluded and it is allowed to choose another serial alternative from a set of matching subject to logical conditions for exception and other related operations to harmonize probabilities and make the single instance.

Note 4. For the third option, a calculation of the q^{min} indicator according to the equation (12) has retained its original form, with an exception that values of the applied probability are changed. Herewith a decrease (probability change) each time occurs by the same value. Logical values of the error according to (12) assume identifying those alternatives, a change to the probability of which is impossible.

Note 5. An origin for a concept of the “tenversion uncertainty” goes back to Latin terms “tension” (tension), “tenuis” (up to) and “vaste” (broad, wide). In the author’s opinion, the tenversion uncertainty complements the interval evaluation with the new informative content. It seems that the probability range causes a case when the interval uncertainty loses complexity of choice identification across the uncertainty’s full length. It seems that the tenversion uncertainty solves this issue, essentially reflecting the interval filling.

It is worth paying attention that the tenversion uncertainty is not included in a cyclic sequence of a change to types of the environmental uncertainty, decision-making, consequences of these decisions and the variational uncertainty. It is important to understand that the tenversion uncertainty accompanies the interval analysis, confirming or non-confirming its acceptability for each specific case.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

Economic Growth and Internet Usage Impact on Publication Productivity among ASEAN's and World's Best Universities

Hossein Gholizadeh¹, Hadi Salehi², Mohamed Amin Embi³, Mahmoud Danaee⁴, Ali Ordi⁵, Farid Habibi Tanha⁶,
Nader Ale Ebrahim⁷ & Noor Azuan Abu Osman¹

¹ Department of Biomedical Engineering, Faculty of Engineering, University of Malaya, 50603 Kuala Lumpur, Malaysia

² Faculty of Literature and Humanities, Najafabad Branch, Islamic Azad University, Najafabad, Isfahan, Iran

³ Faculty of Education, Universiti Kebangsaan Malaysia, Bangi, 43600, Malaysia

⁴ Faculty of Agriculture, Roudehen Branch, Islamic Azad University, Roudehen, Iran

⁵ Universiti Teknologi Malaysia, Advance Informatics School (AIS), Kuala Lumpur, Malaysia

⁶ Department of Financial Sciences, University of Economic Sciences, Tehran, 1593656311, Iran

⁷ Research Support Unit, Centre of Research Services, Institute of Research Management and Monitoring (IPPP), University of Malaya, Malaysia

Correspondence: Hadi Salehi, Faculty of Literature and Humanities, Najafabad Branch, Islamic Azad University, Najafabad, Isfahan, Iran. E-mail: hadisalehi1358@yahoo.com

Received: July 3, 2014 Accepted: July 21, 2014 Online Published: September 1, 2014

doi:10.5539/mas.v8n5p169

URL: <http://dx.doi.org/10.5539/mas.v8n5p169>

Abstract

Measuring the number of papers which are published each year, publication productivity is the factor which shows the reputation of universities and countries. However, the effect of growing economy and using internet on the publication productivity in Asian countries has not been discovered yet. The present research is going to figure out the publication productivity among the elite universities in Asian countries and also ten top universities around the world in the last twenty years (from 1993 to 2012). Furthermore, the current research is aimed to study the relationship among publication, gross domestic product (GDP) and internet usage. It is worth to mention that the publication of the top Ten Malaysian Universities was regarded for the similar period of time. To get the exact numbers of documents like papers, conference articles, review papers and letters which are published by the universities in the last twenty years, the writer of the same paper used the Science Direct database. Moreover, the data for GDP and the number of internet usage was collected through the World Bank database (World Data Bank). To compare all kinds of publications, one-way ANOVA was used and to investigate the impact of economic growth and internet usage on publication productivity, multiple regression analysis was applied. The results showed that the rate of publication growth was 1.9, 20.9, and 65.5 % in top universities in the world, ASEAN countries and Malaysia, respectively. The results also showed that there was a positive and significant correlation between GDP and the number of internet users with the number of publications in ASEAN and Malaysian universities. Internet usage had much more influence in comparison with the GDP in predicting the number of publications among these groups except for top ten Malaysian universities from 2003 to 2012. In summary, publication trends in top ten Malaysian and ASEAN universities are promising. However, policy makers and science managers should spend much more percentage of their GDP on Internet facilities and research studies that their outputs lead to more rapid economic growth and internet usage.

Keywords: ASEAN, publication productivity, documents, internet usage, GDP, Malaysian Universities, publication trend

1. Introduction

Ten countries in Southeast Asia formed the geo-political and economic Association of Southeast Asian Nations. It was first formed on 8 August 1967 by Indonesia, Malaysia, the Philippines, Singapore and Thailand. Some other countries like Brunei, Burma (Myanmar), Cambodia, Laos, and Vietnam became the members of this group and expanded it. The chief purpose of organizing this group is increasing economic growth (Sarel, 1997).

Gross domestic product (GDP) which refers to the market price of all officially realized goods and services which are produced in a country in a specific period of time is the chief tool to measure the economy of a country. Reviewing the previously published papers, the writer found out that there is a relationship between economic growth with education and paper publication. The above mentioned result is not fully support by realistic facts (Jin, 2013; Nelson, 1966; Lucas Jr, 1988; Becker, 2009; Romer, 1990). Positive and significant impact of education on economic growth has been found by Mankiw et al. (1992) and Barro (1991) (Mankiw et al., 1992 & Barro, 1991); however, a possibility of reverse relationship between economic growth and education was shown by Bils and Klenow (2000). Besides, Jin and Jin newly indicated that the effect of publication productivity on economic growth is not the same in different fields. For example, there is a positive relationship between engineering and science with economic growth, while the social sciences do not have the same effect on economic growth (Jin, 2013).

Nowadays, compared with the past twenty years, studying the publication productivity is a main topic for the researchers and students because the findings of the researches can positively affect the whole community (Zain et al., 2009). According to the recent development rules, the numbers of educated employees have been enhanced. This matter helps the economy of the countries to grow very fast (Jin, 2013). It has been found out that those countries which are highly developed like the United States and England are among the world's top productive research universities. The number of publication shows the research productivity and is employed to grade the countries and universities (Yazit and Zainab, 2007, Narin and Hamilton, 1996, Toutkoushian et al., 2003, Liu and Cheng, 2005, Meho and Spurgin). It can also be used to determine author's productivity or the publication productivity of research groups (Liu, 2005; Hart, 2000; Uzun, 2002; Gu, 2001; Fox, 1983). Numerous referring to the previously published papers by the new papers a lot, shows the following identification and also the effect in the field of study. Those review articles which refer to other articles a lot, can give us some facts about the major areas of discipline besides, they can emphasize the increase of specific fields. Moreover, more frequently cited papers are mostly written by famous researchers who can impress future directions of the field by their ideas (Lefavre and O'Brien, 2011, Kelly et al., 2010, Ponce and Lozano, 2010, Joynt and Leonard, 1980).

Several indicators of academic or research performance are used to rank educational institutes and universities. They include alumni and staff winning Nobel Prizes and Fields Medals, highly cited researchers, papers published in Nature and Science, papers indexed in major citation indices, and the per capita academic performance of an institution. The Academic Ranking of World Universities (ARWU) is the first global ranking of universities to be published. Today, ARWU is regarded to be one of the three most influential and widely observed international university rankings, along with the QS World University Rankings and the Times Higher Education World University Rankings. The Academic Ranking of World Universities (ARWU), commonly known as the Shanghai Ranking, is a publication that was founded and compiled by the Shanghai Jiao Tong University to rank universities globally. The rankings have been conducted since 2003 and updated annually.

The current study is mainly going to investigate the amount of publication productivity among the best universities in ASEAN countries and world's top ten universities from 1993 to 2002 and 2003 to 2012. The study also aimed to achieve the following objectives:

- Studying the relationship among publication productivity, gross domestic product (current US \$) and internet users
- Examining the publication direction of ten elite Malaysian Universities in a specific time periods

Since the Science Direct offers about 20% more inclusion than Web of Science, it has been used as the first full-text theoretical database in this research. The researchers think that there is a positive relationship among the economic growth, the numbers of people who can use the internet and also the number of publication of papers in elite Asian universities and also the ten best universities in the whole world.

2. Methodology

ScienceDirect database was used to collect the number of documents including articles, conference papers, review papers, letters and books published in the last two decades from 1993 to 2002 and 2003 to 2012. These data were collected to make a comparison among the top university in each ASEAN country, top ten universities in Malaysia and top ten universities in the world. To find the first university in each ASEAN country and top ten Malaysian universities, we used the number of publications in ScienceDirect database. Moreover, to determine the top ten universities in the world, the Academic Ranking of World Universities (ARWU) was used. Furthermore, in each university, the main subject area (overall), number of papers published in Nature and Science journals, and the most cited papers were identified. The numbers of citations that each paper could receive

during one week were identified as well.

To compare all kinds of publications among top ten universities in the world, ASEAN and Malaysia), one-way ANOVA was applied. As the homogeneity test was not met, the Welch statistic was used to test the equality of means. Moreover, to evaluate the relationship between publications (articles, conference papers, review papers, letters) with GDP and Internet usage, Spearman correlation coefficient test was applied. To investigate economic growth and internet use impact on publication productivity, multiple regression was applied to examine what extent the proposed multiple linear regression model is supported by the research data. The regression examined how well the number of publications could be predicted from GDP and internet usage. In the multiple regression model, GDP and Internet usage were set as the independent variables and the number of publications was considered as the dependent variable.

3. Results

According to the Academic Ranking of World Universities (ARWU), the top ten universities in the world are mostly located in the United States (8 universities) or United Kingdom (2 universities) (see Table 1). Moreover, the first university in each ASEAN country and top ten universities in Malaysia based on the number of publications in ScienceDirect database were listed in Table 1. The main research areas in world's best universities were physics and astronomy (7 universities), medicine (2 universities) and engineering (1 university). In these universities, the average number of papers published in Nature and Science were 1586 and 1419, respectively (see Table 1).

Table 1. The average number of published papers

Institution (country)	established (year)	Overall publication*	Main Subject area	Main Subject area (%) publication)	papers (Nature)	papers (Science)	most cited paper (citation) 22/7/2013	most cited paper (citation) 29/7/2013	most cited paper (publication date)	
Top world's universities	Harvard University (US)	1636	74433	Physics and Astronomy	14.8	1430	2294	10224	10255	1990
	Stanford University (US)	1891	110914	Engineering	13.6	861	1593	6249	6266	2001
	Massachusetts Institute of Technology (MIT) (US)	1861	134794	Physics and Astronomy	20.1	1563	1860	11678	11732	2000
	University of California, Berkeley (US)	1868	158231	Physics and Astronomy	15.3	1864	2233	18659	18757	1965
	University of Cambridge (UK)	1209	135913	Physics and Astronomy	14.9	4099	644	7966	7977	1990
	California Institute of Technology (US)	1891	62675	Physics and Astronomy	26.3	974	1134	8657	8705	1995
	Princeton University (US)	1764	62273	Physics and Astronomy	20.1	754	945	6123	6136	1998
	Columbia University (US)	1754	112569	Medicine	17.9	676	1403	10425	10484	1998
	University of Chicago (US)	1890	90126	Medicine	21.7	980	1560	11741	11777	1953
	University of Oxford (UK)	1096	122553	Physics and Astronomy	15.2	2658	526	10198	10216	2001
Top ASEAN Universities	National University of Singapore (Singapore)	1905	74484	Engineering	17.9	71	0	2171	2180	2003
	University of Malaya (Malaysia)	1949	21563	Medicine	14.9	24	0	445	449	2000
	Mahidol University	1943	20291	Medicine	41.9	0	0	1494	1503	2005

(Thailand)										
Top Malaysian Universities	Institut Teknologi Bandung (Indonesia)	1959	2979	Engineering	23.6	0	5	330	330	2001
	International Rice Research Institute (Philippines)	1960	2955	Agricultural and Biological	54	15	11	857	861	1972
	Vietnam National University (Vietnam)	1945	1230	Computer Science	15.2	0	0	153	153	2000
	UNIVERSITI BRUNEI DARUSSALAM (Brunei)	1985	864	Agricultural and Biological	10.4	0	0	222	224	2003
	Institut Pasteur du Cambodge (Cambodia)	1953	251	Medicine	44.9	0	0	680	687	2009
	National University of Laos (Laos)	1996	178	Agricultural and Biological	26.3	0	0	171	171	2003
	University of Yangon (Myanmar)	1878	109	Chemistry	13.6	0	1	77	77	1999
	University of Malaya	1949	21572	Medicine	14.9	24	0	449	452	2004
	Universiti Sains Malaysia	1962	17054	Material Science	13.3	0	0	808	811	1996
	Universiti Putra Malaysia	1969	16322	Agricultural and Biological	15.7	0	0	453	455	2008
	Universiti Kebangsaan Malaysia	1970	15010	Engineering	15.5	0	0	449	452	2004
	Universiti Teknologi Malaysia	1975	10134	Engineering	26.1	0	0	260	261	2004
	Universiti Teknologi MARA	1956-1965	6784	Engineering	22.5	0	0	305	307	2010
	International Islamic University Malaysia	1983	3995	Engineering	19.8	0	0	100	101	2007
	Multimedia University	1996	3872	Engineering	27	0	0	275	276	2001
	Universiti Teknologi Petronas	1997	3343	Computer science	23.5	0	0	77	77	2008
	Universiti Malaysia Perlis	2001	2321	Engineering	32.6	0	0	137	137	2007

However, top universities in ASEAN countries could averagely publish 11 papers in Nature Journal and 2 papers in Science journal. The average numbers of citations for the most cited papers in each university in these three groups (world, ASEAN and Malaysia) were 10192, 660 and 331, respectively. Furthermore, the resultsshowed 39 citations per week for most cited papers in world's top universities while it was 4 citations per week in ASEAN universities (see Table 1, Figure 1).

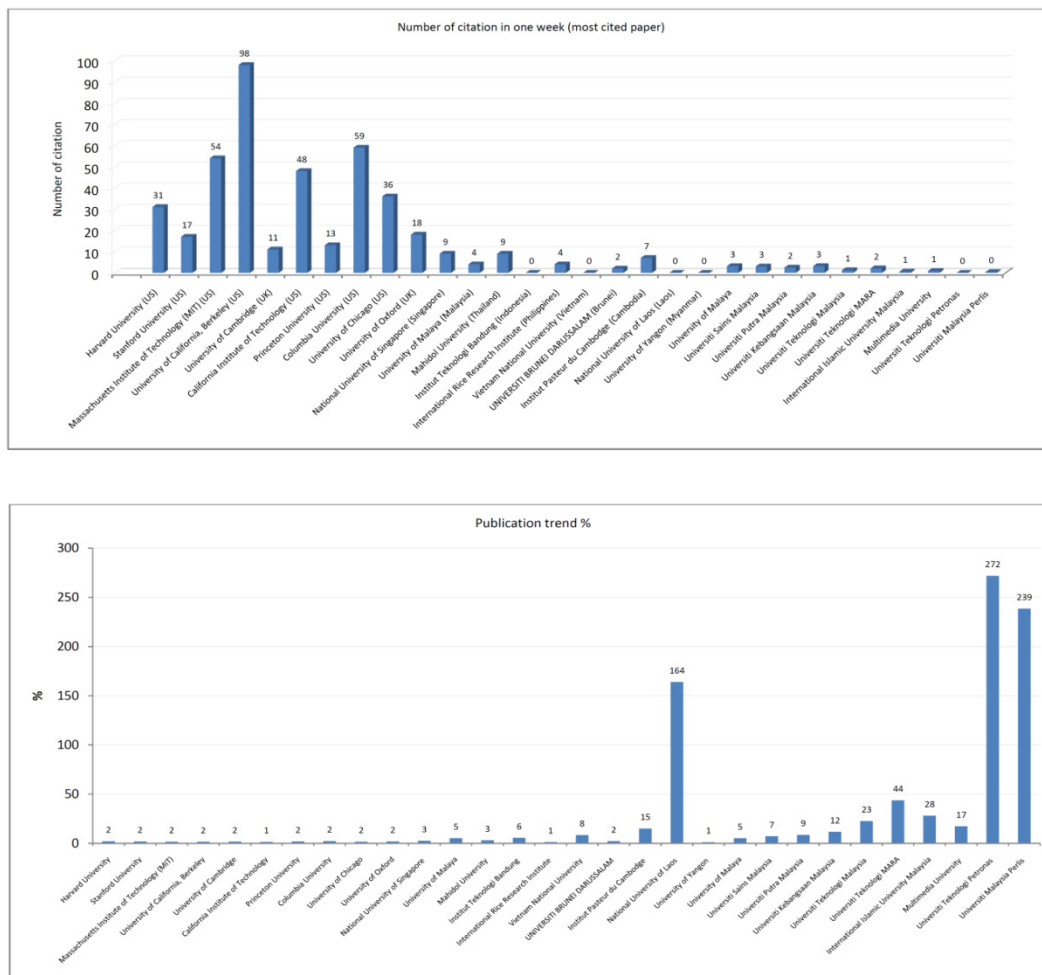


Figure 1. Number of citations in one week for the most cited papers in each university (Top); Publication trend (Down)

University of Singapore with 74484 papers was the first in ASEAN countries and was followed by University of Malaya with 21563 publications among ASEAN countries. Moreover, the University of Yangon in Myanmar had the least publications (109) among ASEAN countries. Interestingly, publication trends in ASEAN countries are promising compared to the top ten universities in the world (see Appendixes 1 and 2). The rate of publication growth between these two decades was 1.9, 20.9 and 65.5 percentage in top universities in the world, ASEAN and Malaysia, respectively (see Figure 1).

To compare all kinds of publications, the results of ANOVA showed that there were significant differences among these three groups. Duncan's multiple range test showed that there was no significant difference between ASEAN and Malaysia universities, while both of them were significantly lower than top ten world universities (see Table 2).

Table 3 shows the correlation between indices (articles, conference papers, review papers, and letters), GDP and internet users for these three groups. The results showed that there was a positive and significant correlation between GDP and the number of internet users with the number of publications in ASEAN and top ten Malaysian universities. However, there was a negative correlation between GDP and internet users with the number of letters published from 1993 to 2002 in Malaysia. Moreover, there was a negative and significant correlation between GDP and the number of articles published in world's top universities (see Table 3).

The R-squared (R²) value presented in Table 4 showed some information about the goodness of fit of a model. In regression, the R² coefficient of determination is a statistical measure of how well the regression line approximates the real data points. The R² of 1 indicates that the regression line perfectly fits the observed data.

The R^2 value of 0.584 for ASEAN universities from 1993 to 2002 implies that the two-predictor model explained about 58.4% of the variance in publications. Table 5 revealed that based on the reported value of the F-statistic, the model fits the data. This means that the slope of the estimated linear regression model line was not equal to zero; thus, it confirmed that there was a linear relationship between publication and the two-predictor variables.

Table 2. A (Welch robust Tests of Equality of Means); B (Pairwise comparison using Tamhane test)

(A)

	1993-2002			2003-2012		
	Welch statistic	df	Sig.	Welch Statistic	df	Sig.
Article	306.521	137.71	P<0.05	222.531	182.888	P<0.05
Conference	119.638	135.955	P<0.05	77.143	188.254	P<0.05
Review	159.91	142.948	P<0.05	221.178	160.595	P<0.05
Letter	143.944	168.747	P<0.05	101.268	156.708	P<0.05
Total publication	319.809	137.21	P<0.05	230.36	180.99	P<0.05

(B)*

Decade		Article	Conference	Review	Letter	Total publication
1993-2002	Top world	2030.73a	264.1a	124.6a	23.93a	2443.36a
	ASEAN	222.98b	28.76b	6.55b	1.65b	259.94b
	Malaysia	66.58c	8.87b	2.13b	0.78b	78.36c
2003-2012	Top world	3304.72a	776.77a	374.79a	58.3a	4514.58a
	ASEAN	574.85b	166.54b	35b	7.04b	783.43b
	Malaysia	509.03b	235.81b	22.39b	3.31b	770.54b

Means with same letter are not significantly different in each column.

*data are log transformed.

Table 3. Correlation coefficient between the number of articles, conference papers, review papers and letters with GDP and internet usage for Top ten universities of the world, ASEAN and Malaysia

		1993-2002		2003-2012	
		GDP	internet	GDP	internet
WOELD	Article	-0.042	.229*	-0.004	.531**
	Conference	0.178	.232*	0.012	0.158
	Review	0.185	.497**	-.443**	-0.062
	Letter	-0.128	0.137	-.256*	.324**
ASEAN	Article	.723**	.611**	.779**	.800**
	Conference	.775**	.599**	.737**	.739**
	Review	.565**	.574**	.518**	.705**
	Letter	.416**	.416**	.415**	.567**
MALAYSIA	Article	0.133	.280**	.624**	.595**
	Conference	.249*	.400**	.879**	.876**
	Review	0.09	0.171	.530**	.442**
	Letter	-0.03	-0.008	.338**	.258*

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

Table 4. Results of multiple linear regression between GDP and internet usage with total number of publications for Top ten universities of the world, ASEAN and Malaysia, respectively

1993-2002

group		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	R2
		B	Std. Error	Beta			
Top world							
	GDP*	-1149.03	279.863	-0.408	-4.106	P<0.05	0.187
	internet	21.03	4.943	0.423	4.254	P<0.05	
ASEAN							
	GDP*	197.843	70.785	0.214	2.795	P<0.05	0.584
	internet	43.194	4.661	0.708	9.267	P<0.05	
Malaysia							
	GDP*	100.412	164.662	0.066	0.61	0.543	0.063
	internet	1.771	0.888	0.216	1.994	P<0.05	

* log transformed.

2003-2012

group		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	R2
		B	Std. Error	Beta			
Top world							
	GDP*	-1461.77	446.388	-0.297	-3.275	P<0.05	0.276
	internet	121.544	25.935	0.425	4.687	P<0.05	
ASEAN							
	GDP*	363.724	161.614	0.177	2.251	P<0.05	0.551
	internet	42.706	4.983	0.676	8.57	P<0.05	
Malaysia							
	GDP*	6219.755	1339.622	1.141	4.643	P<0.05	0.421
	internet	-50.131	23.01	-0.535	-2.179	0.102	

* log transformed.

Standardized regression coefficients are presented in Table 4 to explain the importance of two predictors in predicting the number of publications. Independent variable with a high beta coefficient is highly important in contributing to the prediction of the number of total publications. Based on the beta values obtained, the beta coefficient in the world's top university was -0.408 and -0.297 for GDP and 0.423 and 0.425 for Internet usage between 1993-2002 and 2003-2012, respectively. This means that Internet usage had a much more significant influence than the GDP in predicting the number of publications. The results showed the same issue for ASEAN universities meaning that in these two decades, Internet usage had much more significant effects than the GDP in predicting the number of publications. Interestingly, GDP had more power (1.141) than the number of internet users (-0.535) from 2003 to 2012 in predicting the number of publications in top ten Malaysian universities.

4. Discussion and Conclusion

In this study, a comparison was made among the top university in each ASEAN country, top ten Malaysian universities and the world's top universities regarding the relationship between economic growth and internet usage with publication productivity from 1993 to 2002 and 2003 to 2012. Shanghai Ranking (ARWU) was used to determine the top ten universities in the world as it is the first global ranking of universities to be published and one of the three most influential and widely observed international university rankings. Moreover, the numbers of publications were used to find the first university in each ASEAN country and top ten Malaysian universities because most of these universities were not listed in the ARWU, QS, or the Times Higher Education World

University Rankings.

Publication productivity is an indicator of research output and could be used to rank countries and universities (Yazit and Zainab, 2007, Narin and Hamilton, 1996, Toutkoushian et al., 2003, Liu and Cheng, 2005, Meho and Spurgin). It can also be used to determine author's productivity or the publication productivity of research groups and to assess the productivity of persons in a particular discipline (Liu and Cheng, 2005, Hart, 2000, Uzun, 2002, Gu and Zainab, 2001, Fox, 1983, Yi et al., 2008). World's best universities are mostly working on physics and astronomy. Institutes which are working in these fields could publish more papers and get more citations. They could publish 144.2% and 709.5% more papers compared to ASEAN universities in Nature and Science journals, respectively (see Table 1). They could also receive 9.8 times more citations per week for their most cited papers (see Table 1).

Table 5. ANOVA table

1993

ANOVA ^b							
group	Model		Sum of Squares	df	Mean Square	F	Sig.
world	1	Regression	1.756E7	2	8781193.216	12.390	.000 ^a
		Residual	6.875E7	97	708732.439		
		Total	8.631E7	99			
asean	1	Regression	1.656E7	2	8279590.609	52.212	.000 ^a
		Residual	1.126E7	71	158575.023		
		Total	2.782E7	73			
my	1	Regression	57700.227	2	28850.113	3.281	.042 ^a
		Residual	853014.813	97	8793.967		
		Total	910715.040	99			

a. Predictors: (Constant), internet, log GDP.

b. Dependent Variable: total publication.

2003

ANOVA ^b							
group	Model		Sum of Squares	df	Mean Square	F	Sig.
world	1	Regression	5.958E7	2	2.979E7	17.996	.000 ^a
		Residual	1.440E8	87	1655372.703		
		Total	2.036E8	89			
asean	1	Regression	9.489E7	2	4.745E7	50.153	.000 ^a
		Residual	7.379E7	78	946035.347		
		Total	1.687E8	80			
my	1	Regression	2.040E7	2	1.020E7	33.387	.000 ^a
		Residual	2.658E7	87	305522.329		
		Total	4.698E7	89			

a. Predictors: (Constant), internet, log GDP.

b. Dependent Variable: total publication.

Publication trend from 1993 to 2012 in ASEAN and top ten Malaysian universities with 20.9 and 65.5% growth is promising compared to the world's top ten universities (1.9%). ASEAN and Malaysian universities have been averagely established over the past 66 and 37 years ago, respectively; while the top universities in the world have been averagely stabilised 327 years ago. This can be one of the reasons of high trend in publication productivity in ASEAN and Malaysian universities. In addition, the growth of publications might be overestimated due to the omitted variables, and a reverse causality from GDP to publications will be another possibility to occur. The number of publications in Malaysia increased dramatically after 2007 to 2012 (almost 5 times) (see Appendixes 1 and 2). One of the main reasons for increasing the number of paper publications in Malaysia could be greatly concentrating on enhancing the quality of research in the universities which are specific for researching like University of Malaya. Specified in the 10th Malaysia Plan, 1% of Malaysia GDP will be spent on the development and research projects. Furthermore, the other important reason could be the change of dissertations direction from conventional into research-based. PhD students have enhanced 10 times in the past years (there were about 4000 students in 2002, while they have increased to 40,000 in 2012). The top ten universities in Malaysia are shown in Table 1. It should be mentioned that the first five universities shown are ranked as the research universities and get more government funding.

The findings of this study showed a positive and significant correlation between GDP and the number of publications in ASEAN and Malaysian universities. This finding was in line with the findings of previous research studies that showed positive and significant correlation between education and economic growth (Mankiw et al., 1992; Barro, 1991; Bills and Klenow, 2000). Interestingly, there is a negative and significant correlation between the number of publications (articles and letters) and economic growth in top ten Malaysian universities from 1993 to 2012. Besides, Jin and Jin showed in their recent study that the publication productivity affect the economic growth differently in various fields. In top ten universities located in the United States and United Kingdom, GDP could not have positive effects to increase the number of publications. While the significant effects of GDP on the number of publications were seen in Malaysian universities especially from 2003 to 2012.

However, it should be noted that only the effects of GDP and internet usage on publications of recent years were evaluated in this study. Since the educational effects are accomplished over longer horizons, further investigation of the educational effects using the data that correspond to the stock of publications in earlier years would be important.

Science managers and policy makers in ASEAN countries should spend much more percentage of their GDP on development and research projects as Malaysia does. Moreover, increasing the number of PhD students and changing the university programs to paper-based dissertations could be another solution to increase the number of publications. Furthermore, it is necessary to find out different ways to improve the quality and visibility of the research studies and invest more on research studies that their outputs lead to more rapid economic growth.

References

- Barro, R. J. (1991). Economic growth in a cross section of countries. *The Quarterly Journal of Economics*, 106, 407-443. <http://dx.doi.org/10.2307/2937943>
- Becker, G. S. (2009). *Human capital: A theoretical and empirical analysis, with special reference to education*. University of Chicago Press.
- Bills, M., & Klenow, P. J. (2000). Does schooling cause growth? *American Economic Review*, 1160-1183. <http://dx.doi.org/10.1257/aer.90.5.1160>
- Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, web of science, and Google scholar: Strengths and weaknesses. *The FASEB Journal*, 22, 338-342. <http://dx.doi.org/10.1096/fj.07-9492LSF>
- Fox, M. F. (1983). Publication productivity among scientists: A critical review. *Social Studies of Science*, 13, 285-305. <http://dx.doi.org/10.1177/030631283013002005>
- Gu, Y., & Zainab, A. (2001). Publication productivity of Malaysian researchers in the field of Computer Science and Information Technology. *Malaysian Journal of Library & Information Science*, 6, 1-23.
- Hart, R. L. (2000). Co-authorship in the academic library literature: A survey of attitudes and behaviors. *The Journal of Academic Librarianship*, 26, 339-345. [http://dx.doi.org/10.1016/S0099-1333\(00\)00140-3](http://dx.doi.org/10.1016/S0099-1333(00)00140-3)
- Jin, J. C., & Jin, L. (2013). Research publications and economic growth: Evidence from cross-country regressions. *Applied Economics*, 45, 983-990. <http://dx.doi.org/10.1080/00036846.2011.613785>

- Joynt, R. L., & Leonard, J. A. (1980). Dantrolene sodium suspension in treatment of spastic cerebral palsy. *Developmental Medicine & Child Neurology*, 22, 755-767. <http://dx.doi.org/10.1111/j.1469-8749.1980.tb03742.x>
- Kelly, J., Glynn, R., O'briain, D., Felle, P., & McCabe, J. (2010). The 100 classic papers of orthopaedic surgery a bibliometric analysis. *Journal of Bone & Joint Surgery, British Volume*, 92, 1338-1343. <http://dx.doi.org/10.1302/0301-620X.92B10.24867>
- Lefavre, K. A., & O'brien, P. J. (2011). 100 most cited articles in orthopaedic surgery. *Clinical Orthopaedics and Related Research*, 469, 1487-1497. <http://dx.doi.org/10.1007/s11999-010-1604-1>
- Liu, N. C., & Cheng, Y. (2005). The academic ranking of world universities. *Higher education in Europe*, 30, 127-136. <http://dx.doi.org/10.1080/03797720500260116>
- Lucas, Jr, R. E. (1988). On the mechanics of economic development. *Journal of monetary economics*, 22, 3-42. [http://dx.doi.org/10.1016/0304-3932\(88\)90168-7](http://dx.doi.org/10.1016/0304-3932(88)90168-7)
- Mankiw, N. G., Romer, D., & Weil, D. N. (1992). A contribution to the empirics of economic growth. *The Quarterly Journal of Economics*, 107, 407-437. <http://dx.doi.org/10.2307/2118477>
- Meho, L. I., & Spurgin, K. M. (2005). Ranking the research productivity of LIS faculty and schools: An evaluation of data sources and research methods. *Journal of the American Society for Information Science and Technology*, 56(12), 1314-1331.
- Narin, F., & Hamilton, K. S. (1996). Bibliometric performance measures. *Scientometrics*, 36, 293-310. <http://dx.doi.org/10.1007/BF02129596>
- Nelson, R. R., & Phelps, E. S. (1966). Investment in humans, technological diffusion, and economic growth. *The American Economic Review*, 56, 69-75.
- Ponce, F. A., & Lozano, A. M. (2010). Highly cited works in neurosurgery. Part I: the 100 top-cited papers in neurosurgical journals: A review. *Journal of neurosurgery*, 112, 223-232. <http://dx.doi.org/10.3171/2009.12.JNS091599>
- Romer, P. M. (1990). Endogenous technological change. *Journal of political Economy*, S71-S102. <http://dx.doi.org/10.1086/261725>
- Sarel, M. (1997). Growth and productivity in ASEAN countries. International Monetary Fund.
- Toutkoushian, R. K., Porter, S. R., Danielson, C., & Hollis, P. R. (2003). Using publications counts to measure an institution's research productivity. *Research in Higher Education*, 44, 121-148. <http://dx.doi.org/10.1023/A:1022070227966>
- Uzun, A. (2002). Productivity ratings of institutions based on publication in Scientometrics, Informetrics, and Bibliometrics, 1981-2000. *Scientometrics*, 53, 297-307. <http://dx.doi.org/10.1023/A:1014864827441>
- Yazit, N., & Zainab, A. (2007). Publication productivity of Malaysian authors and institutions in LIS. *Malaysian Journal of Library and Information Science*, 12.
- Yi, H., Ao, X., & Ho, Y. S. (2008). Use of citation per publication as an indicator to evaluate pentachlorophenol research. *Scientometrics*, 75, 67-80. <http://dx.doi.org/10.1007/s11192-007-1849-y>
- Zain, Z. M., Ishak, R., & Ghani, E. K. (2009). The influence of corporate culture on organisational commitment: A study on a Malaysian listed company. *European Journal of Economics, Finance and Administrative Sciences*, 17, 16-26.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

Justification for Remote Control of Construction and Road-Making Machines

Nadegda Savelyevna Sevryugina¹, Eugene Alexandrovich Volkov¹ & Eugene Pavlovich Litovchenko¹

¹Belgorod State Shukhov Technological University, Belgorod, Russia

Correspondence: Nadegda Savelyevna Sevryugina, Belgorod State Shukhov Technological University, Kostytkova Street, 46, 308012, Belgorod, Russia. E-mail: ujinisthebigwolf@yandex.ru

Received: July 3, 2014 Accepted: July 16, 2014 Online Published: September 1, 2014

doi:10.5539/mas.v8n5p179

URL: <http://dx.doi.org/10.5539/mas.v8n5p179>

Abstract

It is evident from the experience of operating the construction machinery (excavators, bulldozers, loaders etc.) that quite often the machinery and its operators are working under severe conditions. Working under threat of rock fall, on unstable or contaminated grounds, or debris handling tends to increase the impact of adverse occupational environment on the operators' health. When the operators are at risk, it is advisable to deploy remote process control technologies on the work sites. This is what brings research and development of modern remote control systems to the top of the chart in order to improve the productivity of machinery, enhance the safety and quality of the jobs carried out. The use of remote control will exclude the adverse impact of aggressive environment during the process operations. The research allowed to establish that the design specifics of the construction and road-making machines enable them to go far beyond the standard process layout giving into a much wider range of application, up to aggressive environments. A crawler excavator model was devised, imitating the visibility range of the process areas for the manual and remote control systems. Implementation of remote control for construction and road-making machines may become a step towards a completely new level of interaction within the man-machine-environment system.

Keywords: aggressive environment, elimination, machine, safety, control system

1. Introduction

The designers developing modern construction and road-making machinery have to take into account a variety of functionality indices determined by psychophysical capabilities of operators, to ensure efficient and safe operation within the man-machine-environment system.

It is due to the fact that, as long the industry development intensifies and new machinery is conceived, the operators's activity is becoming increasingly complex and subject to stresses. The functional specifications of construction and road-making machinery do not at all times meet the requirements of the sites where it is operated. This results in deployment of substandard process layouts, which, in turn, leads to impaired efficiency and increased labor consumption (Khmara and Shatov 2012; Romanovich and Kharlamov, 2009). Such production sites feature specific operational conditions, whether natural and climatic or due to the highest level of man-made impact. Ensuring quality and safety of the operational tasks under such conditions becomes a real challenge, and the environment conditions set thresholds for the operator's work. On great many occasions, when the personnel's actions prove to be wrong, it is not due to poor skill level (though there are many problems on this side as well) but to the mismatch between the machinery's design features and human capabilities (Scopylatov & Yefremov, 2013). The physical environment of the production site has to correlate with the human performance features, and only then one can expect high productivity from him/her (Munipov & Zinchenko, 2001). Certain conditions demand from operators to use their psychophysical capabilities to the utmost extent, which, under adverse operational environment, may provoke erroneous actions, resulting not from poor skill level but from the mismatch between the machinery's design features and the operational environment, on the one side, and human capabilities.

Adverse factors in the operational environment provoke occupational diseases of operators, often resulting in permanent disability (Poroznuk et al., 2012). The operators of construction and road-making machinery are subjected to continuous vibration loads, noises, and dust, which can give rise to hand-arm vibration syndrome, hearing disturbances, diseases of peripheral nervous system, locomotor and respiratory diseases. Unfortunately,

the design of modern digging and road-making machinery fails to ensure protection of operators from adverse factors caused both by operational environment and the machines (Zorin et al., 2009).

A wide range of operational tasks require to limit the presence of operators on the production site (Kajita et al., 2006). Listed below are examples of construction and road-making machinery operation involving increased adverse effects:

- mining and concentrating mills, mines, ore dressing mills, open pits for commonly occurring minerals, construction material works;
- animal burial sites (anthrax);
- waste dumps, sanitary landfills for solid household waste and toxic chemicals;
- demolition of buildings and debris handling;
- mitigation of radiation accidents and incidents involving detection of uncontrolled radiation sources.

Various transport, digging, filling, compacting, and crushing machinery is used at the above sites, all subjected to aggressive environment and quickly going out of service.

If we consider a specific machine as a complex unit of equipment, the principal natural, climatic, and environmental factors produce the following impact on it:

- High temperature: reduced viscosity and modified structure of diesel fuel, lubricants, pressure and process fluids, impaired cooling of internal combustion engines, accelerated ageing of rubber seals and other insulating materials.
- Low temperature: increased viscosity of diesel fuel, congealed lubrication oils and solid greases, frozen condensate in pneumatic systems, reduced toughness of steels, hardened and embrittled rubber seals.
- Increased humidity: accelerated corrosion of steel parts, reduced insulation resistance, water intrusion into fuel and process liquids, mold build-ups.
- Reduced humidity: thickening lubrication oils, drying out seals, fracturing insulation materials.
- Sun radiation: changing coefficient of friction for friction materials, accelerated ageing of polymer coatings.
- Wind: drying of materials, increased heat output of machine parts and extra strain on them.
- Dust: changing coefficient of friction for friction materials, clogging of ducts and reduction of air flows, impaired cooling and ventilation, build-up on heated surfaces reduces heat exchange, and intensely heated items may become a source of ignition.
- Aggressive environment: accelerated deterioration of materials. The following groups of environments is among the most widespread: potent oxidizers (nitric, chromic acids etc.); mineral and organic acids (phosphoric, acetic acids, etc.); alkali; organic compounds (petroleum products, etc.); halogen compounds. Aggressive environments can produce chemical transformation, deterioration, cracking, stiffening, etc.

Therefore, research of machinery operation and operators' activities within the single man-machine-environment system emphasizes the importance of finding new ways to reduce the risk of human errors and enhanced operational safety of construction and road-making machinery.

2. Materials and Methods

Working under complicated operational conditions relies on operation, upkeep, and maintenance costs of machinery. If insufficient capacity, inappropriate or unreliable equipment is selected, early failures may occur which, under urgent work pressure, may prove to be critical.

Special operational conditions for machinery are accounted for at the stage of design and manufacture. Modern equipment is manufactured in various climatic options as regards their fitness for operation in various macroclimatic zones: for cold, moderate, arid, or humid tropical climate. Standard machinery greatly outnumbers the specialized options, which is due to overwhelming proportion of brown field areas with moderate climate, as well as to manufacturing industrial facilities, production cost of machinery etc. In this connection, it is necessary to adapt standard machinery to special operational conditions by means of special refit and by changing their modes of operation (Karakulev et al., 1991). Such necessity arises during operation of standard machinery in climatic areas with high temperature fluctuations or when it is required to operate such machinery in a variety of meteorologic conditions (Kudryavtsev, 2006).

To adapt process systems to their operational conditions, proven methods are used to refit such systems, thus

obtaining high efficiency of standard equipment under special conditions. The solutions improving the efficiency of machinery have to be coordinated against each factor affecting the productivity (purpose, operational environment, operating mode, technical condition, technologies deployed) and the duration of the machinery operation, as well as any possible variations of all these factors. Therefore, the aggregate range of impact for each specific factor builds into the set of positive/adverse factors affecting the productivity of process systems.

Analysis of the man-machine-environment system enables to suggest a way to reduce the impact of aggressive environment on such system:

- Develop a set of activities which have to include the selection and setup of specialized equipment;
- Remove the operator from the potential hazard area where the operations are carried out by implementing remote control;
- Comprehensive integration of technology to improve efficiency, safety, and enhance quality control of the processes and remake them into a single high added value production line.

This is what brings research and development of deploying modern remote control systems in construction and road making machinery to the top of the chart in order to improve the productivity of machinery, enhance the safety and quality control of the jobs carried out. There is a widespread solution for such tasks: a team of equipment enabling two operation modes, direct or remote, depending on the operational conditions (Michael, 2012; Martin, 2011).

3. Informative Part

1) Currently the electronic control systems of construction and road making machinery are monitoring and optimizing the operation of the engine, hydraulics, all sensors and operating controls, and ensure that information is shown on the display. The consistent operation of such electronic control systems is due to digital communication and control features applied. Operators may use the electronic control system to adjust the operating force and receive feedback about the condition of and load on such machinery resulting from interaction with the objects (Safford, 1973; Robert S. et al., 2002). Reliable feedback is ensured between the operator and equipment, to monitor the reaction force when actuators contact the working surface. The existing level of construction and road making machinery and the capabilities of radio electronic features enable creation of a set of radio devices which can be applied to provide remote control over operation of specialized machinery under a variety of conditions.

The capability to handle the necessary process operations is the key functionality for construction machinery remote control systems. It can only become possible subject to a fail-free control of the actuators of such machinery, which requires a homogeneity and optimization of operation for all units and modules. The construction machinery control system operation can divide into the following tasks:

- a) Principal: a set of control features to carry out the machinery operational cycle (its core function).
- b) Auxiliary: a set of auxiliary features enabling control between the operational cycles.
- c) Visual and spatial control of process operations. A system of cameras, microphones, positioning sensors, and data from the electronic control console enables visualization of parameters and positioning of the machinery and of the working members of its actuators.

The remote control system must be capable to handle the above tasks. At the same time, it must ensure feedback regarding the force impacting each of the actuators.

The coordination between the operations of such machinery can be ensured using the Master-Slave system already widely spread in modern construction equipment (Hirabayashi, 2006).

A typical Master-Slave system is a team of coordinated devices consisting of the following systems:

Master: controls one or several other devices (servo units).

Slave: is set up to operate under control, ensuring that the operating forces are applied consistently with the gearing diagrams of the operational equipment and using the installed feedback system, and features a set of information features to ensure detection of mechanical loads during the operations.

Master-Slave system enables control of motions of working members and sends signals describing such motions of each working member, enabling their realtime positioning (Hirabayashi, 2006). Feedback ensures adequate effect, based on the force applied by the operational equipment to the working surfaces. Such interaction ensures the degree of "sensitivity" between the components of the man-machine-environment system, thus promoting the process quality.

Also, there are a wide variety of remote controllers for construction and road-making machines. Examples are listed in the following patents (US №6219589B1/2001, EP №0976879B1/2006, US №5448479/1995, US №6907029B2/2005, WO №091395A2/2008, WO №078811A1/2002, US №7890235B2/2011, US №8272467B1/2012). Common feature in the present inventions is the using the remote control unit or organized remote stationary point for the remote control.

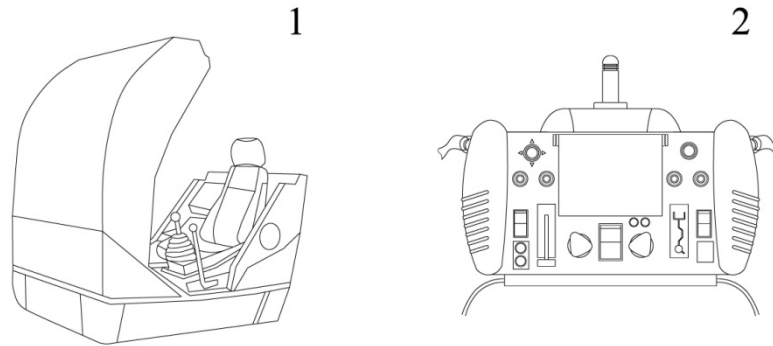


Figure 1. Remote controllers for construction and road-making machines: 1. the remote stationary point for the remote control; 2. the remote control unit

2) Generally speaking, the proposed modern single-purpose remote control systems for construction machinery operate as follows (fig. 2.): The electronic control unit 1 gathers (detects) and generates a control command $uk1(t)$ based on the inputs and source information $e(t)$ about the condition of the controlled equipment 3 taking into consideration the feedback channel 4 data. Then formed information $uk1(t)$ is supplied to the control channel 2, the operation of which is determined by radio RC, or laser channel LC. The control channel 2 is determined by the corresponding output data $uk2(t)$. Then commands are transmitted to the control object 3. As a consequence arising from this distortion $e1(t)$, the information $uk2(t)$ comes to the object may be slightly different from $uk1(t)$. Signals in the facility management are also affected the possibility of distortion $e2(t)$. The feedback channel 4 detects adequate effect, based on the force applied by the control object 3 to the working surfaces, and returns the operational data to the electronic control unit 1.

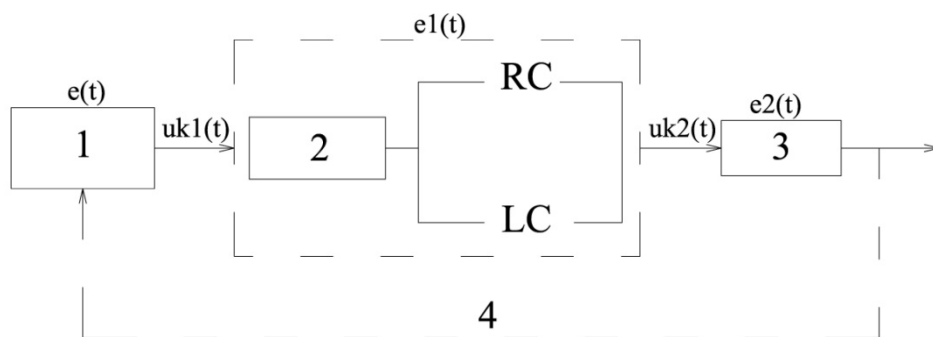


Figure 2. The scheme of a remote control system of construction machines

The necessity to use a wide radio frequency range to ensure a reliable communication channel without cross distortion or jamming, is a serious drawback of remote control systems. The operational conditions may impose limitations on the radio channel if side electronic noises are present (Ustinov et al., 2006).

3) To assess the ergonomic features created by application of construction machinery remote control, a 3D model of a crawler excavator using SolidWorks CAD software. The range of visibility is one of important ergonomic features for any vehicle. The range of visibility means the excavator's design feature describing the objective possibility and conditions for the operator to perceive visual information information for a safe and efficient control of the vehicle. To this end, the visibility diagrams were drawn for various operational conditions and

means of control (manual or remote) for such machinery.

Based on orthogonal drawings of the excavator, a 3D model is built based on the following list of parameters: viewing angle, eye transition and staring time, visibility sector. The principal visibility range indicator is the distribution of attention between the preferred and occasional surveillance items.

- the preferred surveillance items include: operational positions of the shovel, shovel boom, levees, vehicles etc.;
- the occasional surveillance items include: tracks, control console, information panels etc.

And to carry out the visibility range for the excavator under two control modes, visibility diagrams were drawn for civil structure dismantling. Dismantling a building in cities with an excavator is a widespread practice, as certain other controlled demolition techniques (such as pinpoint blast) sometimes are impossible to use. In this case the process visualization for different types of control will enable to feel the difference between the conditions of the two control modes (Volkov, 2013).



Figure 3. Process visibility from the excavator cabin

Inside the excavator cabin, a virtual point is selected to model the operator's glance, the visibility degree from the workplace will depend on the position of the cabin and operator's seat inside it, design of window apertures and quality of glasses and mirrors; it is also necessary to account for the possibility for the operator to change the position of his/her torso (e.g. upfront or leaning forward).



Figure 4. Process visibility from the remote control point

The remote control system grants a freedom of choice for the remote control point, thus enabling the operator to select both a convenient and safe position from which visual monitoring over the process operations is possible.

Therefore, the visibility diagrams enable to assess the adequate visibility of the preferred and occasional surveillance items under different control modes. The operator using remote control is able to continuously monitor the changing conditions of the working area and give a more substantiated judgment on such factors as safety and quality of work, thus improving the efficiency of the machinery operation. The working area image can be delivered using 3D models or with extra cameras and data read from computers, which, in turn, enhances the possibilities to monitor the process system as a whole (Steven & Curtis, 2012; CMT, 2013).

4. Discussion

We established that the design specifics of the construction and road-making machines enable them to go far beyond the standard process layout giving into a much wider range of application, up to aggressive and hazardous environments. Such conditions may become incompatible with the normal psychophysical loads on construction machinery operators, resulting in a variety of adverse effects. There are modern examples of successful deployment of remote control over machinery. In 2010 Brodrene Gjermundshaug Anlegg AS. was busy reclaiming the territory of a former military firing range on the territory of the actual Dovre National Park (Norway). The hazard consisted in the occurrence of many unexploded shells in the ground. One of the operators, Havard Thoessen, said: "It was quite a strange bit of experience, learning to do my normal job sitting in a steel box miles away from the place I am actually working at. It took me about two weeks to get used to the new way of working. First we had some difficulties to retain control over everything, but now there are no more problems" [CM, 2010]. This is the justification for remote control system implementation in operation of construction and road-making machines in aggressive environments, to reduce psychophysical loads on the operators and enhance their safety.

5. Conclusion

Operational and design are the two types of requirements applied to the construction machinery remote control systems.

Operational requirements consist in fail-safe and reliable operation of all remote control systems under the given weather and climatic conditions. The importance of this condition is due to the fact that the current development of equipment mostly targets its improved precision and implementation into control systems of high speed computers assuming an increasing amount of the operators' functions. Such control systems are complex and contain many different components. Whereas a failure of any single component may disturb the operation of the entire system, it is, therefore, of utmost importance that all components and the system as a whole should be highly reliable [Bogomolov, 2008].

Design requirements consist in quality of the installed features' operation. It must have minimal dimensions and weight, resist overloads, and be immune to vibration. These features should be operable under a wide range of temperatures, humidity, and pressure.

Remote control features fitted on the construction machinery will increase its base cost up to 30%. Taking into account the process operations carried out by the machinery and the conditions of such operation, the development of remote control sets for such machinery must be based on the value added. This consideration is viable both to design new machinery and to retrofit the existing equipment. For the latter, to avoid excessive costs, onboard equipment may be installed without any material redesign of the machinery. The consideration of extra costs is overshadowed by the totally different level of safety and comfort offered by the remote control systems.

At this stage of the research we established that the design specifics of the construction and road-making machines enable them to go far beyond the standard process layout giving into a much wider range of application, up to aggressive and hazardous environments. Such conditions may become incompatible with the normal psychophysical loads on construction machinery operators, resulting in a variety of adverse effects. Justification was provided for remote control system implementation in operation of construction and road-making machines in aggressive environments, to reduce psychophysical loads on the operators and enhance their safety.

Further research will focus on the technical aspects of the project safe and effective system of remote control of Construction and Road-Making Machines.

References

- Bogomolov, A. A. (2008). Technical bases of creation of machines. Belgorod: BSTU named after V.G. Shukhov. Cat Mining Tecnology. Retrieved July 5, 2013, from <https://mining.cat.com/miningtechnology>.
- Hirabayashi T. (2006). Name lost. *KENSETSU KIKAI Magazine*, 42(6), 27-32.
- Kajita Shigeo, Awano Katsusuke, Tozawa Shoji, Nishikawa Hiroyasu. Remote radio operating system, and remote operating apparatus, mobile relay station and radio mobile working machine. Patent No.: EP 0 976 879 B1. Date of a Patent: 20.12.2006.
- Karakulev, A.V., Ilyin, M. W., & Markelanec, O. V. (1991). Operation and construction of working equipment and machines. Moscow: Transport.
- Khmara, L. A., & Shatov. S. V. (2012). Determination of mechanization for demolition of ruined buildings on the basis of dam structure analysis. *Mechanization in construction*, 1, 34-38.
- Kudryavtsev, E. M. (2006). *Complex mechanization of construction* (p. 424). Moscow: Publishing house of the Association of civil engineering universities.
- Martin Carlsson. Control system for a remote control work machine. Patent No.: US 2011/0282519 A1. Date of a Patent: Nov, 17.2011.
- Munipov, V. M., & Zinchenko, V. P. (2001). Bases of ergonomics. Moscow: Logos.
- Poroznuk, L. A., Vasylenko, V. A., & Poroznuk, E. C. (2012). Role of environmental audit in the waste management in the Belgorod region (pp. 177-180). Vestnik of the University (issue 4), BSTU named after V.G. Shukhov.
- Robert S., James M., Leif A. Remote Control System. Patent No.: WO 02/061515 A2. Date of a Patent: Jan, 29.2002.
- Romanovich, A. A., & Kharlamov, E. V. (2009). Technical maintenance of construction machines. Belgorod: BSTU named after V.G. Shukhov.
- Scopylatov I. A., & Yefremov O. Y. Personnel management. Bibliotekar.ru. Retrieved January 18, 2013 from www.bibliotekar.ru/biznes-33/24
- Safford, Edward Zenfesty. Radio control manual / S.E.Zenfesty VII, 192 p.p., ill., 1.25 sh «Brif Nat. Bibliogr.» 1973. - Slough: Fonlskam, 1973. -№ 1231: 19.
- Steven E. Nielsen, Curtis Chambers. Virtual white lines for delimiting planned excavation sites. Patent No.: US8, 218,827 B2. Date of a Patent: Jul, 10.2012.
- Ustinov, U. F., Teplyakov, I. M., Cononov, A. A., & Avdeev, U. V. (2006). The problem of constructing remote control systems for construction machines. *Universities's news*, 1, 83-86.
- Volkov, E. A. (2013). Study of ergonomic technological performance machines, as a factor in assessing the effectiveness of system "man-machine-environment". *Industrial safety journal*, 6, 32-34.
- Zorin, V. A., Daugelo, V. A., & Sevrjugina, N. C. (2009). *Safety requirements to ground-based transport systems*. Belgorod: BSTU named. V.G. Shukhov.
- CAT Magazine Issue. (2010). Shaping the future by remote control. Date Views. Retrieved September 14, 2012, from <http://mining.cat.com/miningtechnology>
- Michael A. Staab. (2012). Remotely controlled backhoe. Patent No.: US8,272,467 B1. Date of a Patent: Sep, 25.2012.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

Design of PID Filter Controller with Genetic Algorithm for MIMO System in Modern Power Generation

Anitha Mary. X¹, L. Sivakumar² & J. Jayakumar³

¹ Research scholar, Department of Electronics and Instrumentation Engineering, Karunya University, Coimbatore, India

² Vice Principal and Dean R&D, Sri Krishna College of Engineering and Technology, Coimbatore and Formerly General Manager, Corporate R&D, Bharath Heavy Electricals Limited, Hyderabad, India

³ Associate professor, Department of Electrical and Electronics Engineering, Karunya University, Coimbatore, India

Correspondence: Anitha Mary. X, Research scholar, Department of Electronics and Instrumentation Engineering, Karunya University, Coimbatore, India. E-mail: anithajohnson2003@gmail.com

Received: July 8, 2014 Accepted: July 24, 2014 Online Published: September 1, 2014

doi:10.5539/mas.v8n5p186 URL: <http://dx.doi.org/10.5539/mas.v8n5p186>

Abstract

In this work, a new technique based on Genetic Algorithm for designing multivariable PID filter controller has been developed and applied to gasifier control of ALSTOM benchmark challenge II. The coal gasifier is the main component in Modern power generation. Coal gasifier involves several performance and robustness requirements in addition to actuator constraints under three operating loads (no-load, 50% and 100% load). The proposed GA optimises the tuning parameters of PID constants in terms of robustness and performance. The optimised controller meets all design objectives under all operating conditions. Robustness of the controller is tested for step and sinusoidal pressure disturbances applied at the inlet of throttle valve along with increase and decrease of calorific value of fuel fed-in (coal). Simulation results obtained confirmed the superiority of proposed technique for gasifier problems.

Keywords: ALSTOM gasifier, MIMO system, gasifier performance, gasifier control, PIDF controller, Genetic Algorithm

1. Introduction

Coal gasifier plays an important part in clean coal power generation. It converts coal into syngas under high temperature and pressure. Control of gasifier becomes vital in producing syngas with higher efficiency. In this context, ALSTOM, the UK power generation centre, posed the benchmark challenge II to design controller for gasifier that incorporates pressure disturbance test as well as coal quality variation test. The MATLAB SIMULINK model (Dixon et. al., 2006) given by ALSTOM controller design should satisfy the design objective for output magnitude and rate of constraints at the input under three operating loads (no-load, 50% and 100% load). In this paper, optimised Proportional Integral and Derivative filter controller is designed. Even though, many optimisation algorithms exist, Genetic Algorithm is mainly used to solve global optimization problems existing in power systems. In spite of its high computation time, GA based techniques are highly preferred, because GA works with population of solutions rather than with single solution. In this paper, the parameters of Proportional Integral Derivative controller with filter approach are optimised using Genetic Algorithm and multiobjective problem existing in gasifier is converted into single optimisation problem and can be taken as objective function.

1.1 Gasifier Plant

Gasifier is a chemical reactor with five inputs (coal flow, air flow, steam flow, limestone flow and char extraction flow) and four outputs (pressure, temperature and calorific value of syngas and bedmass). Coal reacts with steam and air to produce low calorific value fuel gas and char. Limestone is added to capture sulphur content in the coal. Oxygen in the fluidising air combine with carbon present in char to form carbon monoxide and carbon dioxide. Though several endothermic and exothermic equations occur, the main equations involved in gasifier are





Equations 1 and 2 are exothermic gasification.

The carbon-dioxide reacts more with carbon to form carbon-monoxide. Also steam reacts with carbon to form carbon-monoxide and hydrogen.



Equations 3 and 4 are endothermic reactions.

The un-reacted char is added to the bed, which is maintained at a constant height and will be removed periodically. The fuel gas is filtered and combusted in a gas turbine to generate electricity. As a result, one of the disturbances is a change of downstream pressure (Pressure test) at the gas turbine throttle valve. The coal quality of syngas also affects the power generation (model error test). The objective of benchmark challenge II is to control the gasifier system with step and sinusoidal pressure disturbances (Psink) with increased and decreased coal quality variations. The objective is to control the gasifier maintaining the steady state value of output variable within the limit as well as maintaining the rate of constraints at the input as shown in table 1 and 2. A group of authors attempted different methods such as Model predictive control (Al Seyab et. al., 2006), H_∞ (Gatley S.L, 2006), optimal PI controllers (Gatley et. al., 2004, Simm et. al., 2006, Koteeswaran et. al., 2014, Xue Y et. al., 2010), Fuzzy gain scheduled controller (Yong wang, et. al., 2009) and suggested suitable controllers. While these methods have provided desirable solutions, certain segments remained unattempted especially with respect to coal quality variations as shown in table 3.

Table 1. Output variables with allowable limits

Output variables	Steady state value for 100% load	Steady state value for 50% load	Steady state value for 0% load	Limits
Calorific value of syngas (CVGAS) in MJ/Kg	4.36	4.49	4.71	± 0.01
Bed mass(MASS) in Kg	10000	10000	10000	± 500
Fuel gas pressure (PGAS) in bar	20.1	15.5	11.2	± 0.1
Fuel gas temperature (TGAS) in K	1223.1	1181.1	1115.1	± 1

Table 2. Input variables with allowable limits

Input variable	Maximum value in Kg/s	Minimum value in Kg/s	Peak rate in kg/s ²
Coal flow (WCOL)	10	0	0.2
Air flow (WAIR)	20	0	1.0
Steam flow (WSTM)	6	0	1.0
Char Extraction (WCHAR)	3.5	0	0.2

Table 3. Various controller methods for ALSTOM benchmark challenge II

S.No.	Authors	Controller methods	Change in calorific value with a wide band of $\pm 18\%$ at three operating loads- 0%, 50% and 100% (Coal quality test) and pressure disturbance test
1	Anthony Simms et. al.,	Multi objective optimization approach	Results are not shown for coal quality test. Results are available only for pressure disturbance test.
2.	Sarah Gatleyet. al.,	H-infinity design approach	Results are not shown for coal quality test. Results are available only for pressure disturbance test.
3.	Wilson et. al.,	state estimators to improve on the base line performance	Results shown for +18% coal quality variations
4.	Y. Cao et. al.,	Model Predictive controller	Results shown for 100% load exceeding the specified output limit

5	Yong Wang	Study on Fuzzy Gain-Scheduled Multiple Mode Predictive Control of ALSTOM Gasifier Problem	Results not shown for coal quality test. Results are available only for pressure disturbance test.
6	R.Kotteswaran and L.Sivakumar	Performance evaluation of optimal PI controller for ALSTOM gasifier during coal quality variations	Results shown for +18% to -7% for sinusoidal disturbance at 0% load .

2. Proposed Genetic Algorithm based PID Filter Controller

PID controllers are widely used for complex chemical processes and engineering systems.

The structure of PID controller is given by

$$C(s) = K_p(1 + \frac{1}{sT_i} + T_d s) = P(1 + I(\frac{1}{s}) + Ds) \quad (5)$$

However, one of the most common problems associated with PID is with the synthesis of derivative action. The ideal derivative has very high gain and susceptible for noise accentuation (Aström et. al., 1995). Hence the authors have chosen PID filter controller whose derivative action is represented as $D = \frac{K_d s}{1+sT_f}$. Here T_f is called filtering time. The transfer function of a PID controller with a filtered derivative is given in equation (6)

$$C(s) = K_p(1 + \frac{1}{sT_i} + \frac{sT_d}{1+s\frac{T_d}{N}}) = P(1 + I(\frac{1}{s}) + D(\frac{Ns}{s+N})) \quad (6)$$

and are schematically shown in Figure 1.

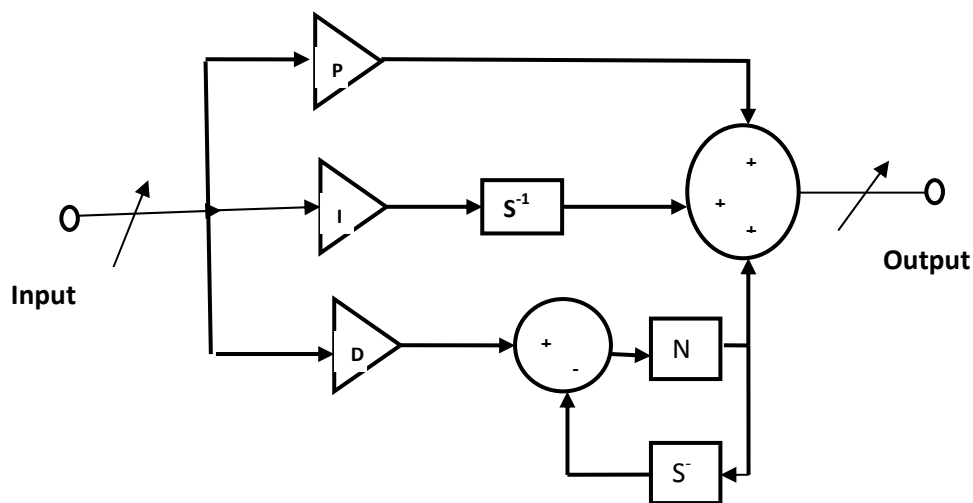


Figure 1. Schematic for PID filter controller

2.1 Problem Formulation and Implementation

PID tuning can be performed using techniques like empirical methods such as Zeigler Nicholas method (Aström et. al., 1995), analytical methods like root locus technique (Blasco et. al., 2000) and optimisation methods such as Lopez and Ciancone methods (Marlin et. al., 1995). The PID values obtained through these methods can be applied to a system operating in a particular operating point. When the system is operating under different operating zones, genetic algorithms can be used to tune PID parameters taking all non linearities and process characteristics into account.

Genetic Algorithms are the optimisation techniques which apply the law of natural selection to achieve population in a search space (Deepa et. al., 2009). The search space is the objective function. They use probabilistic transition method to obtain population of solution called individuals or chromosomes that evolve iteratively. Each iteration is called generation.

2.2 Objective Function for Pressure and Coal Quality Disturbance

For the proposed PID filter controller, step disturbance in Psink is applied to closed loop system and IAE (Integral Absolute errors) are calculated for over 300 seconds. The objective function for step and sinusoidal disturbance in Psink are given in equations (7) and (8).

$$f_1(x)_{\text{step}} = \sum_{j=1}^3 \sum_{i=1}^4 \int_0^{300} |y'_{isp}(t) - y'_i(t)| \quad (7)$$

$$f_2(x)_{\text{sine}} = \sum_{j=1}^3 \sum_{i=1}^4 \int_0^{300} |y'_{isp}(t) - y'_i(t)| \quad (8)$$

similarly the objective function for coal quality change is given in equation (9).

$$f_3(x)_{\text{CV of coal}} = \sum_{j=1}^3 \sum_{i=1}^4 \int_0^{300} |y'_{isp}(t) - y'_i(t)| \quad (9)$$

where $f_1(x)$ step is the objective function for step disturbance of -0.2 bar applied at Psink

$f_2(x)$ sine is the objective function of sinusoidal disturbance of amplitude 0.2 bar and 0.04Hz frequency applied at Psink. $f_3(x)$ cv of coal is the objective function for disturbance at fuel fed-in.

$y'_{isp}(t)$ is the steady state value for output number i at operating load.

$i=1$ means CV of syngas; $i=2$ means bedmass output; $i=3$ means pressure output of the syngas; $i=4$ means temperature output for syngas; also $j=1$ means 100% load; $j=2$ means 50% load and $j=3$ means 0% load.

$y_{ij}(t)$ = measured output value at the three operating loads.

$$D(x) = f_1(x)_{\text{step}} + f_2(x)_{\text{sine}} + f_3(x)_{\text{CV of coal}} \text{ is the fitness value.}$$

The objective is to minimise $D(x)$

2.3 Objective Function for Output Constraints

When the disturbances are applied, the controller must be tuned in such a way that output limits should not exceed.

$$C_{\text{step}} = \frac{\max_i \max_j \|y'_i - y'_{isp}\|}{D_i} \quad (10)$$

$$C_{\text{sine}} = \frac{\max_i \max_j \|y'_i - y'_{isp}\|}{D_i} \quad (11)$$

where y'_i = measured variable for output i at operating point j

y'_{isp} = steady state value for output i at operating point j .

D_i = allowable deviation of output i

Combining equations (10) and (11), the output objective function is given by

$$O = \max(C_{\text{step}}, C_{\text{sine}})$$

Therefore, the overall objective function is to minimise $D(x)$ if $O < 1$.

The procedure for optimising PID filter controller with genetic algorithm is given below:

1. The PID tuning parameters (P,I,D,N) must be encoded in real numbers or vectors or binary strings
2. Population size and limits are noted
3. Normalised Geometric selection is applied to select any random values of parameters based on fitness value.
4. Reproduce the selected parameters to get optimised solution.
5. Arithmetic crossover and uniform mutation are performed to alter the parameters to optimised values.
6. Calculate the fitness value $D(x)$ for each iteration
7. Repeat steps 8-10 for 'n' off springs
8. Using fitness function, find value of error in the Generation.
9. The parameters with highest fitness value are chosen as the final parameter values.
10. If the obtained values are not up to the mark, repeat step 2

The flow chart for GA based PID tuning is shown in Figure 2 and tuning parameters obtained by proposed and other methods are given in Table 4.

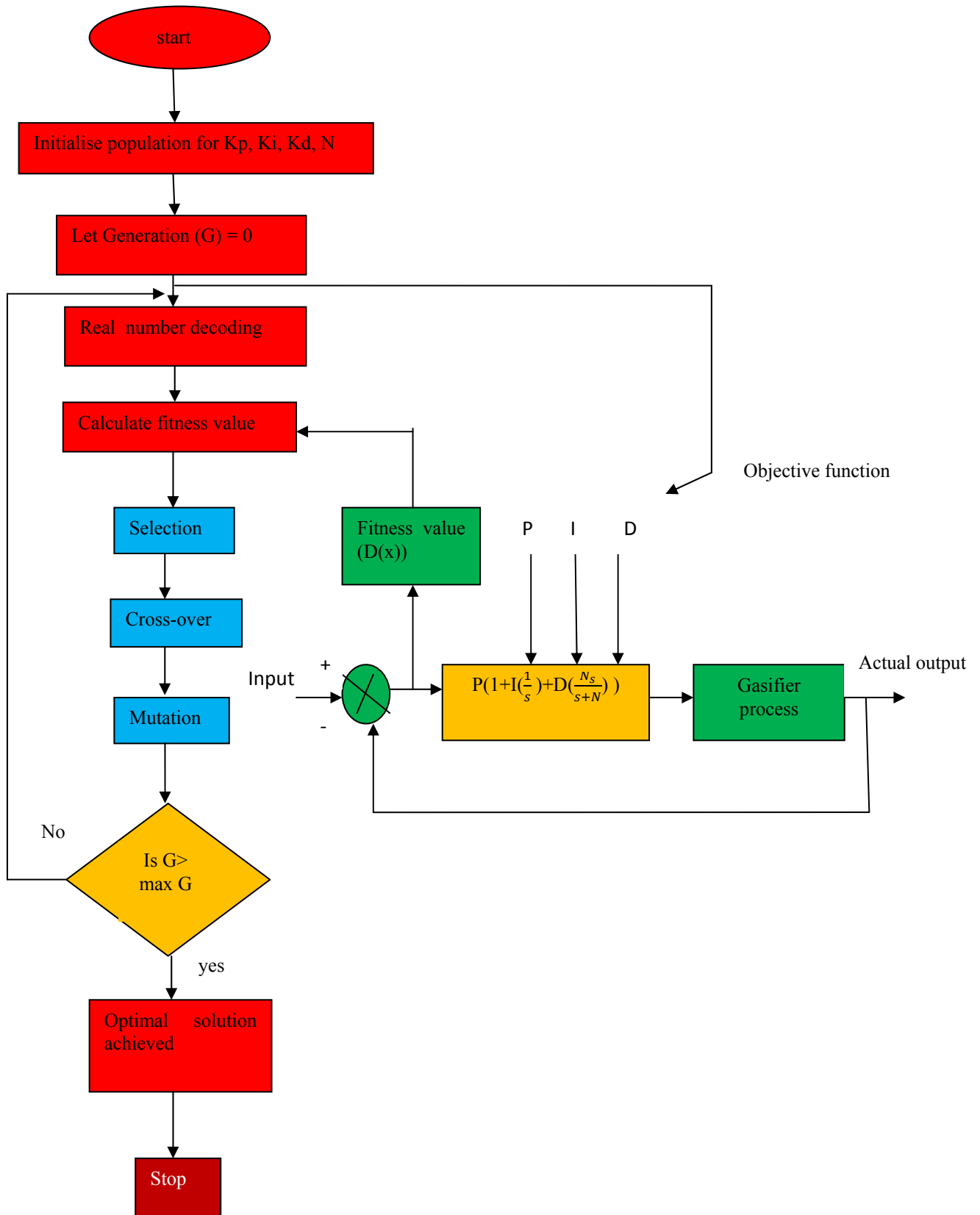


Figure 2. Flow chart for PID tuning using Genetic Algorithm

Table 4. Tuning parameters by various controller methods

Parameter	Baseline PI[Dixon 2006]	Multi objective PI controller [Xue et. al., 2010]	Optimal PI [Kooteswaran et. al., 2014]	Multivariable PID [Farag 2006]	Proposed GA based PIDF controller
CV_Kp	-0.1226e-03	-0.016972	-0.2937e-03	0.000044	-0.002098
CV_Ki	0.80e-03	-0.024813	0.747e-03	0.000068	0.000362
CV_Kd	-	-	-	0	0 N=100
BM_Kp	0.145070	0.18498	0.227116	-0.000367	0.000260
BM_Ki	1.032797	1.741	1.857655	-0.000113	0.000147
BM-Kd	-	-	-	0	0.2163021 N=100
Pr_Kp	0.201e-03	0.0003055	0.1558e-03	1.16e-05	0.000189
Pr_Ki	0.656e-04	0.00001077	0.51e-04	0.000118	0.000011
Pr_Kd	-	-	-	0.00026	0.00001 N=100
Tg_Kp	1.701288	2.2825	1.692696	2.622e-02	1.724918
Tg_Ki	0.009479	0.097237	0.009555	0.3881	0.009927
Tg_Kd	-	-	-	0.512	0.151923 N=0.001574

3. Simulation Results and Discussion

The performance of gasifier during coal quality along with step and sinusoidal disturbance in PSINK is of prime concern according to challenge II. Coal quality is allowed to change incrementally within the range $\pm 18\%$ with respect to design value of the coal, and the transient performance of the gasifier output variables are monitored using MATLAB/ SIMULINK simulation tools. A step disturbance of -0.2 bar from the steady value of Psink and sinusoidal disturbance of 0.2 bar as amplitude and 0.04 Hz are applied along with $\pm 18\%$ calorific value of coal. Further, an auto tuning option has been chosen.

The simulation responses pertaining to the change in calorific value of the fuel along with step and sinusoidal disturbance in Psink are shown in figures 3 -14. For the purpose of analysis, the input and output limits within which the input and output variables should lie during transient region are shown in Table 1 and 2.

3.1 Step Disturbance at Psink Coupled with $\pm 18\%$ CV of Coal Variation

Figure 3 shows that the output variables (pressure, temperature and calorific value of the syngas) are reaching the respective set point values (11.2 bar, 1115.1 K and 4.71 MJ/Kg) corresponding to 0% load. Figure 4 shows that the input variable flow rates for coal air and steam are also within the allowable limits corresponding to 0% load. Similar figures for input and output variables corresponding to 50% and 100% loads are shown in Figure 4-8. It is observed that Coal flow rates are deviating from the allowable band at 100% load for coal quality variation in the negative direction.

3.2 Sinusoidal Disturbance Coupled with Coal Quality Variations

Figure 9 shows that the output variables (pressure, temperature and calorific value of the syngas) are reaching the respective set point values (20 bar, 1223.2 K and 4.36 MJ/Kg) corresponding to 100% load. The input variable flow rates for coal air and steam are also within the allowable limits corresponding to 100% load. Similar figures for input and output variables corresponding to 50% and 0% loads are shown in Figure 10 to Figure 14.

However the following deviations have been observed during the sinusoidal disturbances:

- Coal and steam flow rates are deviating from the allowable band at 0% load for coal quality variation in the negative direction.
- Temperatures of syngas are not within the limit for 100% load when the coal quality is increased to +18%.

It has been observed that GA based PIDF controller applied to lower order modelling provides better results during all situations as compared to (Dixon et. al., 2002, Simm et. al., 2006, Farag et. al., 2006, Kooteswaran et. al., 2014).

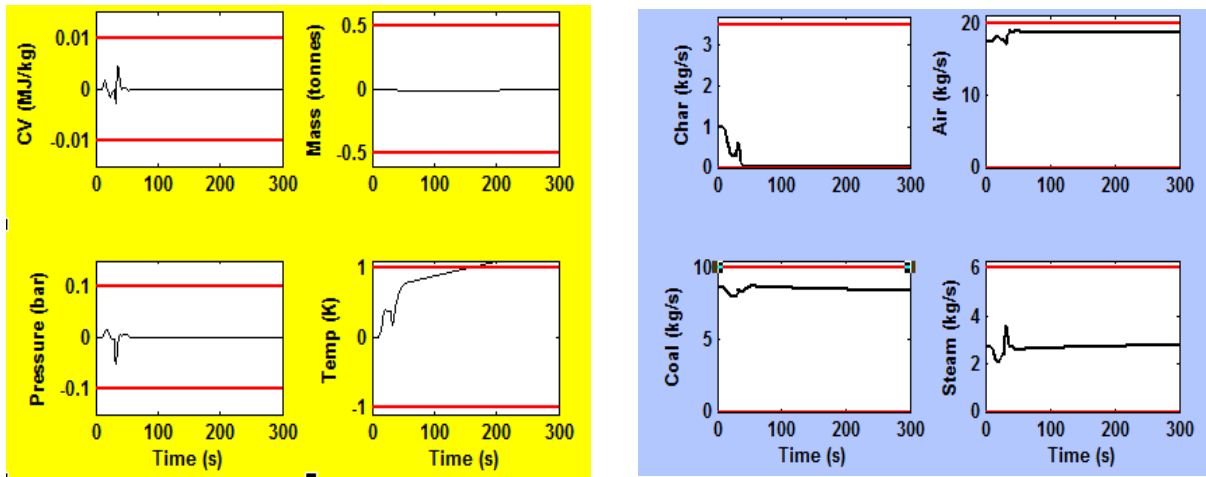


Figure 3. Output and Input response for +18% coal quality change with step disturbance for 100% load

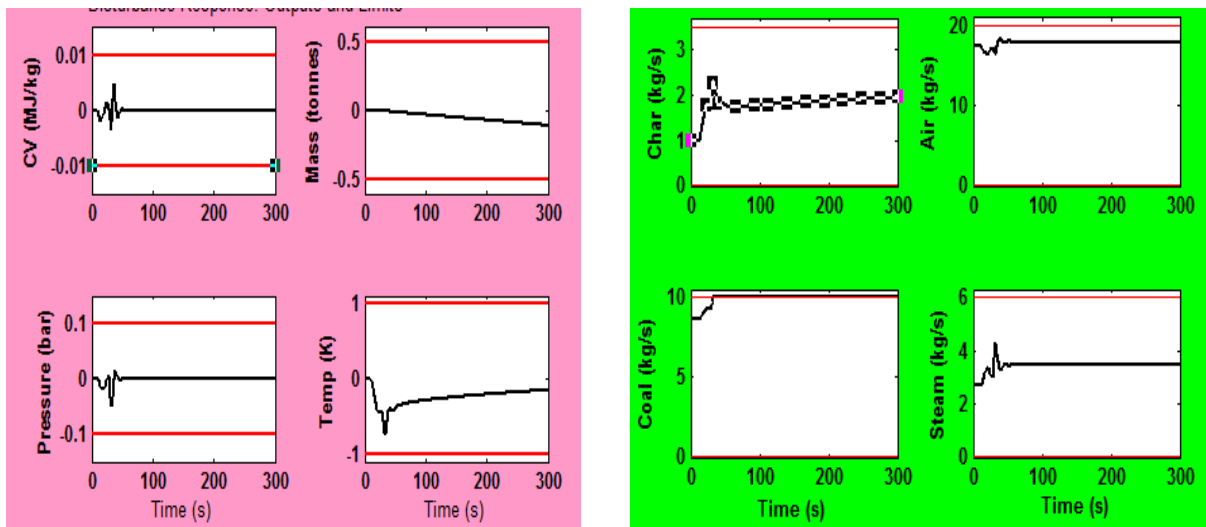


Figure 4. Output and Input response for -18% coal quality change with step disturbance for 100% load

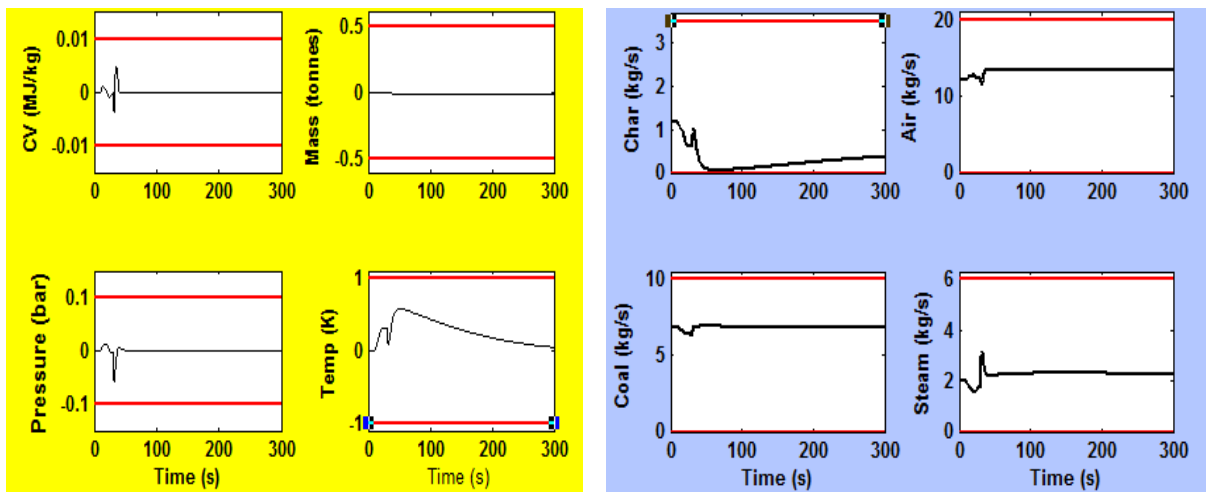


Figure 5. Output and Input response for +18% coal quality change with step disturbance for 50% load

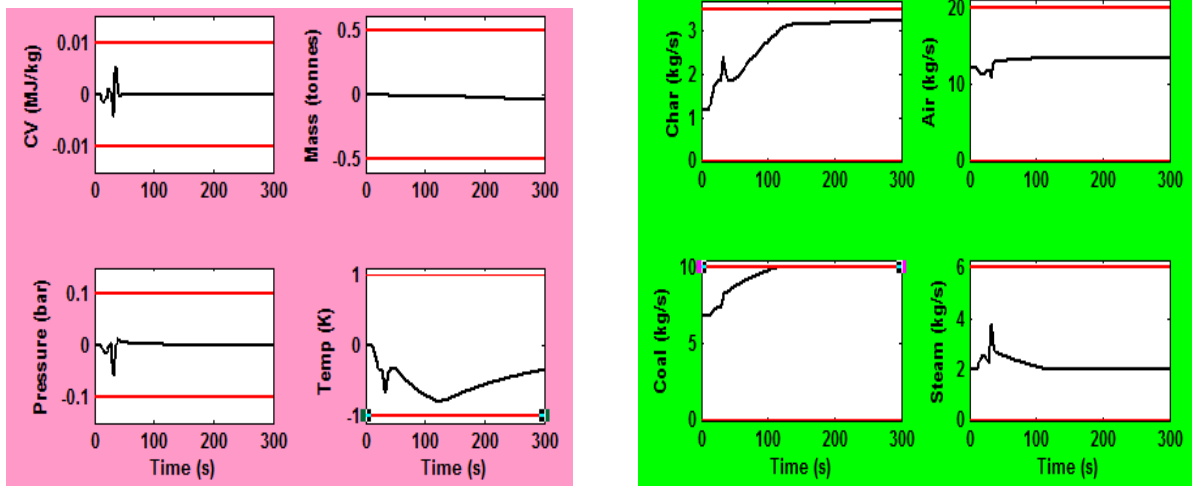


Figure 6. Output and Input response for -18% coal quality change with step disturbance for 50% load

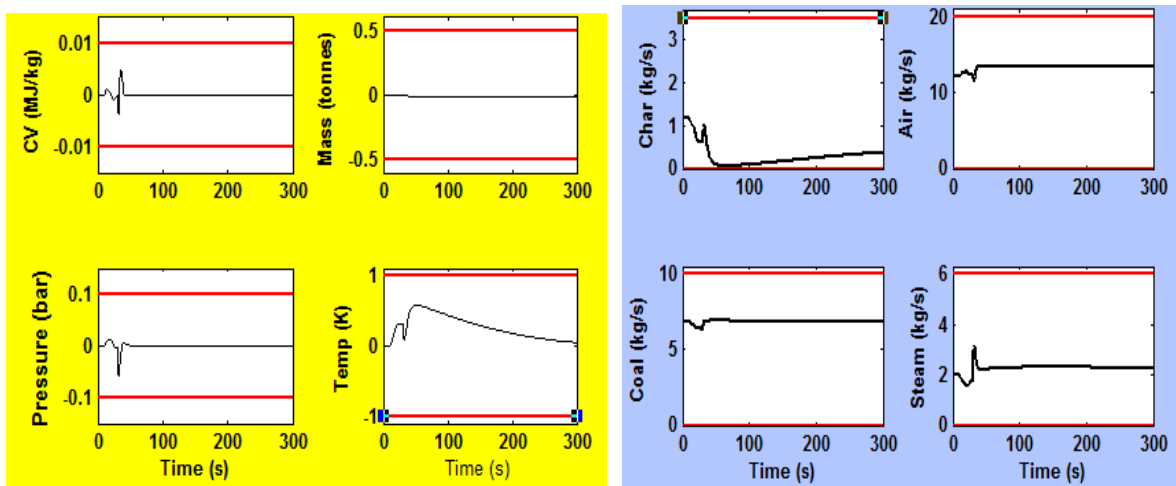


Figure 7. Output and Input response for +18% coal quality change with step disturbance for 0% load

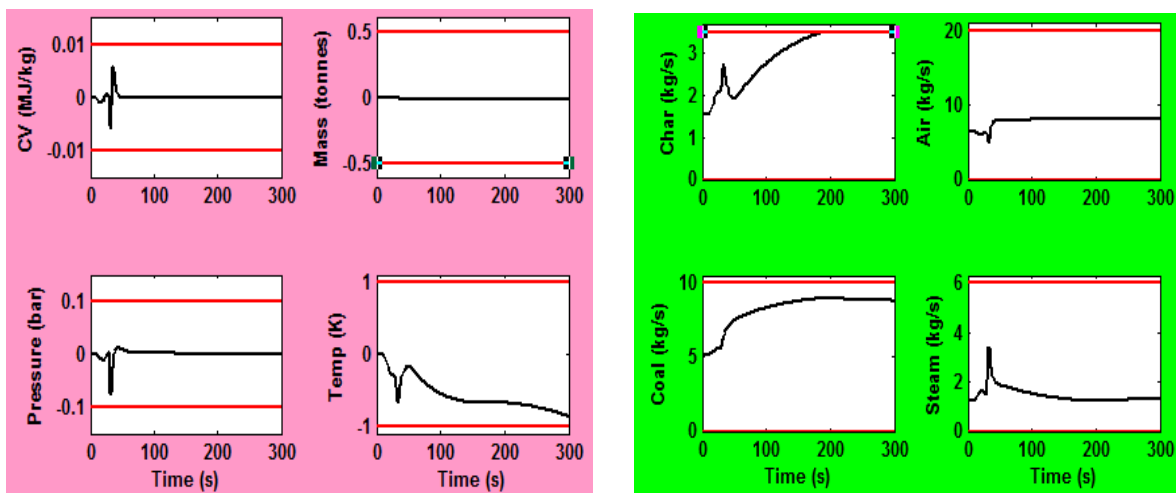


Figure 8. Output and Input response for -18% coal quality change with step disturbance for 0% load

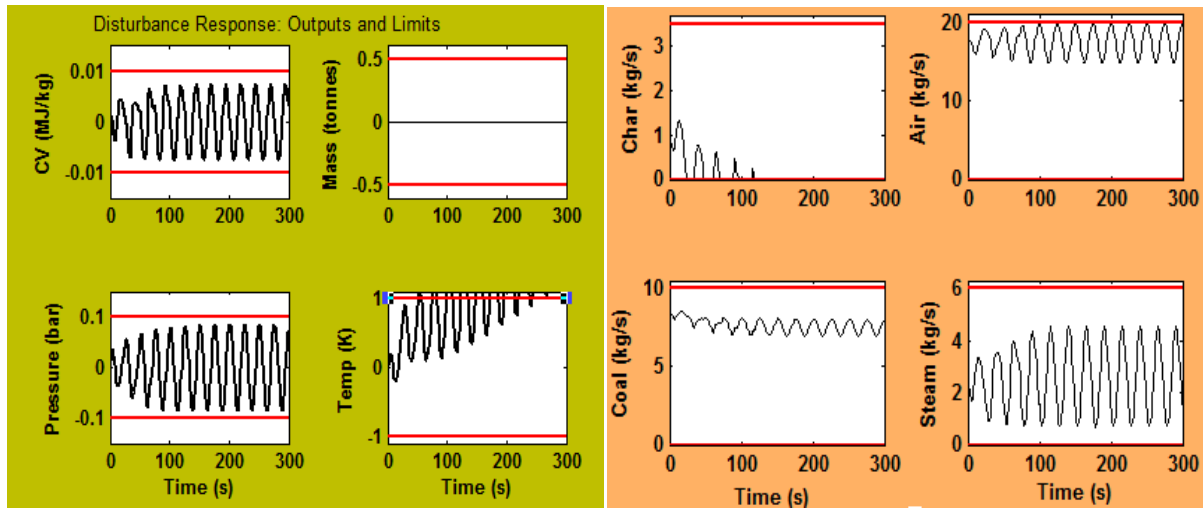


Figure 9. Output and Input response for +18% coal quality change with sinusoidal disturbance for 100% load

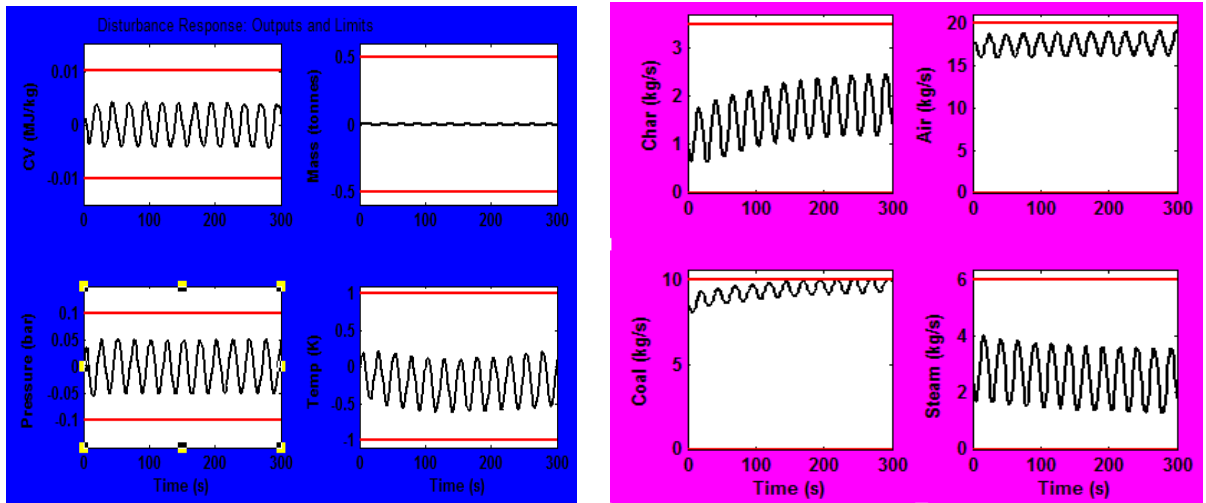


Figure 10. Output and Input response for -18% coal quality change with sinusoidal disturbance for 100% load

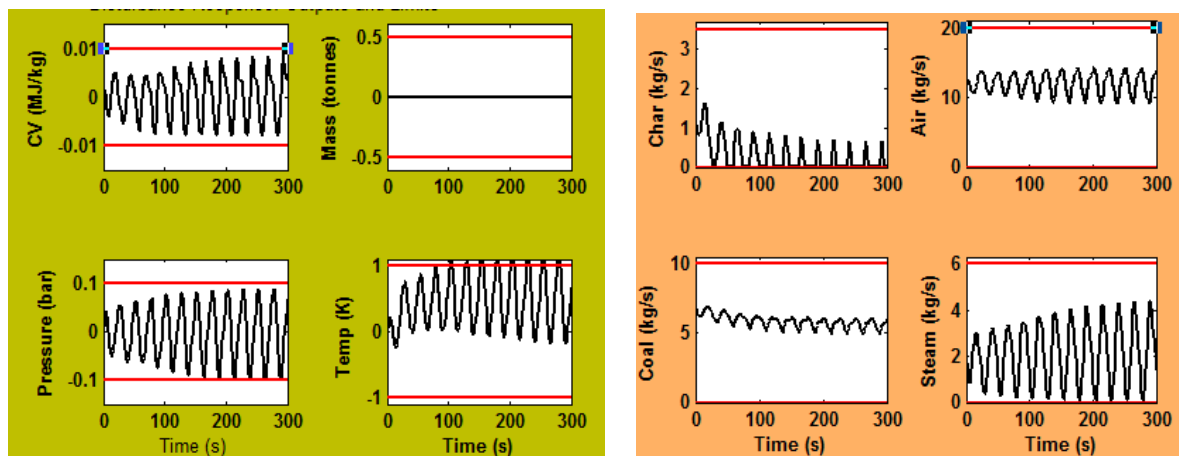


Figure 11. Output and Input response for +18% coal quality change with sinusoidal disturbance for 50% load

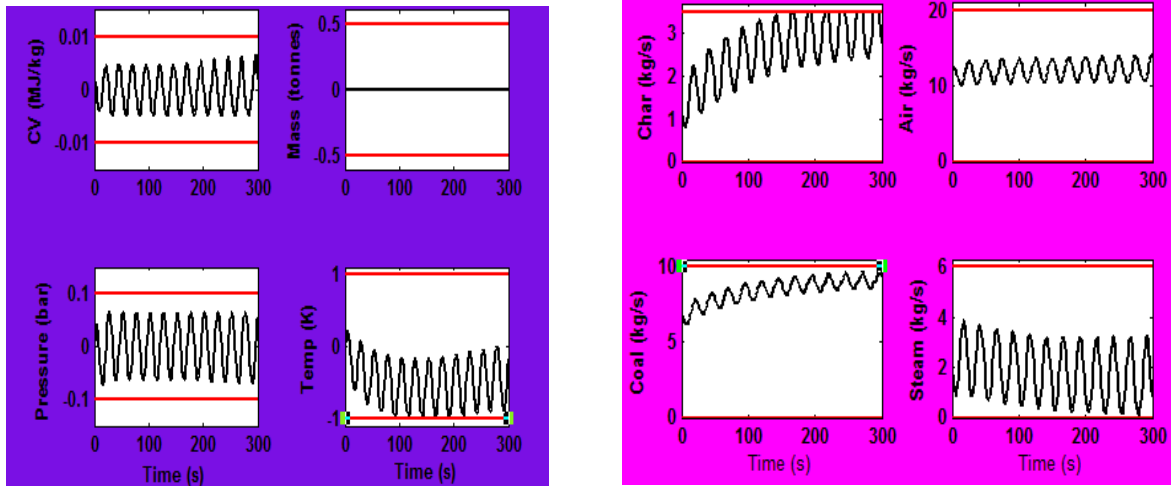


Figure 12. Output and Input response for -18% coal quality change with sinusoidal disturbance for 50% load

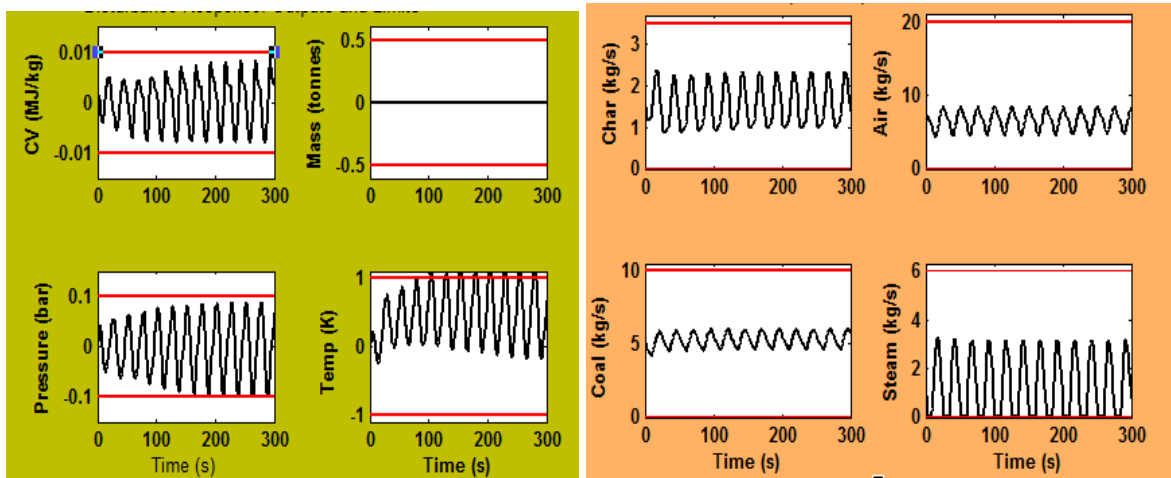


Figure 13. Output and Input response for +18% coal quality change with step disturbance for 0% load

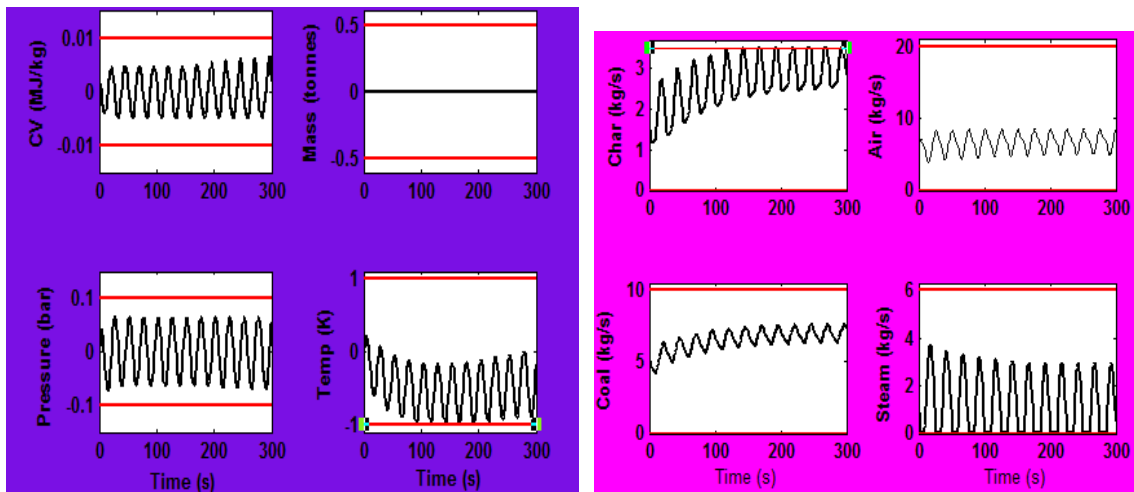


Figure 14. Output and Input response for -18% coal quality change with step disturbance for 0% load

4. Conclusion

Considering the four inputs (char flow rate, coal flow rate, air flow rate and steam flow rate) and four outputs (pressure, temperature, calorific value of syngas and bed mass), an appropriate 4x4 MIMO model has been formed for gasifier. PID controllers have been augmented with these models and a simulation setup has been made in MATLAB environment for various disturbance analysis. The overshoot and under shoot related to various process parameters such as pressure, temperature and calorific value of syngas (the transient performance requirements of gasifier) are found to be well within the limits during step and sinusoidal variations at Psink (output side) along with coal quality variations for different loads (0%, 50% and 100%). The variation and rate of variation of the input variables (manipulated variables) are also found to be well within the specified limits. These results fulfilled the requirement of challenge problem II.

Acknowledgement

The authors would like to thank Prof. R. Dixon, Loughborough university for his support and through email discussions. The authors are grateful to the Management of Karunya University and Sri Krishna College of Engineering and Technology, Coimbatore for their encouragement and support.

References

- Al Seyab, R. K., & Cao, Y. (2006). Nonlinear model predictive control for the ALSTOM gasifier. *Journal of Process Control*, 16(8), 795–808. <http://dx.doi.org/10.1016/j.jprocont.2006.03.003>
- Aström, K. J., & Hägglund, T. (1995). *PID Controllers: Theory, Design, and Tuning*. Instrument Society of America.
- Blasco, F., Xavier, M., Martínez, J., Senent & Sanchis, J. (2000). *Sistemas Automáticos*. Editorial U.P.V.
- Dixon, R., & Pik AW. (2006). Alstom Benchmark challenge II on Gasifier control. *IEE proceedings- control theory and applications*, 153(3), 254-261. <http://dx.doi.org/10.1049/ip-cta:20050062>
- Farag, & Werner, H. (2006). Structure selection and tuning of multivariable PID controllers for industrial benchmark problem. *IEE Proceedings - Control Theory and Applications*, 153(3), 262–267. <http://dx.doi.org/10.1049/ip-cta:20050061>
- Gatley, S. L., Bates, D. G., & Postleth, W. (2004). H infinity control and anti windup compensation of the nonlinear ALSTOM gasifier problem. *Proceedings of the institution of Mechanical Engineers*, 214, 453-468.
- Marlin, T. E. (1995). *Process Control. Designing Processes and Control Systems for Dynamic Performance*. Mc Graw-Hill.
- Kotteeswaran, R., & Sivakumar, L. (2014). Performance Evaluation of Optimal PI Controller for ALSTOM Gasifier during Coal Quality variations. *Journal of Process Control*, 24(1), 17-26.
- Sivanandam, S. N., & Deepa, S. N. (2009). A Comparative Study Using Genetic Algorithm and Particle Swarm Optimization for Lower Order System Modelling International Journal of the Computer. *The Internet and Management*, 17(3), 1-10.
- Simm, A., & Liu, GP. (2006). Improving the performance of the ALSTOM baseline controller using multiobjective optimisation. *IEE Proceedings - Control Theory and Applications*, 153(3), 286–292. <http://dx.doi.org/10.1049/ip-cta:20050131>
- Xue, Y, Li, D., & Gao, F, (2010). Multi-objective optimization and selection for the PI control of ALSTOM gasifier problem. *Control Engineering Practice*, 18(1), 67-76. <http://dx.doi.org/10.1016/j.conengprac.2009.09.004>
- Yong, W., & Junhong, Y. (2009). *Study on fuzzy gain scheduled multiple mode predictive control of ALSTOM gasifier problem*, Proceeding of 2009 IEEE International conference on Mechatronics and Automation August 9-12 Changchun, china.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

About the Method of Analysis of Economic Correlations by Differentiation of Spline Models

Ruslan Hizrailevich Ilyasov¹

¹The Chechen State University, Russia

Correspondence: Ruslan Hizrailevich Ilyasov, The Chechen State University, Sheripova Street, 32, Grozny, Chechen Republic 364907, Russia.

Received: July 3, 2014 Accepted: July 5, 2014 Online Published: September 4, 2014

doi:10.5539/mas.v8n5p197

URL: <http://dx.doi.org/10.5539/mas.v8n5p197>

Abstract

The article considers spline approximation as one of efficient methods of modeling economic dynamics. Spline approximation of economic dynamics allows carrying out qualitative and accurate transition from discrete values of a lattice function to a continuous model of a process, which allows calculating values of a studied index at any time point (interpolation). Spline representation improves the quality of economic dynamics modeling while saving the real values of the studied process at each time point. In this article, differentiation of spline models is used for analysis of the economic indexes growth rate. Correlations are detected and itemized by comparison of derivatives. The possibility of detecting "latent trends" is demonstrated by differentiation of spline models of the dynamics using the example of economic indexes of the oil and gas market of Russia. For example, in the first case, we consider spline models of the dynamics of export prices for oil and natural gas. Here, the correlation of the studied indexes is obvious and is detected by both calculation of the correlation ratio and visualization of the studied rows of dynamics with spline models. As an opposite example, we consider the dynamics of the volumes of oil and natural gas export. In this case, we gain the correlation ratio close to zero, which is to evidence absence of correlation. Modeling of the studied dynamics with cubic splines also does not detect any correlation between the dynamics of volumes of the oil and gas export. Our assumptions about "latent trends" are also confirmed by differentiation of spline models – the correlation between the change rate of the volumes of the oil and gas export is detected. Use of spline functions at economic dynamics modeling is determined with such positive properties of theirs as continuity, flexibility, differentiability, the property of minimal curve, etc.

Keywords: correlation, approximation, dynamics, splines, derivative, rate, export, oil, natural gas

1. Introduction

Modern processes in the economy flow with permanently changing accelerations being exposed to the impact of an unstable set of factors. For analysis of correlations, the classical econometrics (Pinto, 2011) uses the popular correlation method. Depending on the value and the sign of the correlation ratio, the conclusions on presence and proximity of the correlation are made. However, the correlation ratio allows evaluating the interdependence of the studied indexes only in general terms, i.e. on the average for a whole row or a part of the row.

The values of the indexes of the economic process are usually represented in the form of discrete values of a temporal row, i.e. are described with a certain lattice function. However, the lattice, discrete, or table representations of the economic variables do not allow carrying out qualitative analysis of the continuous economic dynamics for a number of reasons, namely:

- there is no term "derivative", representing the trends of an economic process;
- the conceptual absence of interpolative "smoothing" of the lattice, i.e. the impossibility to find the values of the function within the intervals between the nodes;
- the impossibility to forecast the behavior of the economic indexes through extrapolation of the continuous analytical model;
- The impossibility to determine the phase relations in each economic process, build phase patterns, find phase parametric interdependencies between the behavior of certain indexes, etc.

For the analysis of the continuous inertial dynamics of the economic process, the necessity rises to build such a

model, which could detect unambiguously interpreted analytical and quantitative correlations, allow deducing formal consequences from them, detect the properties, construct laws, and forecast the development of the process. Such opportunities for a researcher are found at modeling economic dynamics with splines (Wang and Yu, 2014; Oh, Kim and Hong, 2008; Sekigawa, 2011).

Splines are piecewise-polynomial functions determined on an interval $[a, b]$ and having a finite number of continuous derivatives within this interval. The word “spline” means a flexible ruler used for drawing smooth curves through the given plots on a plane.

Spline appeared in the approximation theory as auxiliary means for proving the Weierstrass theorem and direct theorems of the theory of polynomial approximations as far back as in the times of Lebesgue and Jackson. Spline-based approximations (Schoenberg, 1968) also naturally occur at researching quadrature formulas. Spline approximation (Schoenberg, 1970) was also systematically studied and actively propagated by Schoenberg (Schoenberg, 1979) in the problems of numerical mathematics (Schoenberg, 1983).

2. Materials and Methods

Use of spline models can efficiently complement the correlative approach to the correlations analysis. Comparison of the first two derivative cubic splines approximating the studied dynamics can detect the “latent trends” of the economic dynamics.

Mathematical splines are a set of correlated “pieces” of power polynomials, which represent a process in intervals between nodes. Splines consist of intervals of a small order polynomial, which converge in the given nodes of the process (the nodes of its lattice function). A mathematical spline of q^{th} order is continuous and has a $(q - 1)$ continuous derivative, q^{th} derivative can be subject to discontinuity in the junction plots (lattice nodes) with the finite jump (Ahlberg, Nilson and Walsh, 1972). The pieces of a spline converge in the nodes in an optimal way, so that the values of the function and all its derivatives to the left and to the right would coincide, i.e. $f^{(k)}(x_i + 0) = f^{(k)}(x_i - 0)$, $k = 0, 1, 2 \dots q$, $i = 1 \dots n$. Then $f(x), f'(x), f''(x), \dots$ become continuous functions throughout the whole interval $[x_1, \dots, x_n]$.

At modeling continuous economic dynamics, the model accurately goes through the discrete nodes of the process; thus, an approximating polynomial with further interpolation of the function values at any plots between the nodes is obtained.

The following sequence of actions is suggested in the economics as the method of analysis of correlations by differentiation of spline models (Ilyasov, 2009):

- 1) Economic processes are analyzed for presence of correlation dependence;
- 2) Mathematical models of the economic dynamics are built by approximation with cubic splines (Ilyasov, 2008);
- 3) By differentiation of the obtained spline models, we obtain the models of the growth rate of the studied dynamics;
- 4) By visualization of the derivative spline models (Ilyasov, 2013) of the studied dynamics, we make conclusions about the presence of correlation and about the correspondence of the results of the analysis to the calculated correlation ratios.

3. Results

As the empirical base of the research, we take the economic indexes of the oil and gas sphere. At that, in one case we are going to consider the situation with strong correlative dependence, which is close to functional. In the other case, we will take two rows of dynamics, for which the correlation ratio is close to zero.

For example, multiple researches evidence the presence of strong correlation between export prices for natural gas and oil. Having analyzed the data on the average export prices for crude oil and natural gas in 2000-2012, we found that the correlation ratio between the said pair of prices is equal to $r = 0.982301$. The obtained value of the correlation ratio proves strong correlation between average export prices for crude oil and natural gas, which is close to functional.

The next step is to build mathematical models of dynamics of export prices for oil and gas. We are going to carry out the modeling by approximation with cubic splines.

Table 1. Export of crude oil and natural gas by the Russian Federation during 2000-2012

Years	Crude oil		Natural gas	
	Quantity, MT	Average export price, USD/BL	Quantity, BCM	Average export price, USD/TCM
2000	144.4	23.94	193.9	85.84
2001	164.5	20.78	180.9	98.25
2002	189.5	21.02	185.5	85.69
2003	228.0	23.81	189.4	105.51
2004	260.3	31.02	200.4	109.05
2005	252.5	45.21	209.2	151.36
2006	248.4	56.32	202.8	216.00
2007	258.6	64.28	191.9	233.66
2008	243.1	90.68	195.4	353.69
2009	247.5	55.61	168.4	249.27
2010	250.7	74.11	177.8	268.48
2011	244.5	101.74	189.7	338.88
2012	240.0	103.14	178.7	348.33

The main property of splines, which allows proposing this approach for analysis of correlations, is their continuous nature and differentiability. By differentiation of spline functions, we obtain a more sensitive tool for analysis – the derivatives of the spline functions. Involvement of derivatives in the process of modeling trends of an economic process is more preferable rather than table (lattice) representation of data and use of such concepts only as finite differences, gains, growth rates, etc. The necessity of such an approach is obvious, as economic processes are indeed continuous and inertial.

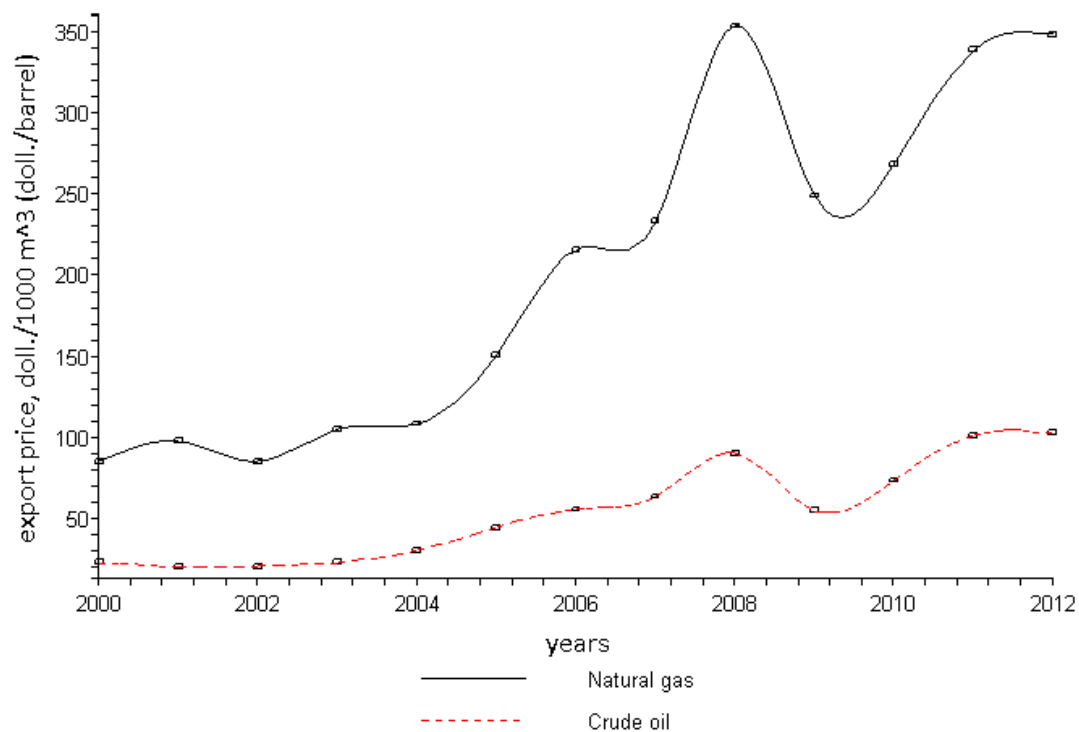


Figure 1. The dynamics of average export prices for natural gas and crude oil between 2000 and 2012. Approximation with cubic splines

Comparison of the curves of dynamics of the export prices for natural gas and oil in Figure 1 represented by cubic splines shows that the prices for gas and the prices for oil demonstrate synchronous fluctuations, which evidences existence of stable correlation between the specified pair of prices. The previously found correlation ratio also confirms the high degree of association closeness.

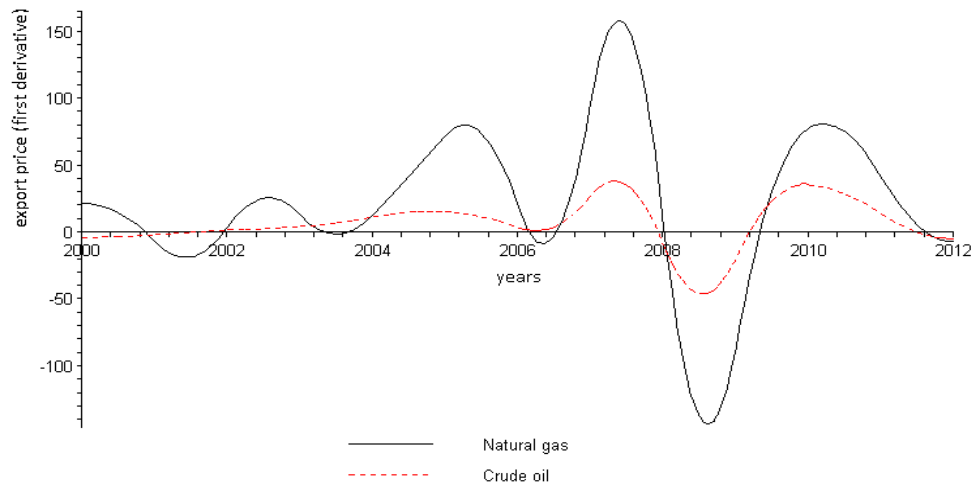


Figure 2. First derivatives of cubic splines approximating the trends of export prices for crude oil (the broken line) and natural gas (the full line). The data on prices are between 2000 and 2012

The next move is to compare the behavior of the first derivatives of the models representing the dynamics of export prices for natural gas and crude oil. The first derivative has rather logical economic content and meaning. It is a continuous trend of the process and indicates the rate of change of the studied index.

The dynamics of the price growth rate, as it is seen in Figure 2, also demonstrates synchronous and proportional amplitude of fluctuations. In this case, we do not see any considerable differences in the results of the correlation analysis based on both the correlation method and the method of comparison of the growth rate of the studied pair of prices – a strong positive correlation is observed.

More interesting situation is when results of the correlation analysis do not match the conclusions made as a result of comparison of the first two derivative spline models. As an example, let us consider the dynamics of the volumes of the natural gas and crude oil export from Russia during 2000-2012.

In this case, the correlation ratio equals to $r = 0.156027$. With such a value of the correlation ratio, econometrics denies existence of any correlation between the studied dynamics rows.

The data in Table 1 demonstrate the complex nature of the studied dynamics – there are intervals here with both oppositely directed dynamics and synchronization of the growth directions. In this case, we cannot identify any regularities in the fluctuations of the studied dynamics using econometric methods.

We will build the models of the dynamics of volumes of the oil and gas export by approximation with cubic splines. Then, we will carry out transition to the models of their growth rate by differentiation of the spline models, which will allow detecting the “latent trends”. Figure 3 demonstrates the asynchronous behavior of the fluctuations of the export volumes growth rate – as the volumes of oil export decrease, the volumes of gas export grow. In this case, two assumptions can be made:

- 1) Due to the interchangeability of oil and gas, decrease in the export of one of the energy sources is compensated with proportional increase in the volumes of export of the other one.
- 2) A chart of the first derivative spline model of the crude oil export volumes dynamics can be built by shifting the chart of the first derivative spline model of the natural gas export volumes dynamics one year backwards.

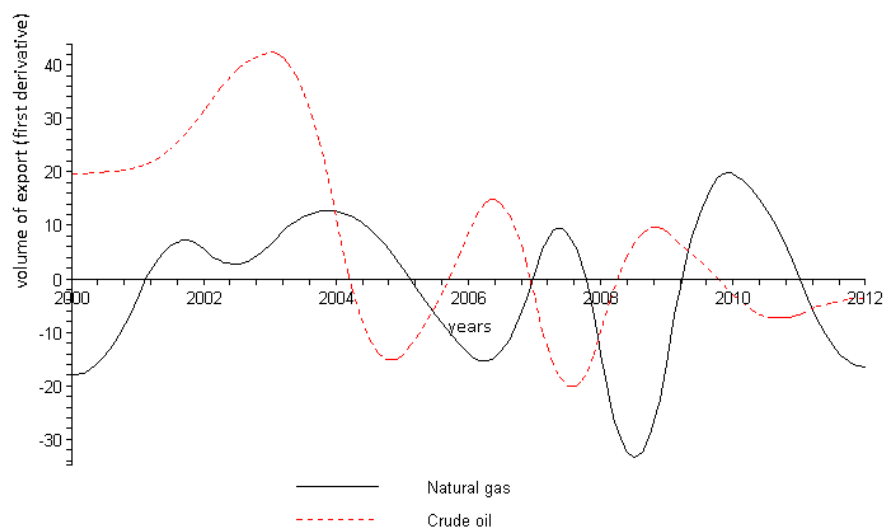


Figure 3. The first derivatives of cubic splines approximating trends of volumes of the crude oil and natural gas export from 2000 until 2012

In order to substantiate the first assumption, we will find the correlation ratio between the values of the first derivative spline models of the dynamics of volumes of the crude oil and natural gas export using the data of Table 2. The obtained result $r = 0.118777$ does not confirm the first assumption.

In order to verify the second assumption, we will find the correlation ratio between the values of the first derivative spline models of the dynamics of volumes of the crude oil and natural gas export shifted one year backwards. The obtained value $r = 0.734616$ indicates quite close correlation between the volumes of the oil and natural gas export. Thus, the dynamics of the volumes of the natural gas export by the Russian Federation reacts to the decrease in volumes of crude oil export by proportional decrease with a delay of up to one year.

Probably, the revealed correlation could have been detected at calculation of the correlation ratio between the original values of dynamics of the volumes of the crude oil and natural gas export when shifting the latter one year backwards. However, in this case, we only obtain the value $r = 0.286268882$. It is obvious that spline approximation and transition by differentiation to the models of the dynamics growth rate were efficient.

Table 2. Export of crude oil and natural gas by the Russian Federation during 2000-2012

Years	Crude oil		Natural gas	
	Quantity, MT	Values of the first derivatives of spline models	Quantity, BCM	Values of the first derivatives of spline models
2000	144.4	19.67	193.9	-17.89
2001	164.5	20.98	180.9	-2.91
2002	189.5	31.99	185.5	5.40
2003	228.0	42.34	189.4	6.76
2004	260.3	10.66	200.4	12.64
2005	252.5	-13.87	209.2	1.91
2006	248.4	9.36	202.8	-14.00
2007	258.6	-4.42	191.9	1.94
2008	243.1	-9.09	195.4	-15.12
2009	247.5	8.69	168.4	-13.76
2010	250.7	-2.97	177.8	19.56
2011	244.5	-6.37	189.7	-0.44
2012	240.0	-3.53	178.7	-16.43

4. Discussion

The modern approaches to the analysis of economic dynamics (Lemmens, Croux and Stremersch, 2012) require applying efficient tools, which would not be limited to detecting common regularities and correlations and would provide the researcher with the opportunity to detect also the "latent trends". This would allow to solve problems related to usage of the analysis results at planning and forecasting more efficiently, thus minimizing alternative losses of market players. One of such tools is the spline approximation approach (Zhanlav and Mijiddorj, 2010) to mathematical modeling of dynamics (Sarfraz, Zawwar and Nisar, 2010).

Among all the spline functions, the cubic splines were preferred, as they have such an attractive optimization property as internal optimality (Korn and Korn, 1973). This property of cubic splines is called the property of minimal curve or of minimal standard. For cubic spline functions $S_{\Delta}(y,x)$, it is expressed by the Holiday's theorem, which shows that spline buildup minimizes the integral:

$$\int_a^b |f''(x)|^2 dx \rightarrow \min$$

Approximation with splines (Valenzuela and Pasadas, 2011; Kano, Fujioka and Martin, 2011) allowed obtaining qualitative models without the typical of econometrics procedures of smoothing and averaging the values of original dynamics (Aatola, Ollikainen and Toppinen, 2013; Benedictow, Fjaertoft and Lofsnes, 2013).

The special role of derivatives in mathematics and physics allows finding their efficient application also in economics at analysis, search for correlations, and forecast of economic behavior. The economic rationale of the first derivative is that it represents the trend or the rate of growth of an economic index; the second derivatives evidence changes in the growth rate, i.e. acceleration or slowdown of the index growth. Involvement of derivatives in the process of modeling the economic conjuncture allows seeing its trends much sooner and managing them, especially if we take into account the continuity and inertia of economic behavior. For qualitative description of processes, it is important to involve continuous analytical models, which have the properties of differentiability, smoothness, flatness, etc.

5. Conclusion

The method of correlations analysis suggested by the authors has a number of deficiencies, which can be eliminated in the long view:

- 1) There is no mechanism of correlation of values of the correlation ratios calculated for the derivative spline models with the correlation ratios for the original rows of dynamics;
- 2) The complexity of the method for an economist who does not have sufficient mathematical skills;
- 3) The greater labor input of the method if compared to the standard correlation analysis.

Despite the specified and other probable deficiencies, the authors believe that the value of the method resides in the sensitivity of this approach to local changes in the economic dynamics. It is also important for the modern economy to identify regularities in the fluctuations of the growth rate, and probably in the accelerations. The parameters of the interdependencies of economic indexes can vary greatly inside even a short interval; they cannot be qualitatively described with a single polynomial regression. As a response to it, spline representation of a parametric pattern of the correlation also demonstrates local reaction to the fluctuations of the factor attribute against the background of the general trend.

The continuity and differentiability of spline functions along with the open dynamic pattern also identify the "latent trends" of the process through the growth rate (the first derivative) and growth acceleration (the second derivative). The analysis of extremes of the first and second derivatives will be able to detect both event-related elements and small crises in the original dynamics.

References

- Aatola, P., Ollikainen, M., & Toppinen, A. (2013). Price determination in the EU ETS market: Theory and econometric analysis with market fundamentals. *Energy Economics*, 36, 380-395. <http://dx.doi.org/10.1016/j.eneco.2012.09.009>
- Ahlberg, J., Nilson, J., & Walsh, L. (1972). *The theory of splines and their applications*. Academic Press, p. 318.
- Benedictow, A., Fjaertoft, D., & Lofsnes, O. (2013). Oil dependency of the Russian economy: An econometric

- analysis. *Economic Modelling*, 32, 400-428. <http://dx.doi.org/10.1016/j.econmod.2013.02.016>
- Ilyasov R. (2013). Phase analysis of the cyclical share of natural gas in the structure of export of the Russian Federation. *Polysubject Network Electronic Scientific Journal of Kuban State Agrarian University*, 4, 88. Retrieved March 13, 2014, from http://www.elibrary.ru/author_items.asp?authorid=537030
- Ilyasov, R. (2008). Spline-analysis “thin” structures of interdependence of the export prices for natural gas and oil. *Scientific-technical Bulletin of SPbSPU. Economic Science*, 6(68), 348-352.
- Ilyasov, R. (2009). Phase spline-analysis as a method of diagnostics of cyclicity in the economy. *Modern High Technologies*, 1, 32-36.
- Kano, H., Fujioka, H., & Martin, C. (2011). Optimal smoothing and interpolating splines with constraints. *Applied Mathematics and Computation*, 5, 1831-1844. <http://dx.doi.org/10.1016/j.amc.2011.06.067>
- Korn, G., & Korn, T. (1973). *Mathematical Handbook for Scientists and engineers ditch*. Moscow: Nauka. p. 832.
- Lemmens, A., Croux, Ch., & Stremersch, S. (2012). Dynamics in the international market segmentation of new product growth. *International Journal of Research in Marketing*, 1, 81-92. <http://dx.doi.org/10.1016/j.ijresmar.2011.06.003>.
- Oh, H., Kim, J. G., & Hong, W. (2008). The piecewise polynomial partition of unity functions for the generalized finite element methods. *Computer Methods in Applied Mechanics and Engineering*, 197(45-48), 3702-3711. <http://dx.doi.org/10.1016/j.cma.2008.02.035>
- Pinto, H. (2011). The role of econometrics in economic science: An essay about the monopolization of economic methodology by econometric methods. *The Journal of Socio-Economics*, 40(4), 436-443. <http://dx.doi.org/10.1016/j.socec.2011.04.011>
- Sarfraz, M., Zawwar, M., & Nisar, A. (2010). Positive data modeling using spline function. *Applied Mathematics and Computation*, 7, 2036-2049.
- Schoenberg, I. J. (1968). On the Ahlberg-Nilson extension of spline interpolation: The g-splines and their optimal properties. *Journal of Mathematical Analysis and Applications*, 21(1), 207-231.
- Schoenberg, I. J. (1970). A second look at approximate quadrature formulae and spline interpolation. *Advances in Mathematics*, 4(3), 277-300.
- Schoenberg, I. J. (1979). On Cardinal Spline Smoothing. *North-Holland Mathematics Studies*, 35(4), 383-407.
- Schoenberg, I. J. (1983). A new approach to Euler splines. *Journal of Approximation Theory*, 39(4), 324-337.
- Sekigawa, H. (2011). Computing the nearest polynomial with a zero in a given domain by using piecewise rational functions. *Journal of Symbolic Computation*, 46(12), 1318-1335. <http://dx.doi.org/10.1016/j.jsc.2011.08.012>.
- Valenzuela, O., & Pasadas, M. (2011). Fuzzy data approximation using smoothing cubic splines: Similarity and error analysis. *Applied Mathematical Modelling*, 5, 2122-2144. <http://dx.doi.org/10.1016/j.apm.2010.11.046>.
- Wang, L., & Yu, M. (2014). Favards theorem of piecewise continuous almost periodic functions and its application. *Journal of Mathematical Analysis and Applications*, 413(1), 35-46. <http://dx.doi.org/10.1016/j.jmaa.2013.11.029>
- Zhanlav, T., & Mijiddorj, R. (2010). The local integro cubic splines and their approximation properties. *Applied Mathematics and Computation*, 7, 2215-2219. <http://dx.doi.org/10.1016/j.amc.2014.01.043>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

An Auto-Landing Solution for a Drop Test RLV Demonstrator

Peng Yong-Tao¹, Wang Yue-Ping¹, Wei Wen-Ling¹ & Wang Xiao-Ting¹

¹ Flight Automatic Control Research Institute, Xi'an, Shanxi, China

Correspondence: Wang Xiao-Ting, Flight Automatic Control Research Institute, Xi'an, Shanxi, China. Tel: 86-159-0295-6193. E-mail: lopermanster@gmail.com

Received: July 2, 2014 Accepted: June 7, 2014 Online Published: September 9, 2014

doi:10.5539/mas.v8n5p204

URL: <http://dx.doi.org/10.5539/mas.v8n5p204>

Abstract

Unpowered drop test is very important for reusable launch vehicle (RLV) autoland technology development. One of the challenges is to design an autoland trajectory with enough robustness against uncertainties of drop conditions, aerodynamic characteristic and disturbances from control system and environment. In this paper, a solution including trajectory generation and control design is proposed for a drop test RLV demonstrator. Firstly, the drop test and vertical flight trajectory are introduced. Also, parts of the drop flight, segments of landing trajectory and trajectory design parameters in groups are shown. Secondly, an online trajectory generation method including self-adapted capture segment plan and landing trajectory optimization following UAV auto-landing experience are illustrated in detail by designing groups of parameters. Then, simple but practical gain schedule control laws are presented. Finally, mathematic simulation and analysis based on both RSS and Monte Carlo methods indicate that the solution proposed has shown an acceptable robustness and can provide enough capability for the demonstrator to land safely.

Keywords: Reusable Launch Vehicle (RLV), auto-landing, on-line trajectory plan, RSS analysis, Monte Carlo Analysis

1. Introduction

Unpowered drop test is always used for validating the auto-landing ability of a newly developed RLV. The vehicle will be dropped away from the carrier about 7-8 km high above the ground, and slide along an arranged trajectory until landing on the pointed runway. Without propulsion system on board, energy control of such a low lift-to-drag vehicle is very difficult. That is why the vertical auto-landing trajectory should be carefully designed before the test, to ensure the safety and enhance the robustness and reliability.

Shuttle unpowered auto-landing trajectory design method (Tsikalas, 1982) is worth learning because of the successful development and applications of Shuttle. The nominal trajectory comprises four segments: steep glideslope, circular flare, exponential flare and shallow glideslope. During the steep glideslope segment, energy and trajectory dispersions caused by TAEM flight will be removed by establishing and stabilizing expected velocity. The circular flare employs a circular arc to linearly increase the flight path angle from steep to shallow. The exponential flare is necessary for avoiding acceleration discontinuity during the transition onto the shallow glideslope. And in the shallow glideslope, the vehicle flies with a relatively small vertical velocity until the landing flare sets the craft on the runway.

With the same segment settings, an Auto-landing I-load Program (ALIP) is developed for X-34 (Barton and Tragesser, 1999), which enforces physical constraints such as loads, vertical descent rate, continuity, and smoothness reduces the design problem to a two-point boundary value problem (TBVP) with conditions on the initial and final dynamic pressure. The ALIP is able to decrease design time and add substantial robustness to off nominal conditions.

In this paper, a developed auto-landing solution for an unpowered drop test RLV demonstrator is proposed, including a drop test trajectory design method and classic control laws. In the second and third section, an on-line trajectory generation method based on the ALIP is illustrated, which is improved for better safeties in these two aspects: 1) to increase robustness of guidance against drop conditions, a self-adapted capture segment is added in front of the steep glideslope segment; 2) to ensure the touchdown performance, the position and dynamic pressure of the final plan point of TBVP, which is set to be the starting point of the shallow glideslope segment, is determined according to UAV auto-landing trajectory design experience. In the fourth section, classic

gain scheduling control laws which are simple but focus on engineering are introduced. In the last section, the Root Sum Square (RSS) Method and Monte Carlo Method are used for simulation analysis and evaluation of the proposed solution in multi-disturbance circumstances.

2. Vertical Flight Trajectory Design

The design of vertical auto-landing trajectory of a drop RLV demonstrator is a two-point boundary value problem, with the initial and ending states specified. The design would process the velocity profile depending on the height profile, in order to meet the constrained state of touchdown. With the vehicle pointing directly to the runway, the whole flight comprises 4 parts: free flight, attitude stabilizing, capturing, and landing. The landing segment also comprises 6 segments: Steep Glideslope, Circular Flare, Exponential Decay, Shallow Glideslope, Landing Flare and runway running, as shown in Fig. 1.

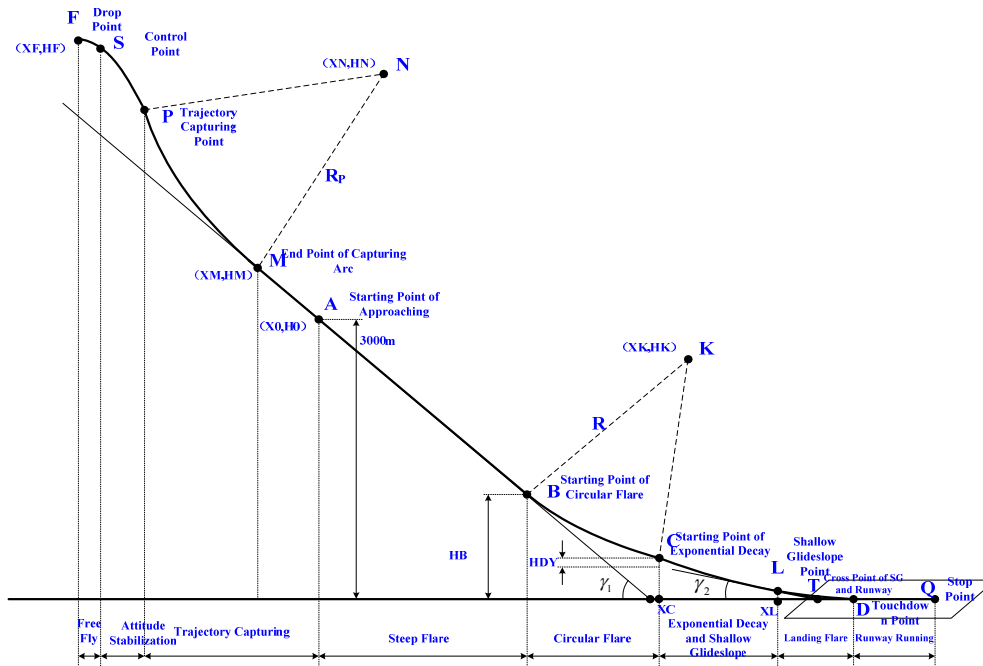


Figure 1. Vertical flight trajectory of drop test with the demonstrator pointing toward the runway

- During the Free Flight Segment, the vehicle will not have any control input, in order to prevent the interference between the carrier aircraft and the demonstrator, ensuring the safety of carrier aircraft;
- During the attitude stabilization segment, the attitude of the demonstrator will be controlled to a prescribed value;
- During the Capturing segment, a trajectory would be generated according to the current position of the demonstrator, the height and horizontal displacement would be controlled, with the reduce of the initial position error;
- During the Steep Glideslope Segment, a trajectory following will be performed, stabilizing the demonstrator flight in the glideslope;
- The segments of Circular Flare and Exponential Decay will control the demonstrator from a large path angle to a smaller one;
- Vertical velocity control will be performed at the final Shallow Glideslope segment, to satisfy the constrained descending rate of touchdown.

Considering the flight states in each segment, the segments contain the need of trajectory design starting from the Capturing Segment, and ending at the Exponential Decay segment. The mathematical representation of the autolanding trajectory is in ground axes, with the prescribed touchdown point D as the origin, using the direction pointing to the farther end of runway as X, and the lateral direction to the right side of the runway as Z, up as Y. The relationship between vertical altitude (h) of the demonstrator and the downrange distance, X, can be

expressed as the following equations.

Circular part of Trajectory Capturing:

$$H = H_N - \sqrt{R_p^2 - (X - X_N)^2} \quad X < X_M \quad (1)$$

Landing Steep glideslope:

$$H = H_A + \tan \gamma_1 (X - X_A) \quad X < X_A \quad (2)$$

Steep glideslope:

$$H = H_B + \tan \gamma_1 (X_B - X) \quad X < X_B \quad (3)$$

Circular flare:

$$H = H_K - \sqrt{R^2 - (X - X_K)^2} \quad X < X_C \quad (4)$$

Exponential Decay and shallow glideslope:

$$H = X \cdot \tan \gamma_2 + H_{DY} \cdot e^{(X_C - X)/\sigma} \quad X < X_L \quad (5)$$

The description of parameters in the equations is presented in Tab. 1, the first 10 parameters define the trajectory geometry of landing segments, and parameters 11-14 define the trajectory of capturing segment. The last parameter indicates the drop point of the demonstrator.

Table 1. Trajectory design parameters

Symbol	Unit	Description
γ_1	rad	Steep glideslope
γ_2	rad	Shallow glideslope
$L(X_L, H_L)$	m	Horizontal coordinate of starting point of Shallow Glideslope segment
R	m	Radius of the arc
$K(X_K, H_K)$	m	The coordinate of origin K
$C(X_C, H_C)$	m	The coordinate of starting point of Exponential Segment
H_{DY}	m	Proportion Factor of Exponential Flare
$B(X_B, H_B)$	m	Coordinate of starting point of the arc
σ	m	Decay rate of the Exponential Flare
$A(X_A, H_A)$	m	Coordinate of the touchdown point
R_p	m	The radius of the arc of Capturing Segment
$N(X_N, H_N)$	m	The coordinate of origin of the arc of Capturing Segment
$M(X_M, H_M)$	m	The coordinate of the finishing point of arc trajectory
$P(X_P, H_P)$	m	The coordinate of the starting point of Capture Segment
$F(X_F, H_F)$	m	Coordinate of the drop point

The parameters in Tab. 1 can be grouped into 4:

Group 1: Steep glideslope, Shallow glideslope, Horizontal coordinate of starting point of Shallow Glideslope segment can be determined considering the aerodynamics of the demonstrator, flight envelop constrains and engineering experience.

Group 2: X coordinate of Starting point of Exponential Decay (X_C), Decay rate of the Exponential Flare (σ) need iteration of optimization based on the restriction.

Group3: Proportion Factor of Exponential Flare H_{DY} , Origin of the circular flare arc K, the radius of the arc R, the altitude of the starting point of circular flare H_B , and the touchdown point A can be determined considering the continuous of trajectory and the restriction of X_C and σ .

Group 4: Coordinate of the starting point of Capturing Segment N, the radius of the arc R_p , the finishing point of arc trajectory M can be generated by the online self-adaption algorithm based on the current position (P) of the demonstrator at the end of the attitude stabilization segment.

3. Parameter Selection

Offline trajectory design needs strict restriction of dropping window, while the online trajectory design reduces its dependence on the pre-designed trajectory, expanding the envelop of initial value constraints, providing the trajectory phase with the maximum robustness, with enhancing the safety and reliability of the demonstrator. So the self-adaption method is used to calculate the capturing segment parameters while the offline method is used in the calculation of parameters in landing segments.

3.1 Group 1

a) γ_1

Supposing the dynamic pressure and the path angle in Steep Glideslope stay the same, selected steep glideslope angle maintain the balance of gravity component and the drag. The motion equations of the demonstrator are given by:

$$\begin{cases} qSC_L = mg\cos\gamma_1 \\ qSC_D = -mg\sin\gamma_1 \end{cases} \quad (6)$$

where q is the dynamic pressure, S is the wing span, C_L is the lift coefficient, C_D is the drag coefficient, m is the demonstrator mass, g the gravity acceleration, and the path angle γ_1 is given by

$$\tan\gamma_1 = -\frac{C_D}{C_L} \quad (7)$$

Obviously, the path angle γ_1 at any altitude is inverse proportion to the lift-to-drag ratio, meaning that the smaller the lift-to-drag ratio is, the larger $|\gamma_1|$ will be, and the bigger the effect on γ_1 of the lift-to-drag ratio uncertainty will be. Thus, the path angle γ_1 better be picked as small as possible from which can meet the needs of the energy and stability requirements.

b) γ_2

Shallow glideslope angle γ_2 is determined by the velocity and descending rate of touchdown, the relationship of touchdown velocity, descending velocity and shallow glideslope angle γ_2 can be expressed as:

$$\dot{H} = V\sin\gamma_2 \quad (8)$$

The smaller $|\gamma_2|$ is, the smaller descending rate will be, and the easier realized the flare will be. Considering the demonstrator is accelerated in this segment, the parameter γ_2 should be picked according to the descending rate of the landing flare starting velocity of common UAV.

c) Horizontal coordinate of starting point of flare

Horizontal coordinate of starting point of flare X_L , can be determined by the prescribed duration of the landing flare segment. The landing flare segment begins at altitude of 15 meters, and ends after 5 seconds, so X_L can be given by:

$$X_L = \int_0^5 V(t) \cdot \cos(\gamma(t)) \cdot dt \quad (9)$$

So, X_L is determined by the velocity and path angle of the landing flare segment.

3.2 Group 2

The iteration procedure of solving parameters of group 2 is shown in Fig. 2. In every iteration step, with the initial altitude H_A and dynamic pressure q_A of the landing trajectory already determined by experience, trajectory geometry parameters, including coordinates of Points A, B, C, K and L, is derived from the guessed abscissa of starting point (X_C) and decay rate (σ) of exponential flare segment. Then a simulation (Barton and Tragesser, 1999) based on the generated trajectory is carried out, and the dynamic pressure in Point L and the normal overload at point C is calculated. If the dynamic pressure in Point L exceeds the constraint, parameter σ will be adjusted for the next iteration. If not, when the normal overload at point C exceeds the constraint, parameter X_C will be adjusted for the next iteration. Once the new parameter σ or X_C adjusted, next iteration step begins.

The adjustment of σ is trying to find a realizable flight trajectory under the constraints of starting point of landing and starting point of flare. The adjustment of X_C is processed in the whole flight trajectory, with little influence of experience factor of designers. If normal overload is added to the consideration of design, the change of normal overload will be smaller, making the trajectory smoother.

3.3 Group 3

After parameters of groups 1 and 2 and the initial altitude of landing segments are determined, H_{DY} , R , γ_4 , and the coordinate of points A, B and K can be determined with derivation of geometry relation.

The proportion factor of the exponential function is given by:

$$H_{DY} = \frac{X_L \cdot \tan \gamma_2}{e^{(X_C - X_L)/\sigma}} \tag{10}$$

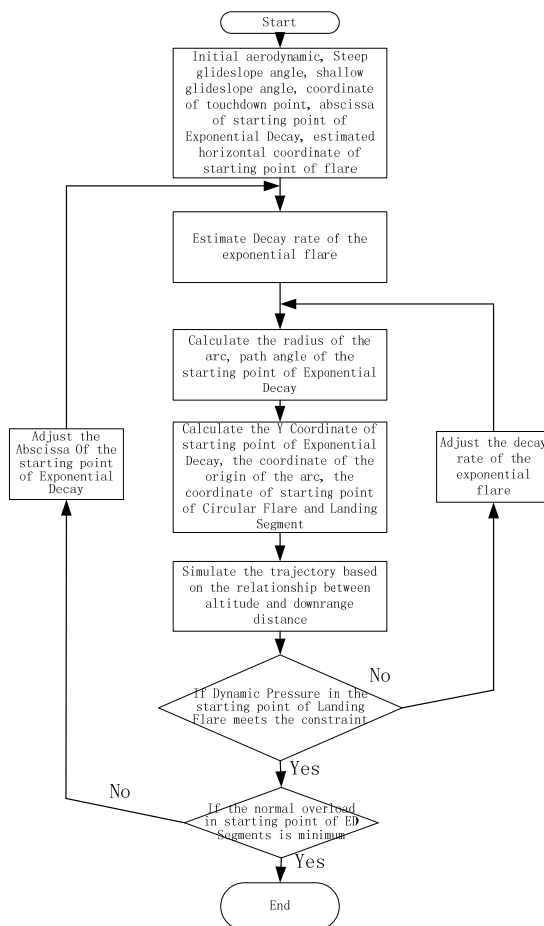


Fig. 2. Flow-chart of solving the altitude phase

The radius R is given by:

$$R = \frac{\sigma^2(1 + (\tan \gamma_2 - \frac{HDY}{\sigma})^2)^{\frac{3}{2}}}{HDY} \quad (11)$$

The path angle at cross point of the Circular Flare and Exponential Decay (γ_4) is given by:

$$\gamma_4 = \tan^{-1}(\tan \gamma_2 - \frac{H_{DY}}{\sigma}) \quad (12)$$

The coordinate of point C is given by:

$$H_C = X_C \cdot \tan \gamma_2 + H_{DY} \quad (13)$$

The coordinate of point B is given by:

$$X_B = X_C + R \cdot (\sin \gamma_1 - \sin \gamma_4) \quad (14)$$

$$H_B = H_C + R \cdot (\cos \gamma_1 - \cos \gamma_4) \quad (15)$$

The coordinate of point K is given by:

$$X_K = X_B - R \cdot \sin \gamma_1 \quad (16)$$

$$H_K = H_B + R \cdot \cos \gamma_1 \quad (17)$$

The coordinate of point A is given by:

$$X_A = X_B + \frac{(H_A - H_B)}{\tan \gamma_1} \quad (18)$$

3.4 Group 4

After the parameters of landing segment determined as above and the altitude of point M picked based on experience, coordinates of point M and N and the radius R_p can be calculated online with self-adapted algorithm based on demonstrator position, once its orientation is stabilized and the velocity is below constraint in capturing segment.

a) The coordinate of point M is given by:

$$X_M = X_A + (H_M - H_A) / \tan \gamma_1 \quad (19)$$

b) The radius R_p is given by:

$$R_p = \frac{\sqrt{(X_p - X_M)^2 + (H_p - H_M)^2}}{\gamma_1 - \tan^{-1}(\frac{H_p - H_M}{X_p - X_M})} \quad (20)$$

c) The coordinate of point N is given by:

$$X_N = X_M - R_p \cdot \sin \gamma_1 \quad (21)$$

$$H_N = H_M + R_p \cdot \cos \gamma_1 \quad (22)$$

4. Gain Scheduling Control

The gain scheduling control laws of the demonstrator are illustrated in Fig. 3 and Fig. 4. In the attitude stabilization segment, the pitch angle is controlled. While in the capturing segment, the altitude and vertical velocity are controlled, with the altitude command calculated according to equation 1 to 5 and the vertical velocity command generated following changes of the altitude command.

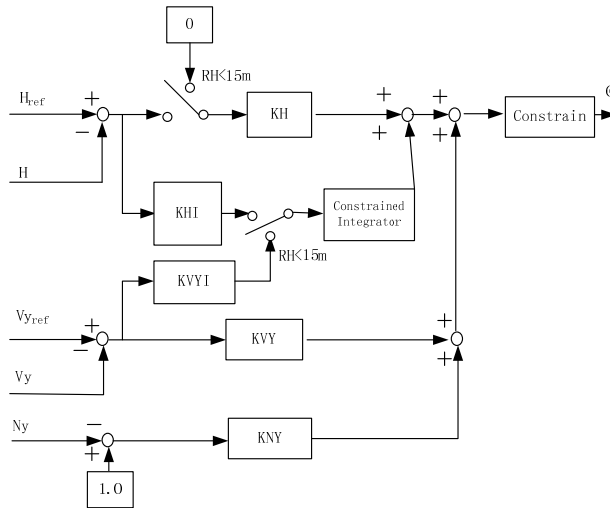


Fig. 3. Vertical Controller

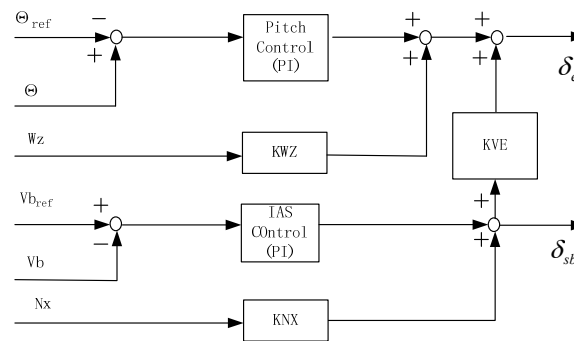


Fig. 4. Pitch Controller

The demonstrator use IAS control (PI) in the velocity control loop, and the command is predetermined offline. The IAS is achieved by using the speedbrake, which will generate an additional moment to the demonstrator. Thus, a cross-link of speedbrake command (δ_{sbcmd}) and elevator command (δ_{ecmd}) is introduced to the pitch control law. Besides, the pitch angle rate, the normal overload and the axial overload are also introduced, to enhance the damping of the system.

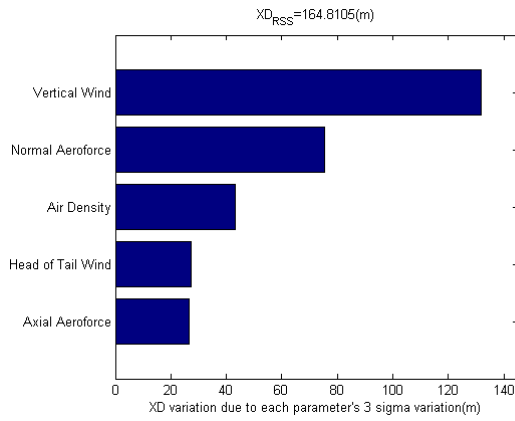
5. Simulation and Analysis

In order to verify the auto-landing solution proposed in this paper, two methods are applied to conduct flight simulations and performance evaluations based on the demonstrator. One is RSS analysis method, which is used to find primary factors that can exacerbate the flight performance and the right direction to improve the solution, by evaluating how different uncertain factors affects the flight performance. The other is Monte Carlo method, in order to evaluate the robustness of the solution by finding out flight performance boundaries with existence of multiple uncertain factors.

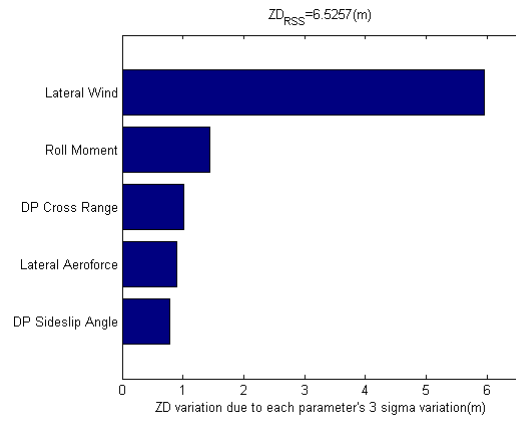
Table 2. Flight Performance and the RSS analysis results

Type	Name	Symbol	Unit	Min	Max	Normal Value	RSS
	Down Range departure	XD	m	200	200	0	164.81
	Cross Range departure	ZD	m	-20	20	0	6.53
	Slapdown	Sink Rate	VyD	m/s	-3	0	-0.8
Performance	Ground Velocity	VD	m/s	0	97.22	85.79	10.12
	Roll Attitude	GamaD	deg	-5	5	0.02	3.17
	Pitch Attitude	ThetaD	deg	0	10	7	0.74
Rollout Performance	Running Distance	XH	m	0	1300	633.17	265.91
	Maximum Lateral Taxiing Deviation	ZH	m	-20	20	0	21.25

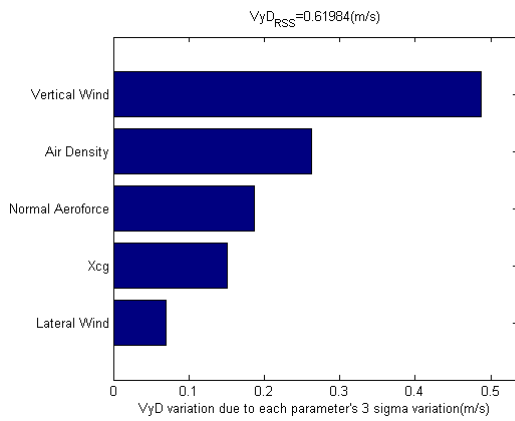
Inflight Performance	Max Dynamic Pressure	QU	pa	0	13000	11175.46	803.49
	Max Normal Overload	NyU	g	0	2.5	1.59	0.62
	Max Angle of Attack	AlphaU	deg	-5	18	8.47	4.26
	Max Angle of Sideslip	BetaU	deg	-5	5	0.46	1.82



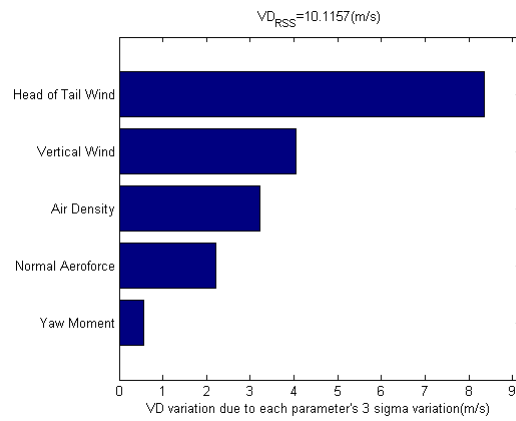
(a)



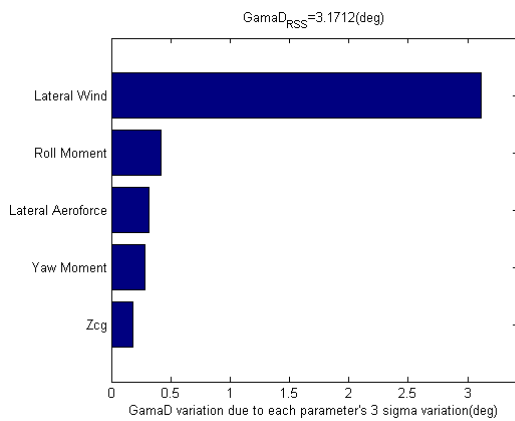
(b)



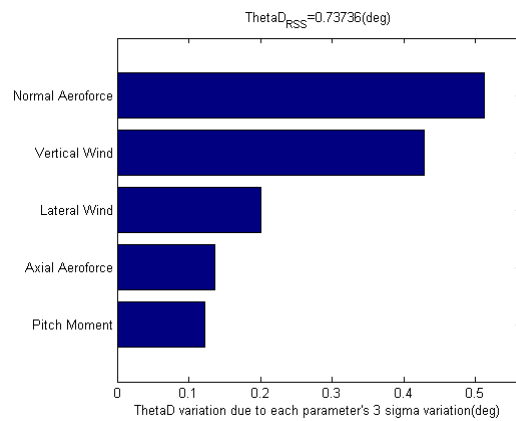
(c)



(d)



(e)



(f)

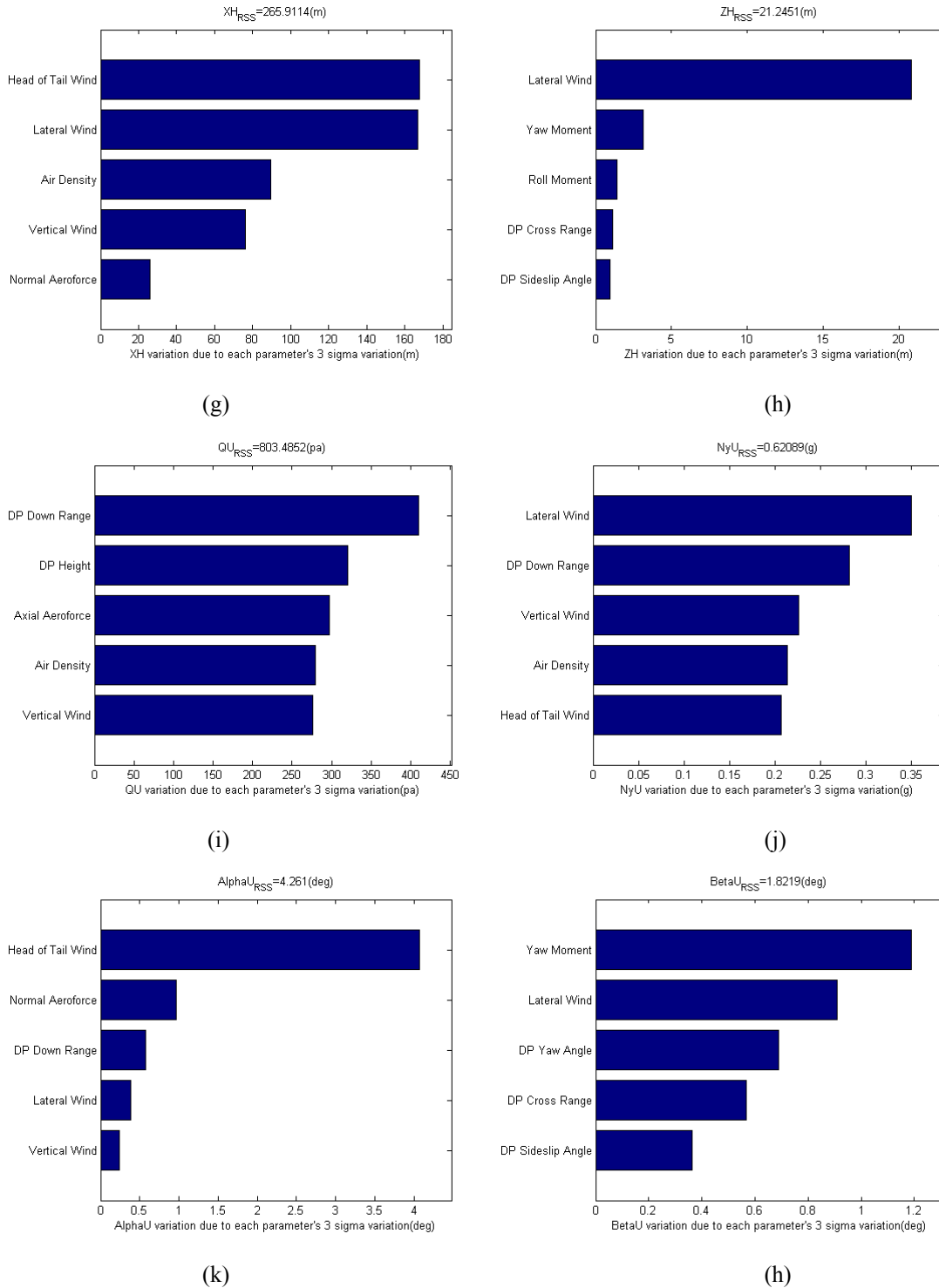


Fig. 5. Flight performance RSS value and the first five disturbances

In RSS analysis, we primarily researched on the influences of 37 uncertain factors belonging to four categories, which are flight quality characteristics, aero-dynamic parameters, initial drop conditions, and turbulence; we also did our research on 12 flight performance items belonging to three categories, which are touchdown performance, taxiing performance, and in-flight performance. The definitions, units, expected boundary values, standard values and RSS computation results of all these performances are listed in Tab. 2. Fig. 5 shows the RSS analysis results of each performance item as well as the 5 most influential ones. We can see from the results that the velocity at touchdown point and the maximum lateral taxiing deviation are close to or even have exceeded the

expected performance boundary, the main influential factor of which is the wind field. Besides, the other items of flight performance are acceptable after being disturbed. Therefore, this design still needs to be improved. If not, more rigorous wind field condition of the unpowered drop test should be proposed for the safety of the autoland; if improved, the anti-wind capability of the aircraft would be increased.

As for the Monte Carlo method, we added sensor measurement errors in 10000 times of simulation. Fig.6 shows the statistical results of the primary flight performance parameters. Star signs are used to mark the expected performance boundary. We can see from the results that the average of the touchdown points, 'x' marks, is slightly to the right of the standard trajectory. The percentage of the touchdown points that are inside the expected area is 97.2%. The other flight parameters in addition to that can basically satisfy the flight performance boundary. Fig. 7 shows the vertical and lateral flight trajectories, which can apparently converge to the standard trajectory with the existence of multiple uncertain factors. This indicates that the solution has shown an acceptable robustness and can provide enough capability for the demonstrator to land softly.

Results of the aforementioned two simulations have shown that this autoland solution is feasible and acceptable. And if the anti-wind capability can be increased, the flight performance would be better improved.

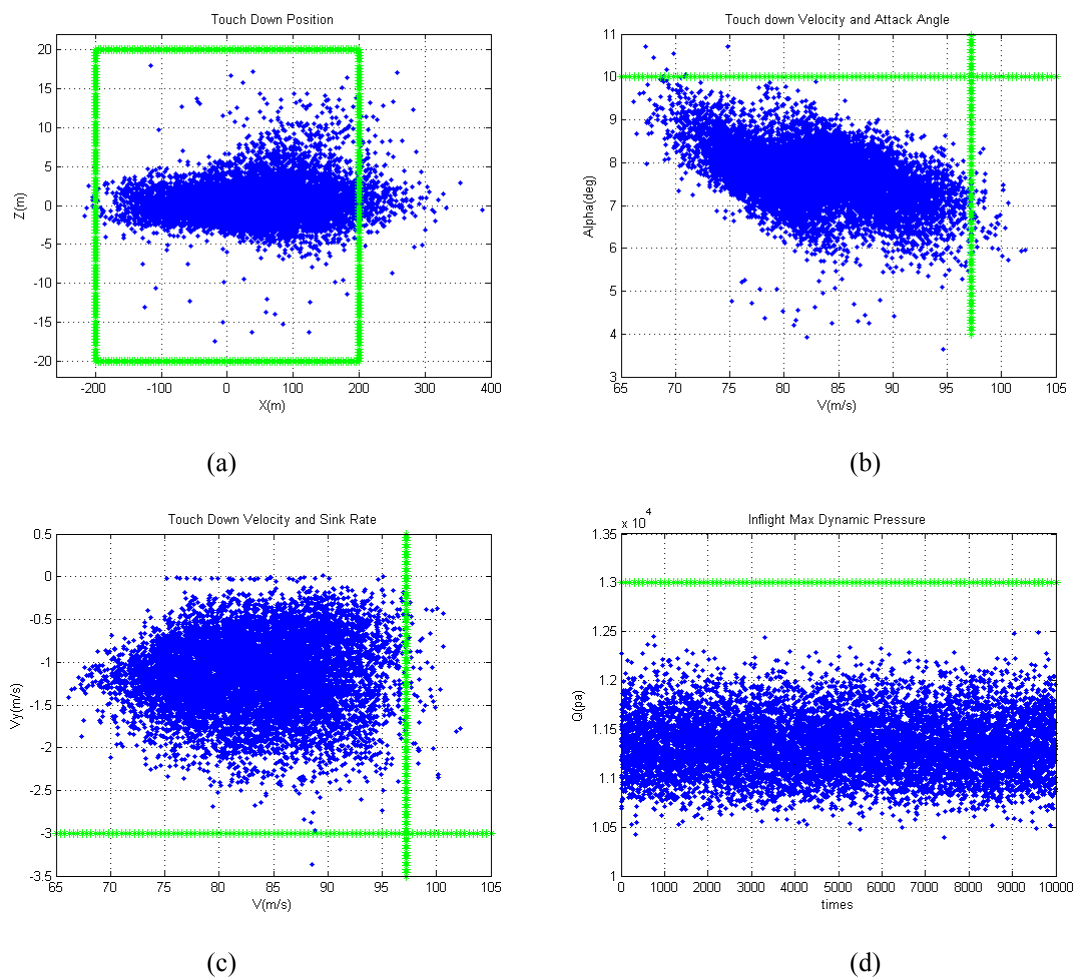


Fig. 6. Primary flight performance boundaries and Monte Carlo simulation results

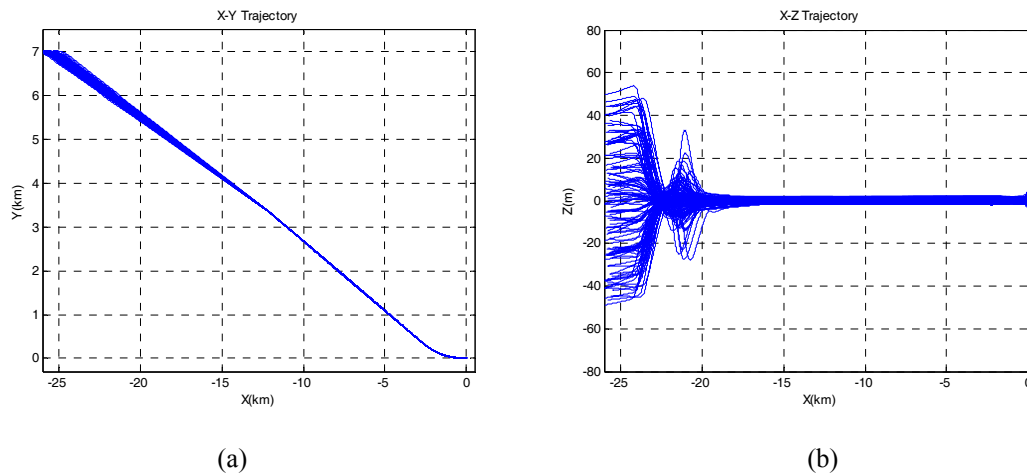


Fig. 7. Autolanding trajectories results

6. Conclusions

To ensure a safe auto-landing of a unpowered drop test vehicle, a new auto-landing solution is proposed including an on-line trajectory generation method and gain scheduling control laws.

In the generation of autolanding segments, a self-adapted capture segment is considered, an iteration method optimizing the abscissa of starting point of flare segment and decay rate of exponential flare is introduced to reduce effects of human decision, and the end of shallow steep is set to be the 15m high to improve touch down performance, all of which lead to a more robust trajectory generation process. Besides, the classic gain scheduling control laws are feasible and reliable with simple but useful control mode and logic.

RSS and Monte Carlo Methods are used for simulation analysis and evaluation of the proposed solution in multi-disturbance circumstances. Analysis results prove that the autoland solution is effective and robust to generate a smoothed and energy controllable trajectory for a safe autolanding with multiple uncertain factors.

References

- Ali Heydari, Sivasubramanya Balakrishnan. (2011). Optimal Online Path Planning for Approach and Landing Guidance. *AIAA Atmospheric Flight Mechanics Conference, Portland, Oregon*. <http://dx.doi.org/10.2514/6.2011-6641>
- Ander R. G. (2001). *Onboard Trajectory Generation for Unpowered Landing of Autonomous Reusable Launch Vehicle*. MS Thesis, MIT.
- C. A. K. (2004). Unpowered approach and landing guidance using trajectory planning. *Journal of Guidance, Control, and Dynamics*, 27, 967-974. <http://dx.doi.org/10.2514/1.7877>
- C. A. K. (2007). Unpowered Approach and Landing Guidance with Normal Acceleration Limitations. *Journal of Guidance, Control, and Dynamics*, 30, 882-885. <http://dx.doi.org/10.2514/1.28081>
- Christina T. C. (1998). Design of a robust integrated guidance and control algorithm for the landing of an autonomous reusable launch vehicle. MS Thesis, MIT.
- Gregg, H., Barton, S., & Tragesser, G. (1999). Autolanding Trajectory Design for the X-34. *AIAA atmospheric flight conference and exhibit, Portland, Oregon*. <http://dx.doi.org/10.2514/6.1999-4161>
- Gust, M. T. (1982). Space shuttle Autoland design. *AIAA Guidance and Control Conference San Diego, California*. <http://dx.doi.org/10.2514/6.1982-1604>
- Schierman, J. D., & Ward, D. G. (2001). A reconfigurable guidance approach for reusable launch Vehicle. *AIAA Guidance, Navigation, and Control Conference and Exhibit, Montreal, Canada*. <http://dx.doi.org/10.2514/6.2001-4429>
- Masaaki Y., Yoshikazu M., & Hirofumi T. (1999). Simulation analysis of the HOPE-X demonstrator. *International Space Planes and Hypersonic Systems and Technologies Conference, Norfolk, VA, U.S.A.* <http://dx.doi.org/10.2514/6.1999-4875>

- Miyazawa, Y., & Motoda, T. (1999). Longitudinal Landing Control Law for an Autonomous Reentry Vehicle. *Journal of guidance, control, and dynamics*, 22(6), 791-800. <http://dx.doi.org/10.2514/2.4480>
- Motoda, T., & Miyazawa, Y. (1998). ALFLEX flight simulation analysis and flight testing. *AIAA Aerospace Sciences Meeting and Exhibit, Reno, NV, U.S.A.* <http://dx.doi.org/10.2514/6.1998-301>
- Schierman, J. D., Hull, J. R., & Ward, D. G. (2002). Adaptive guidance with trajectory reshaping for reusable launch vehicle. *AIAA Guidance, Navigation, and Control Conference and Exhibit, Montreal, Canada.* <http://dx.doi.org/10.2514/6.2002-4458>
- Shen, H. (2004). Autolanding Trajectory Design of Space Shuttle. *Flight Dynamics*, 22(1), 4-7. (in Chinese).
- Sun, C. (2008). *Research on Terminal Area Energy Management and Autolanding Technology for Reusable Launch Vehicle*. MS Thesis, Nanjing University of Aeronautics and Astronautics (in Chinese).
- Sun, C., & Huang, Y. (2008). Online Trajectory Generation of Shuttle. *Missile and Guidance*, 28(4), 261-265. (in Chinese).
- Toshikazu, M., Yoshikazu, M., Kazutoshi, I., & Tatsushi, I. (1999). Automatic landing flight experiment flight simulation analysis and flight testing. *Journal of Spacecraft and Rockets*, 36, 554-560. <http://dx.doi.org/10.2514/3.27199>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

Constants of Metal Rubber Material

Alexander Mikhailovich Ulanov¹

¹ Samara State Aerospace University (SSAU), 34, Moskovskoye shosse, Samara, 443086, Russia

Correspondence: Alexander Mikhailovich Ulanov, Samara State Aerospace University (SSAU), 34, Moskovskoye shosse, Samara, 443086, Russia. E-mail: alexulanov@mail.ru

Received: July 3, 2014 Accepted: August 4, 2014 Online Published: September 26, 2014

doi:10.5539/mas.v8n5p216

URL: <http://dx.doi.org/10.5539/mas.v8n5p216>

Abstract

Damping materials made of pressing wire, such as Metal Rubber (MR material), “metal-flex”, “spring cushion” etc. are used widely in vibration protection systems. They have high strength and damping, however they are non-linear and anisotropic. To use contemporary finite element software for calculation of vibration insulators made of wire damping material one should know constants of this material. Till this time this problem almost isn't researched, there is only a few data in linear approximation. A Young modulus for pressure and bending, shear modulus, Poisson ratio, friction force for pressure and for shear, friction coefficient between wire material and steel plate are obtained for MR material made of stainless steel wire for different load directions by static experiment in the present research. An influence of deformation, relative density, wire diameter, pressing force, deformation in other direction on these constants is considered. Peculiarities of pressing wire material deformation process are discussed: a difference of Young modulus for pressure and bending, a stabilization of hysteretic loop for one direction during load in another direction, an energy dissipation coefficient for different load directions. Results of the present research not only allow calculation of vibration isolator made of MR material by finite element software, but give a method for research of other pressing wire damping materials.

Keywords: MR material, Young modulus, shear modulus, friction force, Poisson ratio

1. Introduction

MR material (MR – Metal Rubber – is manufactured by cold pressure of wire spiral) are used widely for elastic-damping elements of vibration insulators (Ao et al, 2005; Jiang et al, 2008; Xia et al, 2009). It has high strength, high energy dissipation coefficient, ability to work under high and low temperature, in aggressive media, in vacuum etc. Analogous wire materials are “metal-flex”, “spring cushion”, “wire mesh damper” (Gildas, 1989; Kozian and Schmoll, 1998; Al-Khateeb, 2002). These wire damping materials may be used in combination with squeeze film damper (Jiang et al, 2005). However these materials have large non-linearity and anisotropy (Ao et al, 2003). The only way to calculate elastic-damping element made of wire material by contemporary finite element software is to consider this material as anisotropic continuous media (Ulanov and Ponomarev, 2009) and to know constants of this material for different axis direction and dependencies of these constants on parameters of wire material and load. A present paper continues works (Yan et al, 2010a) and (Yan et al, 2010b) and contains a number of constant which is maximally possible in the present time.

2. Method

To obtain constants of MR material a static load experiment was used in the present research. Elastic-damping element with size 25×25×25 mm was used. To research deformation of shear and tension this element was glued to plane surface. X axis is a direction of pressing during MR elastic-damping element manufacturing. Y and Z axis are perpendicular to direction of pressing. Compression direction is positive.

Parameters of wire material are diameter of wire d_w and relative density of material $\bar{\rho}$. The relative density is ratio of density of manufactured elastic-damping element to density of material of wire (for MR it is stainless steel usually). Pressure of manufacturing of elastic-damping element depends on its relative density:

$$\sigma_{press} \approx 576\bar{\rho}^{1.7} \quad (1)$$

Hysteretic loop of damping elements made of MR material is significantly nonlinear (Fig. 1). To make results more common it is better to use relative coordinates such as stress and relative deformation.

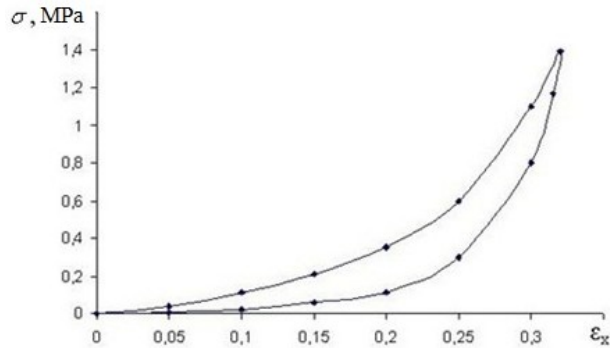


Fig. 1. Hysteretic loop of elastic-damping element used in the present research for X axis direction. $\bar{\rho}=0.18$, $d_w=0.1$ mm

Full stress consists from elastic stress σ_L and friction stress σ_H .

$$\sigma = \sigma_L \pm \sigma_H \tag{2}$$

Here “+” is for load process with stress σ_1 and “-” is for unload process with stress σ_2 . Elastic stress is a middle line of hysteretic loop described by equation

$$\sigma_L = (\sigma_1 + \sigma_2) / 2 \tag{3}$$

Friction stress is difference between load (or unload) process and the middle line, thus

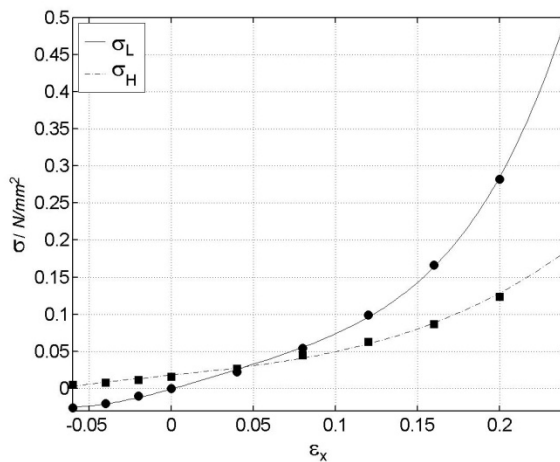
$$\sigma_H = (\sigma_1 - \sigma_2) / 2 \tag{4}$$

Equations (3) and (4) are correct if all wire contacts slip. For the beginning of unloading process it is not so. (Ulanov and Lazutkin, 1997) gives a range for correctness of (3) and (4) as $[0.7A_{min}; 0.7A_{max}]$ (here A_{max} and A_{min} are maximal and minimal amplitude of deformation respectively). For one time loading in experiment A_{max} and A_{min} may be much more than usual working amplitude of vibration isolator, so range $[0.7A_{min}; 0.7A_{max}]$ is sufficient for practice. For X axis a researched range of strain is $\epsilon_x \in [-0.06; 0.24]$, for Y and Z axis $\epsilon \in [-0.06; 0.16]$. Usual working strain of MR material is less (Yan et al, 2013). Range of relative density is $\bar{\rho} \in [0.18; 0.3]$. Range of wire diameter is 0.1...0.25 mm.

3. Results

3.1 Young Modulus and Friction Stress for Pressure and Tension

The elastic stress and friction stress for $\bar{\rho} = 0.18$ and $d_w = 0.1$ mm for X and Y axis direction are presented at Fig 2 (Yan et al, 2010a). Equations (5) – (8) describe dependencies of stress on strain and relative density with an error about 5%.



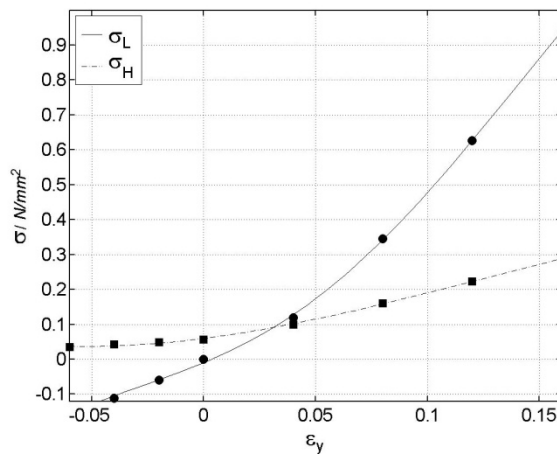


Fig. 2. Elastic stress and friction stress for $\bar{\rho} = 0.18$, $d_w = 0.1$ mm for X and Y axis directions

$$\sigma_{Hx} \approx (0.167 + 2.15\varepsilon_x + 0.641\varepsilon_x^2 + 71.6\varepsilon_x^3)\bar{\rho}^{1.3}, \text{ MPa}; \quad (5)$$

$$\sigma_{Hy} \approx (0.56 + 7.62\varepsilon_y + 60.5\varepsilon_y^2 - 156\varepsilon_y^3)\bar{\rho}^{1.3}, \quad (6)$$

MPa.

$$\sigma_{Lx} \approx (11 + 33\varepsilon_x - 376\varepsilon_x^2 + 2952\varepsilon_x^3)\varepsilon_x\bar{\rho}^{1.7}, \quad (7)$$

MPa;

$$\sigma_{Ly} \approx (49.6 + 262\varepsilon_y + 2560\varepsilon_y^2 - 11660\varepsilon_y^3)\varepsilon_y\bar{\rho}^{1.7} \quad (8)$$

MPa.

First efficient in equations (7) and (8) is analogous to Young modulus E.

Because Z axis is similar to Y axis, it is possible to find stress σ_{Hz} as σ_{Hy} and stress σ_{Lz} as σ_{Ly} .

To obtain dependency of Young modulus on wire diameter the elastic force and friction force for wire diameters 0.15 – 0.25 mm were divided by force for $d_w = 0.1$ mm. These ratios are required coefficients of influence. They are described by functions

$$f_{Hx} = -87.8d_w^2 + 30.8d_w - 1.19 \quad (9)$$

$$f_{Hy} = -27.2d_w^2 + 9.56d_w + 0.322 \quad (10)$$

$$f_{Lx} = -47.1d_w^2 + 16.5d_w - 0.174 \quad (11)$$

$$f_{Ly} = -7.1d_w^2 + 1.90d_w + 0.881 \quad (12)$$

Thus to take into account an influence of wire diameter it is enough to multiply the function (5) by coefficient (9), function (6) by coefficient (10) etc.

3.2 Young Modulus for Bending

To obtain Young modulus for bending a process of deformation of ring-shape elastic-damping element was considered (Xia et al, 2005). For a deformation of ring made of linear material under load P is known (Timoshenko S. and J.N. Goodier, 1951):

$$E_b = 0.149 \frac{P \cdot R^3}{\delta \cdot J_x} = 0.149 C_p \cdot \frac{R^3}{J_x} \quad (13)$$

here δ is deformation, R is radius of ring middle line, J_x is ring section inertia moment, C_p is ring stiffness. For little deformation $C_p \approx (|T_1| + |T_2|) / (|a_1| + |a_2|)$. It is possible to obtain segments T_1, T_2, a_1 and a_2 experimentally by a hysteretic loop (Fig 3).

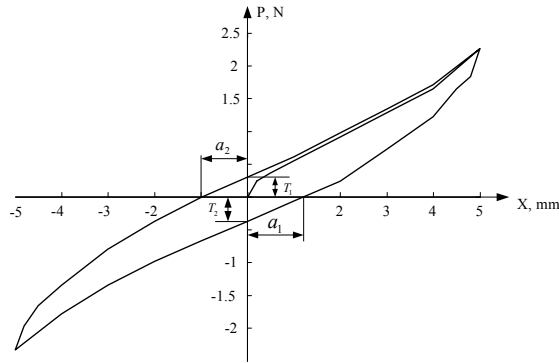


Fig. 3. Obtaining of T_1, T_2, a_1 and a_2 from hysteretic loop

Dependency of Young modulus on $\bar{\rho}$ and d_w is

$$E \approx 10.5d_w(54.1\bar{\rho} - 1.38) \text{ (MPa)} \tag{14}$$

3.3 Shear Modulus and Friction Stress for Shear

In a case of shear the hysteretic loop (if to exclude its ends) is more near to linear (Fig. 4), thus it is possible to use a linear dependency for its description.

$$\tau = G\gamma \pm \tau_H \tag{15}$$

here τ_H is friction stress for shear.

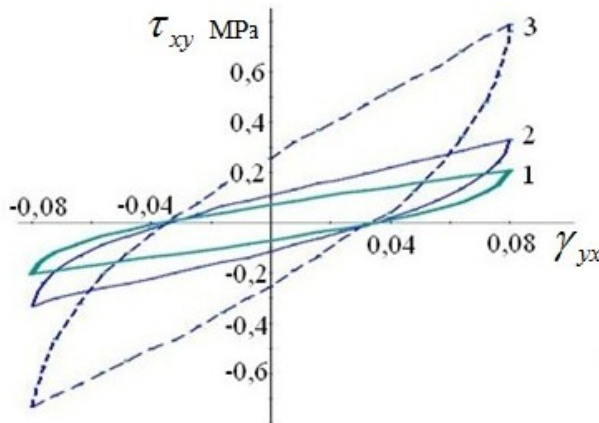


Fig. 4. Hysteretic loop of researched elastic-damping element for shear ($\bar{\rho} = 0.18, d_w = 0.1$ mm) and its dependency of shear in XY axis direction on strain in X axis direction. 1 - $\epsilon_x = 0$; 2 - $\epsilon_x = 0.08$; $\epsilon_x = 0.2$

Dependencies on relative density (for $d_w = 0.1$ mm and the range of deformation $\gamma \in [-0.12; 0.12]$, which excludes ends of the loop) for shear direction XY are

$$G_{yx} \approx 6.65\bar{\rho}^{1.7} \text{ MPa}, \tau_{Hxy} \approx 0.151\bar{\rho}^{1.3} \text{ MPa}. \tag{16}$$

For shear direction YZ

$$G_{yz} \approx 18.3\bar{\rho}^{1.7}, \text{ MPa}, \tau_{Hyz} \approx 0.301\bar{\rho}^{1.3}, \text{ MPa}. \tag{17}$$

Shear modulus and shear friction stress are independent on wire diameter. It is shown in (Yan et al, 2010b) that it is possible if during shear most of wires in MR material work on shear but not on bending.

3.4 Dependency of Constants on Load in Other Direction

Pressure in one axis direction increases force in each wire contact. Of this reason a stiffness and friction should increase in other axis direction. Dependency of shear in XY axis direction on strain in X axis direction is presented on Fig. 4.

In this case if strain $\varepsilon_x > 0$, it is necessary to multiply shear modulus G_{yx} and shear friction stress τ_{Hyx} on $(1 + 22\varepsilon_x^{1.4})$.

Influence of pressure in X axis direction on elastic and friction properties in Y direction is researched for analogous wire material in (Choudhry, 2004).

Because shear changes force in wire contacts insignificantly, there is no influence of shear deformation in Y direction on friction stress in X direction. An elastic stress in this case should be multiplied on

$$(1 + 1.25\gamma_{yx}).$$

3.5 Poisson Ratio

For load in X axis direction the Poisson ratio is $\mu_{yx} = \mu_{zx} \leq 0.03$.

Usually it isn't necessary to take into account so little value, thus it is possible to assume that element made of MR material during its deformation in manufacturing pressing direction has no any pressure on its sides.

If a deformation is in Y direction, one should differ two cases: initial load and multi-time deformation. For initial load there are hysteretic loops in X and Z directions too (Fig. 5).

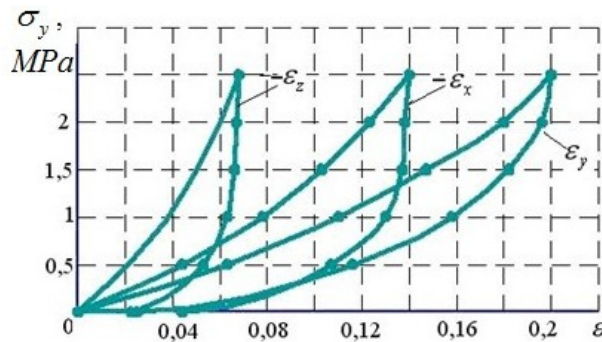


Fig. 5. Hysteretic loops in X and Z axis direction for initial load in Y axis direction

It is possible to recalculate these loops in coordinates $\varepsilon_y - \varepsilon_x$ (or $\varepsilon_y - \varepsilon_z$ respectively). Recalculated loops are described by equations

$$\varepsilon_x \approx -\mu_{x1}\varepsilon_y + \varepsilon'_{x1}; \quad \varepsilon_z \approx -\mu_{z1}\varepsilon_y + \varepsilon'_{z1}, \quad (18)$$

(here part with "+" is taken into account only for unloading process). In these equations coefficients μ are analogous to Poisson ratio, coefficients ε' are residual deformation.

For free surfaces of specimens equations (18) transforms into

$$\begin{aligned} \varepsilon_x &\approx -0,67\varepsilon_y + (0,32 - 0,83\bar{\rho})\varepsilon_{y\max}; \\ \varepsilon_z &\approx -0,33\varepsilon_y + (0,16 - 0,41\bar{\rho})\varepsilon_{y\max} \end{aligned} \quad (19)$$

If MR is glued to the load surfaces, coefficients of equations (15) are a little bit different:

$$\begin{aligned} \varepsilon_x &\approx -0,67\varepsilon_y + (0,28 - 0,72\bar{\rho})\varepsilon_{y\max}; \\ \varepsilon_z &\approx -0,33\varepsilon_y + (0,14 - 0,36\bar{\rho})\varepsilon_{y\max}. \end{aligned} \quad (20)$$

For multi-time compression in Y direction the structure of MR material becomes stabile. After 80...100 deformation cycles residual deformation stabilize near values

$$\begin{aligned} \varepsilon'_{x\max} &\approx (0,92 - 2,3\bar{\rho})\varepsilon_{y\max}; \\ \varepsilon'_{z\max} &\approx (0,45 - 1,17\bar{\rho})\varepsilon_{y\max}. \end{aligned} \quad (21)$$

Hysteretic loops for X and Z direction in this case are absent, dependencies (15) transform to

$$\varepsilon_x \approx -0,49\varepsilon_y; \quad \varepsilon_z \approx -0,31\varepsilon_y \quad (22)$$

Thus Poisson ratio for multi-time compression are $\mu_{xy} = 0,49$ and $\mu_{yz} = 0,31$.

Poisson ratio for tension in a range of possible strain $-0.05 < \varepsilon < 0$ (if tension is more, structure of MR material will be destructed) is equal to 0.

3.6 Friction Coefficient

Beside elastic-damping elements made of MR material, vibration insulator has many other steel details (cups for elastic-damping elements, unload springs, bolts for connection etc). Some energy of vibration dissipates on friction between these details and elastic-damping element. This process depends on friction coefficient f .

This coefficient was obtained as a ratio between a force which moves an elastic-damping element along plane steel surface and a force which presses the element to this surface. The coefficient is independent on density and force direction (X axis or Y axis). If pressing stress is less than 0.25 MPa the value of coefficient is 0.1, its maximal value is 0.125 for pressing stress 0.75 MPa. If pressing stress is higher, the friction coefficient continue to be constant.

4. Discussion

To explain obtained results one need to consider a structure of wire material (Zhu et al, 2011). A reason of increasing of Young modulus and friction stress for pressure is increasing of number of wire contacts during pressing.

There are two processes which form the dependency of Young modulus and friction stress for pressure on wire diameter. When the wire diameter increases, inertia moment of each wire increases too, stiffness of each wire and sum of elastic force increases of this reason. However from another side, if the wire diameter increases for the same density of MR, number of wire becomes less in a unit of volume. It means that number of wire contacts decreases too. Length of parts of wire (curve beams) between contacts becomes more and stiffness of these curve beams decreases. Two these processes together provides maximum for f_L coefficient.

Two different processes exist for friction force too. From one side, when the number of contacts decreases, sum of friction force in all contacts decreases too. From another side, when the number of contacts decreases, contact force in each contact increases, and friction force in each contact increases too. Two these processes together provides maximum for f_H coefficient.

For MR material the Young modulus for bending is more than for pressure or tension, and friction stress is significantly less. During the bending a detail has neutral axis near which a strain is near to zero. In these conditions the wires in MR material almost have no sleep. It makes stiffness of elastic-damping element more and energy dissipation in it less.

For example, for the same $\bar{\rho} = 0.18$ and $d_w = 0.1$ mm and deformation in Y axis direction (it takes place for ring-shape elastic-damping element, because during its manufacturing the ring is pressed in the direction of the ring axis) initial value of elastic stress near $\varepsilon_y = 0$ for pressure is 2.69 MPa, for bending it is 8.7 MPa, friction stress for pressure is 0.06 MPa, for bending is 0.005 MPa (Yan et al. 2010a).

Equation for friction stress for bending (in contrast to Young modulus) isn't obtained in the present research. This stress depends not only on density and wire diameter but on thickness of MR element and amplitude of bend. Perhaps it is possible to solve this problem theoretically: to consider the bending element as a sum of many layers with compression or tension in each of them with different amplitude. Thus this problem needs future research.

Equations (5) – (12) and (16) – (17) allows approximate calculation of area of hysteretic loop (any mistake will be in the ends of loop, because many wire contacts still didn't sleep there) and area under a line of elastic force. A first value is an energy dissipated during one cycle of loading, second one is potential energy of deformation, their ratio is energy dissipation coefficient ψ . Thus it is possible to compare this coefficient for different load directions. If to take only first parts in equations (5) – (8) (let we suppose that $\bar{\rho}$ and d_w are the same and strain is little), ratio of elastic stress for Y and X axis directions is about 49.6:11=4.5, ratio of friction forces is about 0.56:0.167=3.35, therefore ratio $\Psi_x : \Psi_y \approx 1/(3.35 : 4.5) \approx 1.34$.

Thus the energy dissipation coefficient is more in direction of X axis.

Because the hysteretic loop for shear is near to linear, its area (with any mistake for the ends of loop) is about $2\tau_H \gamma$, and potential energy of deformation is $G\gamma^2 / 2$. Thus its ratio is $\tau_H / G\gamma$.

For the same shear angle γ from equations (16) and (17)

$$\psi_{yx} : \psi_{yz} \approx \frac{0.151\bar{\rho}^{1.3}18.3\bar{\rho}^{1.7}}{6.65\bar{\rho}^{1.7}0.301\bar{\rho}^{1.3}} \approx 1.38.$$

It is follow from equations (16) and (17) that ψ for shear is proportional to $(\bar{\rho})^{-0.4}$.

Values of ψ_x and ψ_{yx} are comparable.

For $\varepsilon = \gamma = 0.1$ ratio $\psi_{yx} : \psi_x \approx 1.12$.

For $\varepsilon = \gamma = 0.15$ ratio $\psi_{yx} : \psi_x \approx 0.82$.

Unfortunately till this time that few researchers who work on wire damping materials (Andrés and Chirathadam, 2011; Li et al, 2010; Tian et al, 2008) obtained for researched materials the linear stiffness and energy dissipation coefficient only, and for one-dimensional load process. It isn't enough to describe a behavior of material in contemporary vibration protection systems.

5. Conclusions

Results of the present research give constants for calculation of vibration isolators made of MR material by finite element software. It gives a method for research of other pressing wire damping materials (such as "metal-flex", "spring cushion", "wire mesh damper" etc).

All these results are presented for MR material made of stainless steel wire. However sometimes another materials are used in MR (for example, copper wire to increase heat conductivity). Thus the constants will depend on properties of wire material and on part of materials in a wire body. This problem needs future research.

If MR material is used in squeeze film damper, it is necessary to research how its constants will change in other media (such as oil).

If MR material is too thin, during its deformation some wires don't meet other wire, and increasing of number of contacts doesn't take place. All results above are correct if H/d_w ratio is more than 100 (here H if thickness of element in load direction). Elastic-damping elements for pipeline supports have H/d_w ratio about 10...30, it is necessary an additional research of its properties.

All results above are for new MR material. An important problem is a life-time of vibration insulators made of MR material. Constants of material will change during its wearing. It depends (Ao et al, 2006) on number of load cycles, vibration stress, static preload stress, density of material, wire diameter, working temperature of vibration insulator. The present paper gives a method for future researches of these influences.

6. Support

This work was supported by the Ministry of education and science of the Russian Federation in the framework of the implementation of the Program of increasing the competitiveness of SSAU among the world's leading scientific and educational centers for 2013-2020 years.

References

- Al-Khateeb, E. M. (2002). Design, Modelling and Experimental Investigation of Wire Mesh Vibration Dampers. Department of Mechanical Engineering, Texas A & M University, pp. 215.
- Andrés, L. S., & Chirathadam, T. A. (2011). Metal mesh foil bearing: Effect of motion amplitude, rotor speed, static load, and excitation frequency on force coefficients. *Proceedings of the ASME Turbo Expo 6 (Parts A and B)*: 465-476.
- Ao, H. R., Jiang, H. Y., Wang, S. G., Xia, Y. H., Jia, C. H., & Ulanov, A. M. (2003). Experimental research on anisotropic characteristics of elastic damping elements of metal rubber. *Journal of Materials Engineering*, (9), 11-14.
- Ao, H., Jiang, H., & Ulanov, A. M. (2005). Dry friction damping characteristics of a metallic rubber isolator under two-dimensional loading processes. *Modelling and Simulation in Materials Science and Engineering*, 13(4), 609–620. <http://dx.doi.org/10.1088/0965-0393/13/4/011>
- Ao, H., Jiang, H., & Ulanov, A. M. (2006). Estimation of the fatigue lifetime of metal rubber isolator with dry friction damping. *Key Engineering Materials*, 326-328 II: 949 – 952. <http://dx.doi.org/10.4028/www.scientific.net/KEM.326-328.949>
- Choudhry, V. V. (2004). Experimental Evaluation of Wire Mesh for Design as a Bearing Damper. Department of Mechanical Engineering, Texas A&M University, pp. 86.
- Gildas, P. (1989). Pat. FR2630755, France, B23P17/06; B23P17/00; (IPC1-7): D03D3/02; D04B1/14; D04H3/00; F16F7/00. Spring-type cushion material, spring-type cushion made thereof and system for producing the spring-type cushion material.
- Jiang, H. Y., Hao, D. G., Xia, Y. H., Ulanov, A. M., & Ponomarev, Y. K. (2008). Damping characteristics calculation method of metal dry friction isolators. *Journal of Beijing Institute of Technology (English Edition)*,

17(2), 173-177.

- Jiang, H. Y., Zhang, R. H., Zhao, K. D., & Novikov, D. K. (2005). Analysis of damping characteristics of squeeze film damper with metal rubber. *Journal of Propulsion Technology*, 26(2), 174 – 177.
- Kozian, R., & Schmoll, E. (1998). Pat. DE19629783, Germany, F16F1/373; F16F1/36; (IPC1-7): F16F1/38. Vibration absorber with spring cushion and two end parts.
- Li, D. W., He, Z. B., Ren, G. Q., & Zhang, L. (2010). Study on static deformation characteristics of metal rubber material. *Advanced Materials Research*, 97–101, 826-829. <http://dx.doi.org/10.4028/www.scientific.net/AMR.97-101.826>
- Tian, Z. D., Yao, X. L., Shen, Z. H., & Deng, Z. C. (2008). Mechanical behaviors of a marine vibration- reduction and shock-resistant isolator based on MR. *Journal of Harbin Engineering University*, 29(8), 783-788.
- Timoshenko, S., & Goodier, J. N. (1951). *Theory of Elasticity*. McGraw-Hill Book Company, New York, Toronto, London, pp. 506.
- Ulanov, A. M., & Lazutkin, G. V. (1997). Description of an arbitrary multi-axial loading process for non-linear vibration isolators. *Journal of Sound and Vibration*, 203(5), 903-907. <http://dx.doi.org/10.1006/jsvi.1996.0908>
- Ulanov, A. M., & Ponomarev, Y. K. (2009). Finite element analysis of elastic-hysteretic systems with regard to damping. *Russian Aeronautics*, 52(3), 264-270. <http://dx.doi.org/10.3103/S1068799809030027>
- Xia, Y. H., Jiang, H. Y., Wei, H. D., Yan, H., & Ulanov, A. M. (2009). Shock protection characteristics of metal rubber isolators. *Journal of Vibration and Shock*, 28(1), 72–75.
- Xia, Y., Yan, H., Jiang, H., Ao, H., & Ulanov, A. M. (2005). Determination of elastic modulus of ring-like metal rubber isolator. *Lubrication Engineering*, (3), 34-36, 39.
- Yan, H., Wang, L., Jiang, H. Y., & Ulanov, A. M. (2010a). Analysis of the basic mechanical parameters of metal rubber materials. *ICACTE 2010 – 2010: 3rd International Conference on Advanced Computer Theory and Engineering, Proceedings*, (1), V1396-V1399.
- Yan, H., Wang, L., Jiang, H. Y., & Ulanov, A. M. (2010b). Research on mechanical performance parameter of metal rubber. *Proceedings - 2010 WASE International Conference on Information Engineering, ICIE* (3), 123-126.
- Yan, H., Zhang, W. J., Jiang, H. Y., & Chen, L. (2013). Experimental study of fatigue of metal rubber vibration isolators under pulsating cyclic stress. *Applied Mechanics and Materials*, 385-386: 180-183. <http://dx.doi.org/10.4028/www.scientific.net/AMM.385-386.180>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

Vision-Based Row Detection Algorithms Evaluation for Weeding Cultivator Guidance in Lentil

Hossein Behfar¹, HamidReza Ghasemzadeh¹, Ali Rostami², MirHadi Seyedarabi³ & Mohammad Moghaddam⁴

¹ Department of Agricultural Machinery Engineering, Faculty of Agriculture, University of Tabriz, Tabriz, Iran

² Photonic and Nanocrystal Research Lab., Faculty of Electrical Engineering, University of Tabriz, Tabriz, Iran

³ Department of Communications, Faculty of Electrical Engineering, University of Tabriz, Tabriz, Iran

⁴ Department of Agronomy, Faculty of Electrical Engineering, University of Tabriz, Tabriz, Iran

Correspondence: Hossein Behfar, Department of Agricultural Machinery Engineering, Faculty of Agriculture, University of Tabriz, Tabriz, Iran. E-mail: h_behfar@hotmail.com

Received: March 28, 2014

Accepted: May 4, 2014

Online Published: September 26, 2014

doi:10.5539/mas.v8n5p224

URL: <http://dx.doi.org/10.5539/mas.v8n5p224>

Abstract

It is important to detect crop rows accurately for field navigation. In order to accurate weeding, cultivator guidance system should detect the crop center line precisely. The methods of vision-based row detection for lentil field were studied. Monochrome and color images were used in this research. The color images are transformed into grey scale images in two different formulas to make comparing among them and find an optimal one. In order to detect the center of the crop row rapidly and effectively, Hough transform and gravity center image processing algorithms were applied to acquired images.

The field crop images were segmented into two parts by using optimal thresholding (plant and soil as a background), then Hough transform was applied on these binary images. Gray scale images were used in gravity center method. The center line detection algorithms were tested for two weed distribution density, include general and intensive. It was observed that both systems successfully detects and calculates the pose and orientation of the crop row on synthetic images. The mean errors between the calculated and manually estimated lines were obtained. Mean errors for Hough transform and gravity center methods were 8 and 10 mm with standard deviations of 7 and 12 mm in general distribution density and 12 and 16mm with standard deviation of 11 and 15mm in high distribution density, respectively. Computational time for Hough transform and gravity center were 0.7 and 0.4 s for general distribution density and 1.2 and 0.8 s for high distribution density, respectively.

Keywords: row detection, lentil, machine vision, weeding cultivator

1. Introduction

The presence of weeds in agricultural fields leads to competition between weeds and planted crops. Therefore, it is necessary to eliminate the weeds for better crop growth. The two widely used methods for weed control are chemical and non-chemical weed control. The inorganic chemicals are mainly used to eliminate the weeds in agricultural fields and to increase and protect the crop production. During a past few years, the primary concern in agriculture was organic production. In organic farming no herbicides are permitted, since they have adverse impacts on the environments, soil health, food safe and they cause water pollution and disorder population of healthy worms, and other soil organisms (Asif, Amir, Israr, & Faraz, 2010).

Non-chemical weed control uses some mechanical technique to remove weeds. Manual weeding require labors for which is expensive and is not always available (Dedousis, 2007). Tractor driven cultivator was substituted for manual weeding. It almost improves operation efficiency and productivity. In mechanical weed control, weeding equipment is able to control the weeds between the rows. The lateral position of weeding cultivator tools relatively to crop rows, are usually controlled by the driver of the tractor. This task needs more concentration and could hardly be maintained during a long period, because driver has to give his attention to both tractor and tools position, therefore it exhausts the driver. The precision of the driver guidance between rows determines width of hoeing unit. If driver couldn't be able to achieve the required driving accuracy, the mentioned width should be decreased, so inter-row untreated area would be increased. Weed control within the row and untreated area

requires a lot of manual labors, if we could decrease untreated area between rows, we will be managed to decrease the time needed for this manual working. Under these circumstances, developing an assistant system that can automatically detect rows and guide tools or vehicles, would be helpful.

GPS-based and vision-based guidance systems are the most promising navigation methods for the autonomous guidance (Wilson, 2000; Hague, Marchant, & Tillett, 2000).

GPS-based guidance system relies on prior predefined path information that had been already obtained by sowing machines. However, the coordinates of the crop rows in the real world are currently unavailable (Bakker, Wouters, & Asselt, 2008). GPS has a common limitation on obtaining local accurate position which is often very important for performing efficient automated field operations. The guidance system based on GPS requires more accuracy than common GPS receivers, RTK-GPS is expensive and isn't available in all over the world.

Navigation along the crop row is based on real-time sensed data. Machine vision can obtain this real-time information by detecting rows. In contrast to a real-time differential global positioning system (RTK-GPS), Machine vision is cheaper and has a higher precision. Furthermore, machine vision can provide local or relative information.

In vision-based guidance, data is provided by images acquired from vehicle-mounted cameras. As most crops are cultivated in rows, it is the key to find guidance information from crop row structure in vision-based guidance systems, to precisely control tool or vehicle.

CPU's speed development in recent years copes with high computational load that is needed to vision-based tools and vehicles guidance. A machine vision-based guidance system manages to achieve accurate navigation in an appropriate time that is necessary for real time control. It also does not require a priori field information

The problem of guiding automatically an implement in a farm using machine vision is not new, many technologies and algorithms of image processing were investigated to find guidance line from the row crops.

Recent developed technologies could be divided into the two general categories: autonomous steering and guidance assistance. The main purpose of them, were to evaluate the position of the tool or vehicle in the field relatively to the rows (Leemans and Destain, 2006).

Billingsley and Schoenfisch (1997) presented a method to steer a tractor by following cotton rows. At first they segmented image pixels by applying thresholding. The line position was determined by regression method. Hague et al. (2000) used Hough transform and Kalman filter methods for localizing the row structure and obtaining experimental automatic vehicle position based on image processing.

Sogaard and Olsen (2003) mounted a camera on a hand operated vehicle and later on a cultivator to evaluate the accuracy of guidance system based on machine vision. At first they divided images into band strip. Then they calculated their center of gravity. The row position was calculated by weighted linear regression. The standard deviation was about 15 mm. Tillett, Hague and Mile (2002) applied a method similar to previous work. They used state vector and Kalman filter instead of regression to determine the row position. They succeed in getting the precision of 16 mm.

The Hough transform is the most common crop row detection algorithm used. The Hough transform algorithm is used in line detection for its high robustness. The Hough transform was first proposed by Marchant and Brivot (1995) to detect the crop center line.

Marchant (1996) applied the Hough transform to infra-red images to detect crop rows. Leemans and Destain (2006) used a mean shift algorithm to search the maximum value in the Hough transform domain, to identify the crop rows. To obtain good results for detecting several crop rows, an adapted Hough transform was combined with a priori knowledge of the spacing between rows, and then tested. Åstrand and Baerveldt (2005) applied a row detection method, which combined the Hough transform and the geometric model. The Hough transform was used for detecting crop row structures, while the geometric model determined the link between the image and the ground coordinates. Bakker et al. (2008) opposed a gray-scale Hough transform based crop row detection algorithm. At first color image were transformed into a gray scale images in their method. Then the Hough transform was applied on these gray images to detect the crop rows. They selected three rectangular sections of crop row spacing, and then he summed up the grey values of the sections and used the grey-scale Hough transform to find the row.

Pla, Sanchiz, Marchant and Brivot (1997) offered an algorithm based on finding the vanishing point of the rows. Tillett et al. (2002) described an algorithm based on the periodic variations of brightness between the plants and soil in the parallel crop rows.

Han et al. (2004) described a guidance line detection algorithm. The algorithm first segmented images using the K-cluster algorithm, then detected rows with the moment algorithm and guidance line was selected based on the cost function.

Kise and Zhang (2008) developed a stereo-vision system capable of performing three dimensional (3D) field mapping for measuring crop height and volume and detecting crop rows in 3D for tractor guidance.

Jiling and Liming (2010) offered a vision-based row guidance method to guide a robot platform which was designed independently to drive through the row crops in a field according to the design concept of open architecture. Their method's accuracy of row guidance was up to $\pm 35\text{mm}$.

Tillett et al. (2002); Åstrand and Baerveldt (2002); Bakker et al. (2008) applied their vision based crop row detection algorithms in a sugar beet field. Tillett and Hague (1999); Sogaard and Olsen (2003) tested their vision guidance methods for cereals. Some of the researchers tested their guidance system for Lettuce and tomato (Slaughter, Chen, & Curley, 1999), cotton (Billingsley & Schoenfisch, 1997; Slaughter et al., 1999) and cauliflower (Marchant & Brivot, 1995), soybean (Kise et al., 2005). But these vision based guidance algorithms were not evaluated for lentil.

Different automatic guidance algorithms have been developed for agricultural applications. But as it was mentioned before above, Hough transform and gravity center were the most successful methods that were applied in most researches.

In this paper, these two successful algorithms were studied on images that were acquired from lentil farm by an experimental platform constructed in University of Tabriz. The objective was to get the best image processing method to navigate weeding units automatically between the inter-row spaces of crop for improving the percentage of treated area, reduce labor cost and time.

2. Materials and Methods

A lentil experimental field for performing the tests was prepared in research farm. Lentil seeds were sown manually to eliminate adverse effect of seed bouncing. Two bars were inserted into the ground on the sides of field and a rope was fastened on them. And lentil seeds were laid in the furrow exactly below the rope. The length of the experimental field was about 40 m. Three rows were sown at distances of 30 cm (figure 1).

Two image acquisition systems were used. First system was a CMOS monochrome camera equipped with a near infra-red band-pass filter. It was used to remove visible wavelength to enhance the images. Second system was a CCD camera with color output to obtain sample images from lentil farm. The resolution of image was 640×480 pixels. The focal length of the camera lens was 8 cm. The cameras were mounted on an experimental platform (figure 1).



Figure 1. The experimental platform and lentil farm

The mobile platform was driven by a DC electric motor over the field. The platform consisted of two main parts: stationary and movable frames. Stationary frame was driven by a DC motor on 4 wheels. The motor speed could be controlled by PWM pulses adjusted by the computer. The cameras were mounted on the movable section of the platform. They were able to get the images and send them to computer to save them. One image frame per second was saved in the computer. The camera was directed downward at a specific angle 30° , and the vertical distance from the camera to the ground was 100 cm. The covered area by one image was 1 m long in row direction and 1.5m wide. After acquiring the images, two crop-row detection algorithms were executed on a computer.

Image processing was performed using an industrial laptop computer having an Intel Dual Core i7-2620M 2.7 GHz CPU 8 GB RAM.

Images were acquired at the Agricultural faculty Farm (Khalatposhan) in Tabriz, during growing season of 2012 and 2013. To test the robustness of the system, they were taken under various weed and soil conditions. In three stage of plant growing tests were undertaken. The crop heights in the images were around 5 (small), 16 (medium) 24 (large) centimeters (approximate 1-3 weeks of growth time).

Suitable ROI (region of interest) was selected to restrict the image processing and to reduce the processing time. This ROI was selected based on calibration process. A striped course textile with 10cm width vertical black and white strips, was used to calibration, as was shown in figure 2. This area should be selected as a rectangle with 30cm width so three vertical strips were selected, image perspective model made the ROI figure to be trapezoidal. The corner coordinates of this Trapezoid were found in the image. These coordinates were used to separate ROI from the image. ROI selection is set manually as shown in figure 2.



Figure 2. Calibration process

2.1 Image Processing

In order to detect the center line of the crop row in field images, several image processing steps should be applied to the field images, which include color transformation, image segmentation, noise removal, edge detection and line parameter calculation.

2.1.1 Transform Color Images to Gray Images

The crop-row image mainly consists of crops and background (such as soil). There is a big difference between the crop and background in color. Therefore, color was taken as the feature. But just gray-scale images could be used in image processing. The gray images with bimodal histograms can be segmented effectively by a threshold segmentation algorithm. The color field images, which were taken under natural conditions, were grayed by three color features.

In the first color transformation method, the color images were transformed into a grey image by emphasizing the green value and decreasing the red and blue value. The green value was larger than the red and blue value. Its principle is shown as equation (1).

$$I_1(x,y)=2\times G-B-R \quad (1)$$

Where G, R, B were equal to the green, red and blue value of point (x, y) respectively in the color image.

$I_1(x, y)$ denotes the grey value of pixel(x, y) in the grey image, and were limited to [0, 255].

The second color transformation was performed using the Intel image processing formula. This formula is used for color to gray conversion in common image processing toolboxes like MATLAB.

$$I_2(x,y)= 0.212671R + 0.715160G + 0.072169B \quad (2)$$

Gray-scale images were stored with 8 bits per sampled pixels, so color conversion decreased the image processing calculation load by two third.

So we had three kinds of gray-scale images: two types of transformed images from color images and images acquired directly from monochrome camera equipped with a NIR band-pass filter.

2.1.2 Segmentation

Segmentation technique was used to separate the soil and crop by their color difference. After segmentation images were partially processed that includes just crops information. Other objects like soil, stone and residues were considered as background.

A threshold segmentation algorithm was applied in this system for its simplicity and speed. The outdoor agricultural navigation system needs to be adaptable to various weather and lighting conditions. It requires the proposed threshold values selection has the robustness to the lighting variety as well, so adaptive global thresholding algorithm was used to perform the segmentation. Constant threshold values didn't show satisfactory result in changeable condition. Otsu automatic threshold value calculation was applied on the image gray values as the flow chart of figure 3. The basic principle of Otsu is looking for the best threshold value to divide grey-level histogram of an image into two classes on the condition that between-classes variance is maximal.

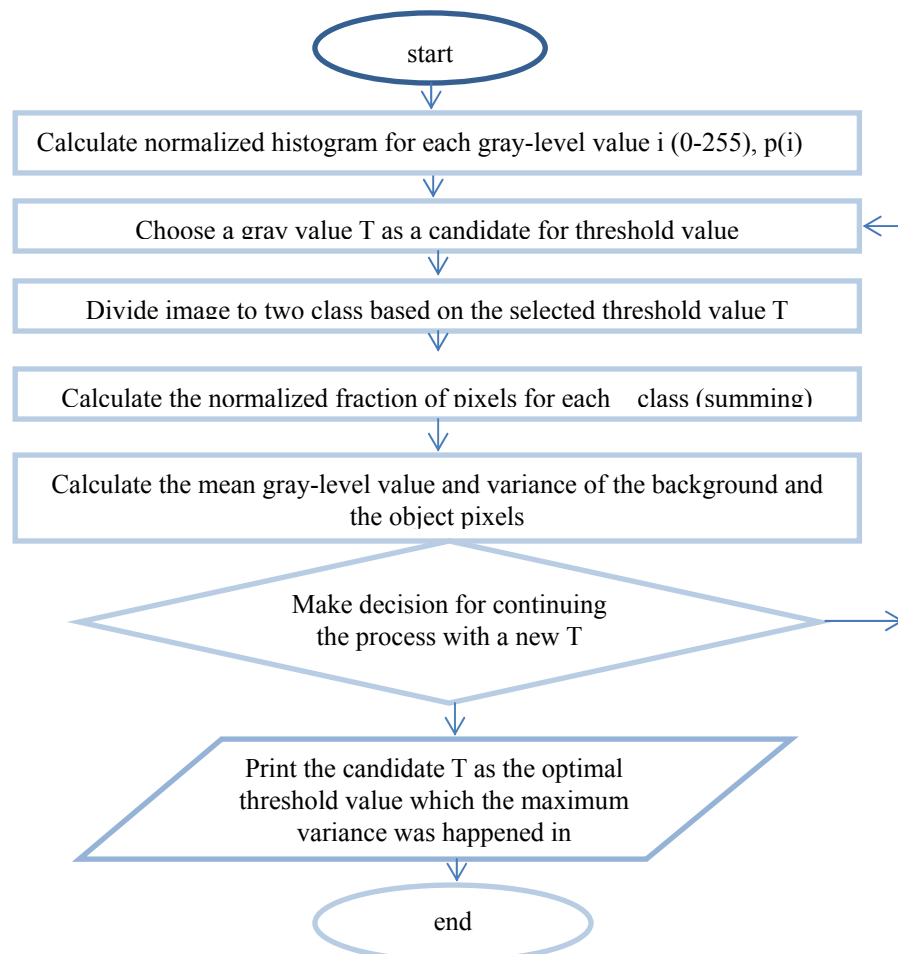


Figure 3. Otsu automatic threshold value detection method

2.1.3 Noise Reduction

To improve the resulted binary images and reduce the fine details in an image the median filtering was used. The median filtering reduced the details in the images by applying the image blurring operation.

2.1.4 Edge Detection

After converting the image into the binary images, edge detection was performed. To optimize and better edge information, Sobel edge detection was applied. The vertical and horizontal Sobel gradient operators were used to get better edge detection results. The Sobel edge detection was computationally simple so it was spent just about 0.05- 0.07s for the Sobel implementation.

2.2 Line Detection

The center line of the crop row should be detected for tool navigation. Center line of the crop row was detected by two algorithm based on the Hough transform and gravity center.

2.2.1 Hough Transform Method

Binary images were used for Hough transform based row position detection method.

The Hough transform is a line detection algorithm based on the relationship between point (x,y) and line ($y=ax+b$). For applying Hough transform algorithm, a set of points in image space were mapped to a set of lines in parameter space. If these points in image space are all located on the line ($y = ax+b$), the mapped lines in parameter space will pass the common point (a, b). This parameter space was not suitable for the studies like our research that 'a' approaches infinity as line approaches the vertical direction. So, normal representation of a line was used. It is named Hough space.

$$X\cos\theta+ysin\theta=p \quad (3)$$

In this study p and θ were position and orientation of the crop row.

After transforming the points in the image space to Hough space, peak detection was performed on all the points of the space that are identified as accumulators. The peak point (p,θ) coordinate was the line parameters that we were looking for.

2.2.2 Gravity Center

Gray scale images were used in gravity center row position detection method. The gray images were divided into a number of horizontal strips as shown in figure 4. Three factors affect the number of strips: existence of minimum one plant in a strip, computational load, and accuracy. In order to reduce the amount of subsequent computations, it may be better choose the least number of strips. Exceeding number of strips causes more accuracy, because it makes more points for fitting line. Drilled crops like cereals and pea don't have any given interval between crops so more strips can be selected. When strip numbers were exceeded more than 10, significant increasing in computational time was observed, so 10 strips were chosen (figure 4). It was just 4 strips were shown in the figure 4 for convenience and better appearance.

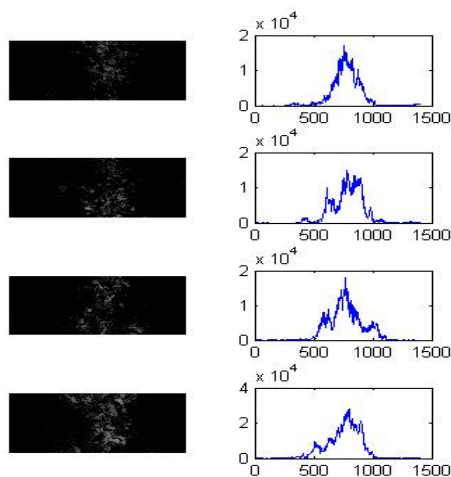


Figure 4. image division and their histogram distribution

The gravity center method was used for getting the center of the row for each image strip. The resolution of image was considered to be $M \times N$ pixels. The image strip size is $M \times k$. $I(i, j)$ was the grey value of the pixel in position (i, j) , $w(j)$ was the result of summation of column j :

$$w(j) = \sum_{i=1}^k I(i, j) \quad j = 1, 2, 3, \dots, M \quad (4)$$

And s was the sum of the grey values of the whole image strip:

$$s = \sum_{j=1}^M w(j) \quad (5)$$

Right section of figure 4 shows the grey values distribution of each column. As can be shown in left section of figure 5, the grey values of crop's area were higher than the background. Weeds roll in the both side of the crop row were the same in the distribution histogram. Weighted center of strips was the position of the column (j) that was encountered the condition of equation 5.

$$\sum_{j=1}^x w(j) \approx \left(\frac{s}{2}\right) \quad (6)$$

This process was continued for all the strips and weighted center points of strips were computed. The best fitted line to these points was estimated by using the regression method. So the crop center line was obtained.

3. Results and Discussion

The performance of the mentioned algorithms, were evaluated against three criteria: accuracy, robustness and computational requirements. The mean errors between the calculated and manually estimated lines pose (ρ) were obtained to detect the method accuracy. The orientation (θ) was used just for the acceptance basis of the process.

Since the main purpose of the study was to achieve real-time automatic guidance of weeding cultivator in lentil, the time requirement was the most important issue.

200 images were randomly chosen to evaluate those criteria. Based on the experiments, the time costs of the image processing were mainly depended on the number of pixels were in the processing. So as were expected, time requirements were lower for images obtained in earlier weeks of growth time plant, because both plant and weed occupied lower area in the images. The average time costs of all levels of Hough transform image processing algorithm (segmentation, filtering and transformation) were 0.5, 0.8, 1.4s for small, medium and large size of plants, respectively.

Both monochrome images acquired with a camera equipped with a band filter and gray-scale showed better robustness in different weather (sunny or cloudy) and soil (dry or wet) conditions. They showed almost constant results. These gray-scale transformation methods were tested by Hough transform method. Mean error were 7, 8 and 15mm and standard deviation were 8, 10 and 19mm for medium size of plants with transformed images with filter, green-emphasized (equation 1) and ordinary gray-scale (equation 2) transformation methods, respectively. The experimental results indicated that the filtered images could overcome the impact of shadows better than others. Green-emphasized images was managed to omit shadow from the images, but sometimes it was observed that it omit some information of plant leafs covered by shadow, so it was confronted some errors.

Segmentation time requirement was omitted in the gravity center and calculation load of this algorithm was so lower than Hough transform, so time costs were about 0.3, 0.45 and 0.7 for small, medium and large size of plants, respectively.

The images that edge detection were applied on them in spite of showing the best time costs (about 2-3s) represented the weakest accuracy with mean error of 25-30mm. So these images weren't used any more for the line detection.

Two center line detection algorithms were tested for two weed distribution density, include general and intensive. Mean errors for Hough transform and gravity center methods were 8 and 10 mm with standard deviations of 7 and 12 mm in general distribution density and 12 and 16mm with standard deviation of 11 and 15mm in high distribution density, respectively. Computational time for Hough transform and gravity center were 0.7 and 0.4 s for general distribution density and 1.2 and 0.8 s for high distribution density, respectively.

4. Conclusion

The experimental result indicated that the band-filtered images had lower mean errors and could overcome the impact of shadows.

It was observed that both algorithms successfully detect and calculate the pose and orientation of the crop row on synthetic images. Hough transform demonstrated better accuracy in the images including non-germinated gaps. The gravity center presented better time costs result. Lentil crops are sown with drill planters so its row seems continuous without gaps between plants in the row. In this ideal situation this algorithm showed better result. But the row with non-germinated gaps showed weak results with higher errors. Time costs for Hough transform algorithm was higher than gravity center method.

References

- Asif, M., Amir, S., Israr, A., & Faraz, M. (2010). A vision system for autonomous weed detection robot. *International Journal of Computer and Electrical Engineering*, 2(3), 1793-8163.
- Åstrand, B., & Baerveldt, A. J. (2002). An agricultural mobile robot with vision-based perception for mechanical weed control. *Autonomous robots*, 13, 21-35. <http://dx.doi.org/10.1023/A:1015674004201>
- Åstrand, B., & Baerveldt, A. J. (2005). A Vision based row-following system for agricultural field machinery. *Mechatronics*, 15, 251-269. <http://dx.doi.org/10.1016/j.mechatronics.2004.05.005>
- Bakker T., Wouters H., & Asselt, K. V. (2008). A vision based row detection system for sugar beet. *Computers and Electronics in Agriculture*, 60(1), 87-95. <http://dx.doi.org/10.1016/j.compag.2007.07.006>
- Billingsley, J., & Schoenfisch, M. (1997). The successful development of a vision guidance system for agriculture. *Computers and Electronics in Agriculture*, 16, 147-163. [http://dx.doi.org/10.1016/S0168-1699\(96\)00034-8](http://dx.doi.org/10.1016/S0168-1699(96)00034-8).
- Dedousis, A. P. (2007). An investigation into the design of precision weed mechanisms for inter and intra-row weed control. Ph.D. thesis, Cranfield University, UK.
- Hague, T, Marchant, J. A., & Tillett, N. D. (2000). Ground based sensing systems for autonomous agricultural vehicles. *Computers and Electronics in Agriculture*, 25(1-2), 11-28. [http://dx.doi.org/10.1016/S0168-1699\(99\)00053-8](http://dx.doi.org/10.1016/S0168-1699(99)00053-8)
- Han, S., Zhang, Q., Ni, B., & Reid, J. F. (2004). A guidance directrix approach to vision based vehicle guidance systems. *Computers and Electronics in Agriculture*, 43, 179-195. <http://dx.doi.org/10.1016/j.compag.2004.01.007>
- Jinlin, X., & Liming, X. (2010). *Autonomous Agricultural Robot and Its Row Guidance* (pp. 725-729). 2010 International Conference on Measuring Technology and Mechatronics Automation. <http://dx.doi.org/10.1109/ICMTMA.2010.251>
- Kise, M., & Zhang, Q. (2008). Development of a stereovision sensing system for 3D crop row structure mapping and tractor guidance. *Biosystems Engineering*, 101, 191-198. <http://dx.doi.org/10.1016/j.biosystemseng.2008.08.001>.
- Kise, M., Zhang, Q., & Rovira-Más, F. (2005). A stereovision based crop row detection method for tractor-automated guidance. *Biosystems Engineering*, 90(4), 357-367. <http://dx.doi.org/10.1016/j.biosystemseng.2004.12.008>
- Leemans, V., & Destain, M. F. (2006). Agricultural tools guidance assistance by using machine vision. 7th National Congress on theoretical and applied Mechanics, Faculté Polytechnique de Mons, Belgium.
- Marchant, J. A. (1996). Tracking row structure in three crops using image analysis. *Computers and Electronics in Agriculture*, 15(2), 161-179. [http://dx.doi.org/10.1016/0168-1699\(96\)00014-2](http://dx.doi.org/10.1016/0168-1699(96)00014-2)
- Marchant, J. A., & Brivot, R. (1995). Real-time tracking of plant rows using a Hough transform. *Real-Time Imaging*, 1(5), 363-371. <http://dx.doi.org/10.1006/rtim.1995.1036>
- Pla, F., Sanchiz, J. M., Marchant, J. A., & Brivot, R. (1997). Building perspective models to guide a row crop navigation vehicle. *Image and Vision Computing*, 15, 465-473. [http://dx.doi.org/10.1016/S0262-8856\(96\)01147-X](http://dx.doi.org/10.1016/S0262-8856(96)01147-X)
- Slaughter, D. C., Chen, P., & Curley, R. (1999). Vision guided precision cultivation. *Precision Agriculture*, 1(2), 199-216. <http://dx.doi.org/10.1023/A:1009963924920>
- Sógaard H. T., & Olsen H. J. (2003). Determination of crop rows by image analysis without segmentation.

- Computers and Electronics in Agriculture*, 38, 141-158. [http://dx.doi.org/10.1016/S0168-1699\(02\)00140-0](http://dx.doi.org/10.1016/S0168-1699(02)00140-0)
- Tillett, N. D., & Hague, T. (1999). Computer-vision based hoe guidance for cereal – an initial trial. *Journal of Agricultural Engineering Research*, 74, 225-236. <http://dx.doi.org/10.1006/jaer.1999.0458>
- Tillett, N. D., Hague, T., & Mile, S. J. (2002). Inter-row vision guidance for mechanical weed control in sugar beet. *Computers and Electronics in Agriculture*, 33, 163-177. [http://dx.doi.org/10.1016/S0168-1699\(02\)00005-4](http://dx.doi.org/10.1016/S0168-1699(02)00005-4)
- Wilson, J. N. (2000). Guidance of agricultural vehicles a historical perspective. *Computers and Electronics in Agriculture*, 25(1-2), 3-9. [http://dx.doi.org/10.1016/S0168-1699\(99\)00052-6](http://dx.doi.org/10.1016/S0168-1699(99)00052-6)

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

H.M. The King's Royally Initiated LERD Project on Community Wastewater Treatment through Small Wetlands and Oxidation Pond in Phetchaburi, Thailand

Kasem Chunkao^{1,2}, Wit Tarnchalanukit^{1,2}, Paiboon Prabuddham^{1,2}, Onanong Phewnil^{1,2}, Surat Bualert¹, Kittichai Duangmal¹, Thanit Pattamapitoon^{1,2} & Chatri Nimpee¹

¹ Department of Environmental Science, Faculty of Environment, Kasetsart University, Bangkok, 10900, Thailand

² The King's Royally Initiated Laem Phak Bia Environmental Research and Development Project, Ban Laem District, Phetchaburi Province, Chaipattana Foundation, 76100, Thailand

Correspondence: Kasem Chunkao, The King's Royally Initiated Laem Phak Bia Environmental Research and Development Project, Phetchaburi Province 76100, Thailand & Department of Environmental Science, Faculty of Environment, Kasetsart University, Bangkok 10900, Thailand. E-mail: prof.kasemc@gmail.com

Received: May 23, 2014 Accepted: June 18, 2014 Online Published: September 26, 2014

doi:10.5539/mas.v8n5p233

URL: <http://dx.doi.org/10.5539/mas.v8n5p233>

Abstract

The research on community wastewater treatment through small aquatic-planting and grass-filtration constructed wetland in cooperating with oxidation ponds is aimed to establish demonstration models for eliminating organic contaminants under nature-by-nature process, simple technology and less expenses. There have been 7 small wetlands on 100-m x 5-m x 0.75-m small vertical flow construction wetlands (VFCW), 7 small wetlands on 25-m x 5-m x 0.75-m small VFCW in cooperating with zero discharge, 2 experimental plots of planted mangrove forest and 5-consecutive oxidation ponds plus 3 herbivore fishes per square meter.

For research operation, some municipal wastewater were pumped about 0.025-cms flow rate from Phetchaburi collected pond through 18.5-km HPDE pipe with separating receivers: firstly, to small grass and aquatic planted wetlands; secondly, to small constructed wetlands; thirdly, to the planted mangrove forest; and finally to the 5-consecutive ponds in descending order of 20 cm by beginning the depth of 2.6 m at the first pond (sedimentation pond) till 1.8 m at the last pond, In basic principles, an influent has to flow continuously at height of one-third depth below surface of oxidation pond on hydraulic retention time (HRT), then flowing over weir crest about 5 cm. The effluent of each oxidation pond was monthly sampled for analyzing the water quality in order to estimate the efficiency of wastewater treatment. In the same procedures, the effluent from small wetlands as grown by aquatic plants as well as zero discharge had to collect for water quality analysis. The results found the wastewater treatment efficiency above 60 percentages for COD, BOD, and TSS. The usable life of plants for maximum wastewater treatment efficiency were specified at 90 days and 45 days for aquatic plants (*Typha angustifolia* Linn. and *Cyperus corymbosus* Rottb.), respectively. It was noticed that small wetland and oxidation pond were suitable for community wastewater treatment and gained benefits from the wastewater treatment system.

Keywords: community, wastewater treatment, small wetlands, oxidation ponds

1. Introduction

It has been wondered why various types of obligatory engineering-device installation for wastewater treatment from municipals, industrial factories, hotels and livestock farms but polluted water is still expanded to every river, stream, canal, lake, reservoir and wetland. In other words, the failure of engineering tools cannot be specified what is the grassroots of problem. It might be just installed the wastewater treatment devices but they were ignored to operate because of economizing the capital, no plan for monitoring, directly smuggling wastewater to public water sources and no employment of environmental technologists. Moreover, Thai traditional human settlement along both riverbanks would be another cause to pollute community wastewater to stream water as the same as livestock-farm behaviors. This is why stream pollution gradual increasing day by day without stopping the situation.

H.M. King Bhumibol has realized how big problems of seriously stream pollution around the country, and

initiated the Royal Laem Phak Bia Environmental Research and Development Project (Royal LERD) with applying the nature-by-nature processes at Laem Phak Bia Sub-District, Ban Laem District, Phetchaburi Province, Thailand in order to find simplicity technology and probable utilization of local materials or the cheapest expenses for encouraging the community wastewater treatment before draining into the streams. Under stated condition, small wetland and oxidation pond could be the most probable technologies for wastewater treatment from households, villages, communities and municipals. These technologies can be applied to high organic waste concentration, i.e. livestock farms, agro-industrial factories and slaughterhouses as well as industrial factories but pretreatment needs before or after draining wastewater or treated wastewater into the small wetlands and/or oxidation ponds which utilizing nature-by-nature processes.

Theoretically, the nature-by-nature processes are implied as photosynthesis, thermo-siphon and thermo-osmosis processes for supporting naturally bacterial digestion of organic waste as the contaminants in wastewater from any point sources. This could be expected to serve need of the Royal LERD project to take small wetland and oxidation pond for disseminating the know-how on wastewater treatment to the concerned point sources around the country.

2. Material and Methods

2.1 Location of the Royal LERD Project

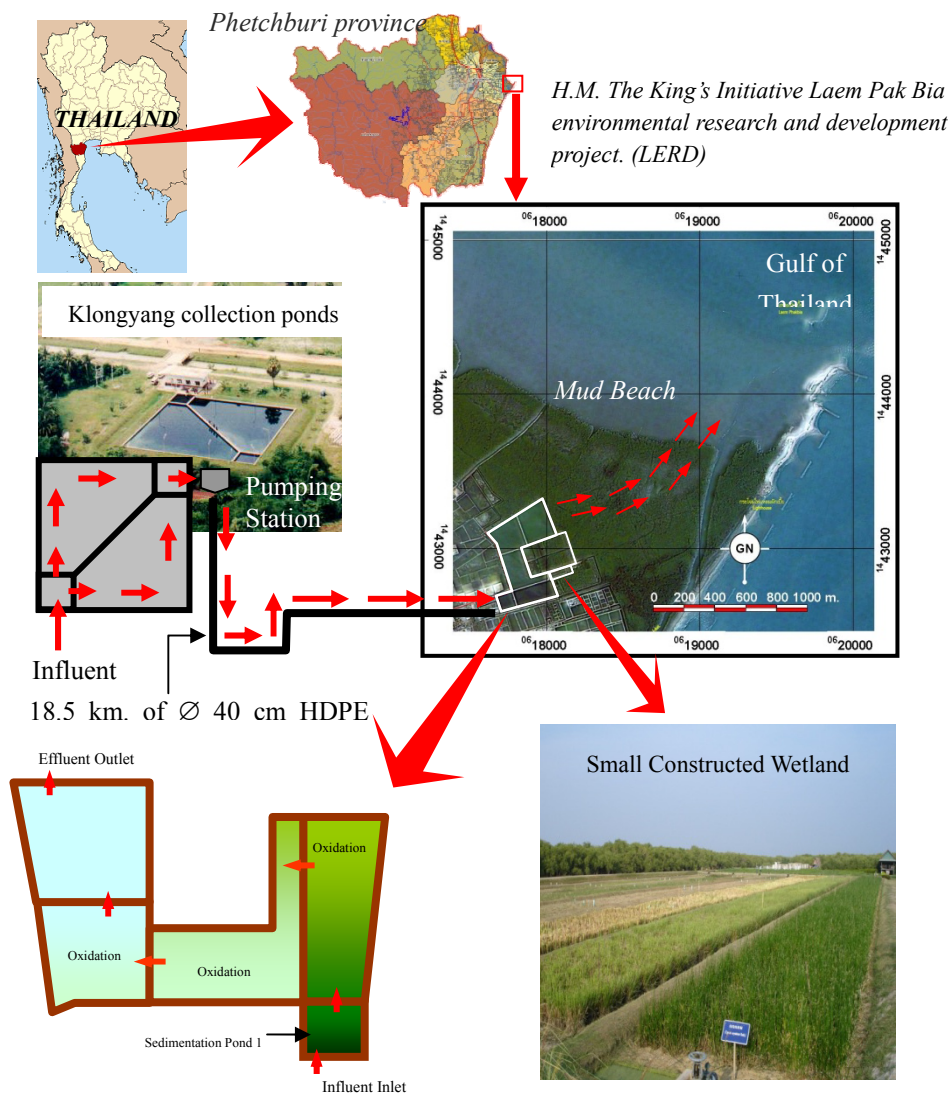


Figure 1. Location of Phetchaburi municipal and the Royal LERD project site at Laem Phak Bia Sub-District, Ban Laem District, Phetchaburi Province

Phetchaburi drainage area is only one system among more 25 rivers in Thailand that composing of one river (Phetchaburi river) to transfer streamflow directly from headwater to the river mouth through five large communities and passing to municipal as a main city before taking the organic pollutants and toxicants to the Gulf of Thailand. For eliminating organic-toxicant wastes in Phetchaburi river, the Royal LERD project has been taken place for decreasing wastewater from fresh-food markets, local sweetmeat factories, municipal sewage, livestock farms, agro-industrial factories and households along the riverbanks by transferring 18.5-km HPDE pipes to the project site at Laem Phak Bia village as seen in Figure 1.

Wastewater from the above point sources flowed through municipal sewerage to four pumping stations which connect to Klongyang collection ponds. Then, wastewater was pumped through 18.5-km HPDE pipes to drain out to two lines: one to small wetlands; and the other to 5-consecutive ponds (Figure 1).

2.2 Small Constructed Wetlands

Small constructed wetlands were made in size of 100-m length, 5-m width and 1-m depth which furnished on each three 300-m-hole pipe in parallel to the bottom in order to allow treated wastewater flowing through outlet. Then after, gravel pavement was performed at about 10-cm depth, another 10-cm soil overtopping, following with 30-cm sandy soils (paddy soil: sand ratio equivalent to 3:1) for growing aquatic plants, and final 30-cm height for wastewater level which flow vertically through soils, sand, and gravel until running out from outlet.

In accordance with research objectives, the small vertical flow constructed wetlands (small VFCW) was divided into four 20-m sections. Each section composed of three 10-cm polyethylene pipes at every section in order to take the vertical flow of treated wastewater to analyze water quality as shown in Figure 2.

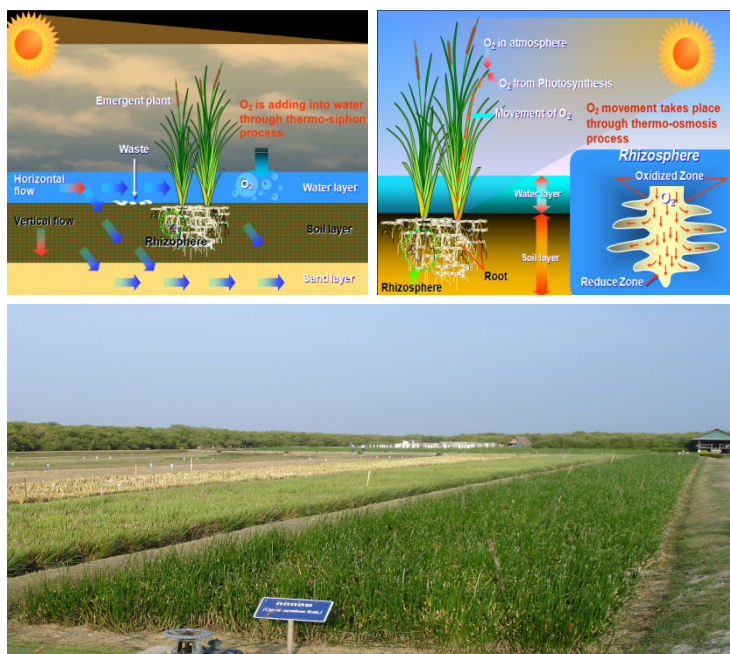


Figure 2. Hypothetical vertical flow constructed wetlands (5 x 100-m surface area and 1-m depth) and filling sandy soils (ratio of paddy soils to sand is equivalent to 3:1) for growing *Typha angustifolia* Linn. and *Cyperus corymbosus* Rottb. as used for community wastewater treatment at the Royal LERD project site

2.3 Five-Consecutive Oxidation Ponds

Construct the consecutive oxidation ponds 1, 2, 3, 4 and 5 with the depth of them about 3, 2.7, 2.5, 2.3, and 2.1 meters, respectively. Then wastewater was flown at the level of 20-cm difference by setting the water depth at 2.7, 2.5, 2.3, 2.1, and 1.9 meters of ponds 1, 2, 3, 4, and 5, equivalent to 75 percent of pond volume. (Figure 1). Finally, continuity flow from Klongyang collection pond about 3,600 cubic meters per day was pumped through 18.5-km HPDE pipes, diverting some part into constructed wetland at setting depth of 30 cm above water surface for 5-day stagnating and 2 days releasing. The other part was flown into 5-consecutive ponds for gradually decreasing the organic waste in wastewater by bacterial digesting process.

2.4 Samples Analysis

Wastewater samples in terms of influent and effluent were collected every ten days, since the year of 2000, for analyzing the water quality indexes, i.e. COD, BOD, TDS, TSS, EC, pH, temperature, salinity, NPK, some heavy metals. At the same time, height measurement of *Typha* and *Cyperus* were conducted until the growth rate equivalent to zero which is the period (90 days of age) of maximum treatment efficiency. *Typha* and *Cyperus* were harvested at age of 90 days for determining the biomass and also to analyze plant nutrients and some heavy metals. Moreover, soil samples were taken before the experiment and after harvesting for analysis of plant nutrients and some heavy metals.

3. Results and Discussion

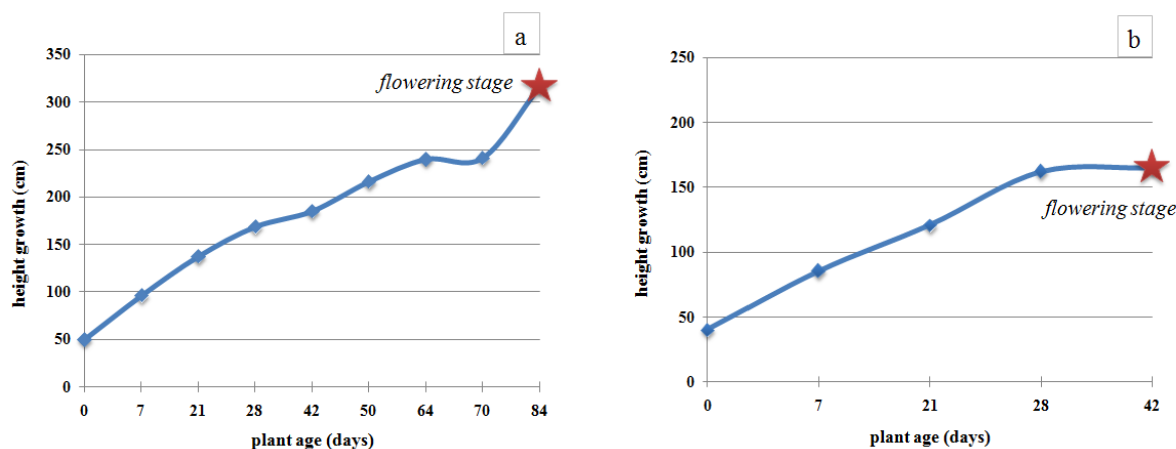
Owing to the Royal LERD project on community wastewater treatment has been launched since 1993 and beginning to collect the samples in the year of 2000 in every ten days for the first 5 years and monthly period after the year of 2003. The better illustration of analyzed data for this study were taken in account with representative climate and wastewater quality will be presented in the following sections.

3.1 Biomass and Accumulative Plant Nutrients

Due to *Typha* and *Cyperus* were planted in the 5-m x 100-m vertical flow constructed wetlands (VFCW) with 35-cm spacing with about 25-cm depth of growing materials (paddy soil-sand ratio is equivalent to 3:1), the measurement of height growth was taken in every seven days as shown in Table 1 and Figure 3. The indicated that average growth rate was more or less zero when their ages reached at about 90 days and 45 days for useful life of *Typha* and *Cyperus*, respectively. The aforesaid statement presented that *Typha* and *Cyperus* grew well in Phetchaburi municipal wastewater and also the maximum effective treatment in constructed wetland concept, although *Typha* was high evapotranspiration rate (Pedescoll et al., 2013; Phewnil et al., 2014)

Table 1. Height growth and biomass of *Typha angustifolia* Linn. and *Cyperus corymbosus* Rottb. as used for community wastewater treatment at the Royal LERD project site in Phetchaburi province Thailand

Plant age (days)	Height growth (cm.)	
	<i>Typha angustifolia</i> Linn.	<i>Cyperus corymbosus</i> Rottb.
7	96.1 ^a	85.5 ^a
21	137.3 ^b	120.9 ^b
28	168.5 ^c	162.1 ^c
42	184.9 ^d	165.1 ^d
50	216.0 ^e	-
64	239.5 ^f	-
70	241.0 ^f	-
84	316.0 ^g	-
Biomass: dry weight (kg/m ³)	0.78 (□ 90 cut off)	0.30 (□ 45 cut off)



(a) *Typha angustifolia* Linn.

(b) *Cyperus corymbosus* Rottb.

Figure 3. Flowering stage of (a) *Typha angustifolia* Linn. and (b) *Cyperus corymbosus* Rottb. as used for community wastewater treatment at the Royal LERD project site in Phetchaburi province Thailand

3.2 Water Quality of Vertical Flow Constructed Wetlands

The vertical flow constructed wetland (VFCW) has been used for 5 functions: firstly, soil media for filtration of larger materials; secondly, supporting unit for bacterial organic digestion processes to become inorganic matters, particularly plant nutrients and some chemicals; thirdly, accumulative chemicals of penetrated wastewater in vertical direction; fourthly, the air filling zone from the processes of thermo-osmosis, thermo-siphon and photosynthesis; and finally, phytoremediation boundary by aquatic plant rhizomes. Such 5-function property of the soil-sand filtration wetlands caused the effect of the water quality indicators of influent greater than the effluent as seen in Table 2 and Figure 4, although the removal efficiency of both aquatic plant species (*Typha* and *Cyperus*) were not different.

Table 2. Water quality indicators of vertical flow constructed wetland (VFCW) as collected at the consecutive distances from head of constructed wetlands for community wastewater treatment in Phetchaburi Province Thailand.

Parameter	Unit	Influent	Effluent		p-value
			<i>Typha angustifolia</i> Linn.	<i>Cyperus corymbosus</i> Rottb.	
pH	(-)	6.4	6.5	6.5	0.422
Total suspended solid (TSS)	mg/L	35.9	18.5	14.8	1.000
Total dissolved solid (TDS)	mg/L	460.0	466.2	449.6	1.000
Dissolved oxygen (DO)	mg/L	0.9	5.0	5.5	1.000
Biochemical oxygen demand (BOD)	mg/L	26.9	10.8	9.8	1.000
Total nitrogen (TN)	mg/L	7.1	4.2	4.6	1.000
Total phosphorus (TP)	mg/L	4.2	2.6	2.1	1.000
Total coliform bacteria (TCB)	MPN/100ml	1.7×10^5	2.8×10^3	3.3×10^2	0.148
Fecal coliform bacteria (FCB)	MPN/100ml	6.8×10^4	9.4×10^2	1.3×10^2	1.000

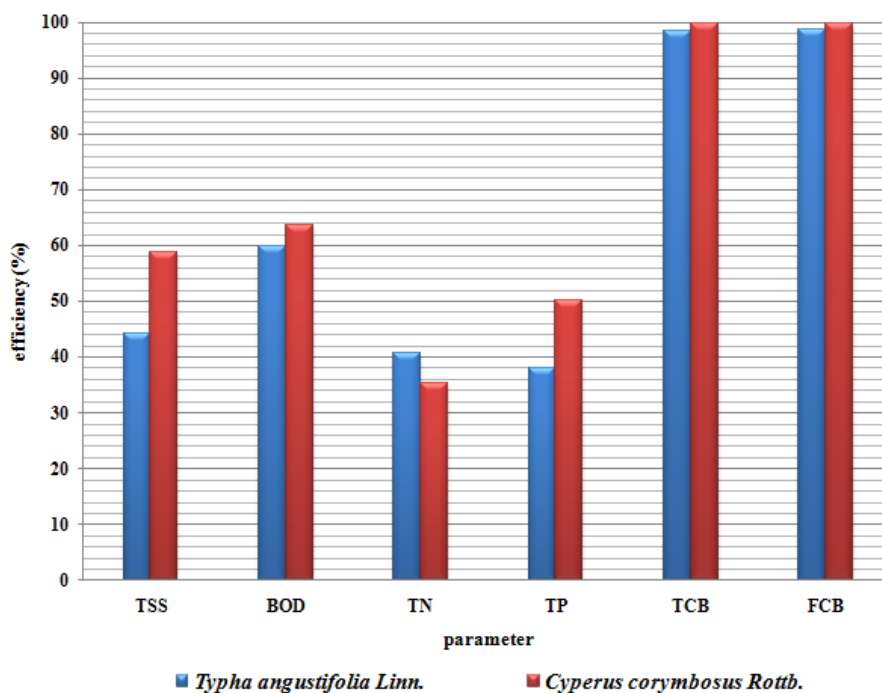


Figure 4. Efficiency of vertical flow constructed wetland (VFCW) as collected at the consecutive distances from head of constructed wetlands for community wastewater treatment in Phetchaburi province Thailand

3.3 Wastewater Treatment Efficiency

Naturally, wetland is the most effective technology for wastewater treatment by not only filtration the larger size of objects and toxicants and also encouraging bacterial organic digestion for becoming to inorganic matter for plant growing.

1) Small Vertical Flow Constructed Wetlands

After analysis of filtrated wastewater samples (3 replications) on 25, 50, 75 and 100 meters from the beginning 0 to 100 m. of those 7 main plots, the results found the averaged values as shown in Table 3. For making clear understanding the function of the vertical flow on grass-aquatic plant filtration plots under the concept of constructed wetland, the water quality indicators was taken an account with BOD and COD as the representative. The results indicated that the values of BOD was gradually decreased from the upper parts of plots to the lowest values at the ends as shown in Figure 5, while COD had the tendency to increase and to get the maximized peak at 50-m length from the inlet, then slowly decreasing until 100-m point of outlet. In principles, the organic matter has to be digested by aerobic and anaerobic processes at the top-layer wastewater and the middle-layer soil growing media during vertical flowing to the bottom of constructed wetlands. The received products from bacterial digestion processes (causing the decrease of BOD) are expected to be inorganic materials as well as some elements and heavy metals which can be removed together with the treated wastewater and causing to increase more concentration of COD at 50-m length, but it trends to decrease because of less product from those organic digestion. In other words, the values of COD are depended on the values of BOD which are related to the rate of bacterial organic digestion.

Table 3. Influences of infiltration process due to vertical flow of Phetchaburi municipal wastewater along with the 100-m plots of constructed wetlands

No.	Plant Species	Distance (m.)	Averaged Values of Compound/Element (mg/l)										Remarks
			COD	BOD	N	P	SS	pH	Cd	Pb	Ni	As	
1	Control Plot	0	95.20	43.70	7.10	4.20	35.90	6.4	0.02	0.13	0.13	17.40	Influent
		25	164.00	5.70	-	-	-	-	0.18	0.76	0.95	5.40	
		50	277.00	7.20	-	-	-	-	0.07	0.42	0.17	5.90	
		75	253.00	6.90	-	-	-	-	0.08	0.33	0.26	2.00	
		100	190.00	5.70	4.80	2.60	40.40	7.6	0.08	0.34	0.18	8.70	
2	<i>Typha</i>	0	95.20	43.70	7.10	4.20	35.90	6.4	0.02	0.13	0.13	17.40	Influent
		25	154.00	4.40	-	-	-	-	0.07	0.36	0.26	2.40	
		50	209.00	2.60	-	-	-	-	0.10	0.55	0.84	7.00	
		75	180.00	1.70	-	-	-	-	0.06	0.43	0.19	4.40	
		100	141.00	2.10	4.20	2.60	18.50	6.5	0.06	0.39	0.16	24.80	
3	<i>Cyperus</i>	0	95.20	43.70	7.10	4.20	35.90	6.4	0.02	0.13	0.13	17.40	Influent
		25	154.00	7.80	-	-	-	-	0.05	0.27	0.24	13.80	
		50	209.00	4.20	-	-	-	-	0.15	0.70	0.83	2.30	
		75	180.00	2.00	-	-	-	-	0.14	0.50	0.89	9.60	
		100	141.00	2.50	4.60	2.20	14.80	6.6	0.14	0.57	0.86	11.60	

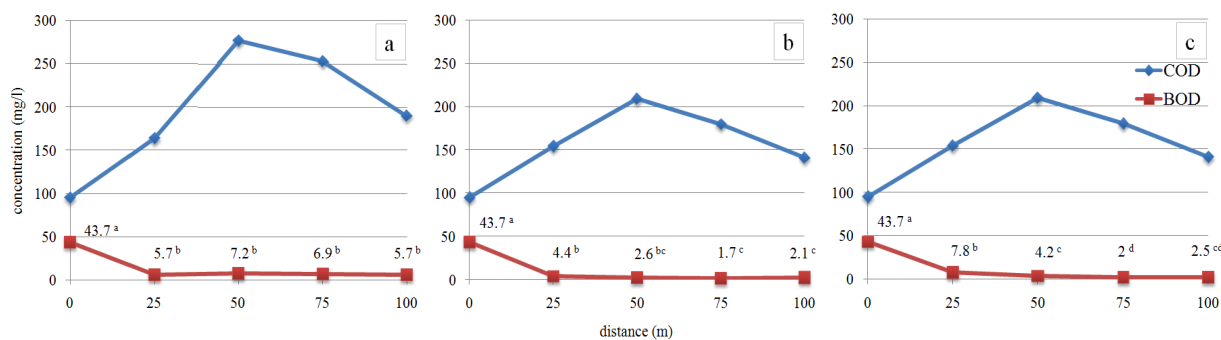


Figure 5. Characters of BOD and COD as released by vertical flow through soil growing media of grass-aquatic plant filtration plots under the concept of constructed wetland; (a) control plot, (b) *Typha angustifolia* Linn. and (c) *Cyperus corymbosus* Rottb

Theoretical speaking, there are hidden two natural processes; firstly, aquatic plant photosynthesis during daytime to provide oxygen as the processing (Sooknah and Wilkie, 2004; Hadad et al., 2006; Tanji et al., 2006), and secondly, thermo-osmosis process produces oxygen to bacteria for organic digestion during photosynthesis process of very young leaves, containing aerenchyma cells, of aquatic plants (Deubigh and Raumann, 1952; Bearnan, 1957; Srivastav and Avasthi, 1975; Grosse, 1989; Buchel and Grosse, 1990; Grofse and Bauch, 1991). The size of the constructed wetland has been used for wastewater treatment from various point sources for long period of time in which it depends on the amount of polluted wastewater, space availability, and topographical characteristics but normally varying on depth 0.3 to 1.0 m, surface area 100 to 600 m²; mostly rectangular-shaped surface area and 1:1.000 slope (Tripathi and Shukla, 1991; Jenssen et al., 1993; Juwarkar et al., 1995; Ahn and Mitsch, 2002; Stottmeister et al., 2003; Maine et al., 2006; Yang et al., 2008; Li et al., 2009; Stefanakis and Tsihrintzis, 2012). The Photosynthetic rate of wetland plants were highly correlated with light intensity and temperature due to influence of oxygen evolving activities and disposal efficiency. The photosynthetic characteristics of wetland species can affect their ability to provide oxygen and remove pollutants (Huang et al., 2010). Furthermore, oxygen transfer and oxygen consumption in constructed wetland revealed with flowing techniques since areal-based oxygen consumption rates in vertical flow system was higher than horizontal flow system (Nivala et al., 2013; Bialowiec et al., 2014). The findings demonstrated that oxygen from thermo-osmosis process was important role for organic digestion in systems.

2) Five-Consecutive Oxidation Ponds

The Oxidation pond treatment system at Laem Phak Bia is wastewater treatment system of the municipality of Phetchaburi province that is a facultative type. It is designed to accept wastewater of about 10,000 m³/day. The system comprises a series of five large, shallow earthen basins: one sedimentation pond, three oxidation ponds, and one stabilization pond. The characteristics of this system are summarized in Table 4. The system Phetchaburi occupies an area of 154,178.5 m². Each pond is separated by an earthen 3-m-width berm a ratio of vertical to horizontal distance of 1:2. Grass is planted at the edge of the pond to protect soil erosion. The outlet structure is a spillway that is constructed as far away as possible from the inlet structure.

Table 4. Characteristics of the model of lagoon treatment system at Laem Phak Bia

Order	Pond	Depth) m(Pond type	Process/Treatment	Retention time (days)
1	Sedimentation	2.5	Anaerobic or Facultative	Sedimentation and primary treatment through facultative digestion	7
2	Oxidation	2.0	Facultative	Addition of O ₂ thorough themo-siphon, facultative digestion and bacteria removal	7
3	Stabilization	1.7	Aerobic	Wastewater polishing and algae removal	7
Total					21

Wastewater from municipality is pumped via HDPE pipe for a distance of 18.5 km to the project area into the sedimentation pond. No additional pumping is necessary, from one pond and enters to the next from the bottom. The BOD concentration was decreased by anaerobic organic digestion process in 18.5-km. HDPE pipe (Poommai et al., 2013). Effluent was discharged after being retained in the system for 21-28 days as the natural purification processes occur. The rectangular weir was introduced to increase the efficiency by oxygen transfer into wastewater as flowing over the weir crest which presented 0.03-m depth showed the highest efficiency of oxygen diffusion and it decreased when water depth increased (Poommai et al., 2012). The quality of treated water meets the effluent wastewater quality standard. The Efficiency of the system is about 50-70% (Table 5).

Table 5. Water quality of lagoon treatment system for Phetchaburi municipal wastewater treatment

Ponds	Sample Stations	Water Quality Indicator (mg/L)							
		BOD	SS	Nitrate	Nitrite	Ammonia	TKN	Phosphate	Potassium
Klongyang	input	31.20	12.80	<0.01	0.013	<0.05	1.930	1.10	14.10
Collection Pond	output	45.07	17.73	0.067	0.013	<0.05	<0.05	3.03	15.77
Sedimentation Pond	input	42.20	16.40	0.037	0.003	<0.05	0.657	2.27	10.57
	output	21.50	21.87	0.153	<0.002	2.437	5.647	2.77	15.87
Oxidation No.1	Pond output	7.33	59.00	0.380	0.297	<0.05	<0.05	1.37	23.10
Oxidation No.2	Pond output	10.87	65.10	0.397	<0.002	<0.05	<0.05	1.37	18.87
Oxidation No.3	Pond output	8.50	38.17	0.137	0.018	<0.05	<0.05	1.07	22.87
Polishing Pond	output	16.40	54.90	0.177	0.006	<0.05	<0.05	0.43	26.60

The treatment processes in lagoon treatment were photosynthesis and thermo-siphon. The thermo-siphon process produces oxygen during evaporation process that causes water surface cooling due to heat being absorbed about 583 g-calories for evaporating 1 g of water, consequently cool surface water moving vertically down to the bottom of wastewater treatment ponds together with free oxygen to bacteria for organic digestion process (Mirmov and Belyakova, 1982; Mevi-Schutz and Grosse, 1988; Ameth and Stichlmair, 2001; Gehlin et al., 2003; Watanabe et al., 2012). Sedimentation pond was the primary treatment of influent where settleable solids were removed. The digestion of organic matters carried out by microorganism under aerobic, facultative and anaerobic conditions in oxidation ponds. Upper portion were an anaerobic zone maintained by oxygen generated by algae and penetration of oxygen from the atmosphere through thermo-siphon process. Symbiotic related exist in this zone. Bacteria use oxygen as an electron acceptor to oxidize the organic wastes to stable product i.e. carbon dioxide, nitrate and phosphate. Algae used these compounds and with sunlight to produce oxygen in photosynthesis. Middle portion were a facultative zone where the volume of oxygen fluctuates based on level of wind action and penetration of sunlight. Therefore, microorganisms in this zone must be capable of adjusting their microbial activity to the change of oxygen level. Bottom portion were an anaerobic zone because stagnant conditions prohibit oxygen transfer to this region. Organic acids and gases were the product from decomposition in this zone and became to carbon source and energy source for microorganism in the aerobic zone. The finally pond was a polishing pond where algae was removed through natural die off process due to lack of nutrients. Furthermore, effluent after treated wastewater flowed over weir crest that finding BOD under standard and also the decreasing of total coliform bacteria and fecal coliform bacteria, particularly the pathogenic bacteria decreasing down to almost zero MPN/100 ml. Because of the effect of solar radiation to dissolved oxygen (DO) and hydrogen peroxide (H₂O₂) which were employed for bacterial organic digestion process in wastewater treatment ponds. The solar radiation showed solar energy between 0-750 W/m² all together with UV-A, UV-B, spectrum and net radiation (Pattamapotoon et al., 2013).

3.4 VFCW Filtration of Toxic Chemicals

It is remarkable that whenever the remediation technique has to be applied for extract some toxic chemicals which might be used for gaining specific profits, particularly for growing cash crops. In order to prove such principles, the roots (rhizomes) and stems of *Typha* and *Cyperus* were sampled for analyzing N, P, K, Cu, Zn, Pb, Cd and Hg, the analyzed results were shown in Table 6. The research results were indicated the amount of N, P, and K higher in stems than in rhizomes accordance with identification of *Typha* and *Cyperus* as hyper-accumulative aquatic plants of macronutrients like N, P, and K in its green stems rather than rhizomes. In contrary, the elements Cu and Zn which are the micronutrients for plant growth found higher values in rhizomes more than stems because of very slow translocation of Cu and Zn from rhizomes to stems, and mostly accumulating around rhizomes (Delgalo et al., 1993). Furthermore, the heavy metals (Pb, Cd and Hg) found in stems more than rhizomes of *Typha* and *Cyperus* since both species has been identified as fast growing aquatic plants that uptaking every elements for short period of time. This is why some heavy metals can be distributed to the green leaves of *Typha* and *Cyperus* as taken for experimenting under the phytoremediation technique and constructed wetland concept.

Table 6. Elements up taking by *Typha* and *Cyperus* on vertical-flow constructed wetlands during 1997-1999 at the Royal LERD project site

No.	Elements	Unit	<i>Typha</i>		<i>Cyperus</i>	
			Root	Stem	Root	Stem
1.	N	percentage by oven	0.87	0.87	0.69	1.11
2.	P	percentage by oven	0.29	0.26	0.22	0.18
3.	K	percentage by oven	0.90	3.42	1.18	3.97
4.	Cu	mg/l	6.60	3.30	9.00	4.00
5.	Zn	mg/l	24.60	16.00	66.30	16.00
6.	Pb	mg/l	42.00	92.30	12.00	36.30
7.	Cd	mg/l	10.30	0.00	7.30	6.40
8.	Hg	µg/l	3.50	3.00	2.40	1.85

Actually, wetland is occurred between the terrestrial and aquatic systems in order to absorb the toxic contaminants by humus, organic matters, and soils before draining away to stream or river. Aquatic plant has eventually to remove contaminants from wastewater and soils as growing units through root system under the osmotic pressure during the photosynthesis processing, then the elements are translocated to accumulate in all parts of vegetative organ, but it depends on the degree of toxic chemical contaminants and aquatic plant species (Tateyama et al., 1967; Reddy et al., 1990; Rai et al., 1994; De Souza et al., 1999; Marin and Ayele, 2003; Pulford and Watson, 2003; Xia and Ma, 2006; Gupta and Sinha, 2007; Wahla et al., 2008; Thaipichitburapa et al., 2010; Zaier et al., 2010; Chunkao et al., 2012).

3.5 Royal LERD Project Benefits

It is noted that the outcomes of BOD treatment efficiency equivalent to 88 % for oxidation ponds and 91 % for constructed wetlands. The main points to produce an effect were not only nature-by-nature process but also the techniques of 5-day stagnating on vertical flow and 2-day releasing the treated wastewater from soil-sand-gravel layering constructed wetlands. The BOD concentration of estuarine water in mangrove forest, over muddy beach and on-site seashore indicated as 5.7, 2.2 and 2.5 mg/L. Surprisingly, they were so much low when compared with wastewater from fresh food market (546.6 mg/L) and in the municipal culverts (164.1 mg/L) on which the constructed wetland was shown high treatment efficiency.

The treated wastewater which obtained from oxidation ponds and constructed wetlands was still treated by the natural mangrove forest and called as "secondary treatment" of such treated wastewater before flowing into muddy beach and seashore in part of the Gulf of Thailand. It could be the vertical flow constructed wetlands and oxidation pond wastewater treatment that worked together with the King's initiative nature-by-nature process causing high rate of bacterial organic digestion processes. Consequently, the abundances of fishes, crams, crabs, muscles, shells and other marine animals have been existed on the Royal LERD project site nearby areas. Benefits from the Royal LERD Project can be described as follows:

- 1) Benefits from solid waste disposal or the concrete box system include compost that can be Utilized to grow flowers. Compost can ideally be mixed with a binding agent and compressed into cubes and used to restore deteriorated mangrove areas. Biological gas, namely methane, from the system was reported to be as high as 60% of the total gas generated under the condition of absence of oxygen. This gas can be utilized for various activities in households such as cooking, lighting and fueling engine.
- 2) Benefits from wastewater treatment systems are utilization of oxidation ponds for aquaculture. Fish farming can be carries out without providing any feedings. The measurement of Nile tilapia (*Oreochromis niloticus*) growth was taken from the third pond in which the water quality was in effluent standard. The fish growth model found that the predicted fish weight revealed with biochemical oxygen demand (BOD), Dissolved oxygen demand (DO), water temperature, concentration of plankton, and ammonia (Dampin et al., 2012) Fish are also safe for consumption. Treated wastewater can be used to grow plants and grass that withstand flooding. The productivity of the grass was also found to be high enough to feed cattle.
- 3) Benefits from sludge include usage of sludge with soil as a substrate to grow flowers and plants. Supakata et al. (2011) showed that using moist sewage sludge as a new source for growing rice was an alternative community wastewater treatment with agricultural benefit and ease management (saved time and space). The finding presented moist sewage sludge piled at the depth of 30 cm could be used to grow rice (*Oryza sativa*) with sufficient nitrogen. Furthermore, Semvimol et al. (2014) presented that gas volume from sludge of oxidation

ponds for community wastewater treatment in which organic matters of both units were digested through the nature-by-nature process in anaerobic condition produced gas 1.8 ml/g (oven dry weight), methane concentration in range 545,686-9,560,606 mg/L.

Socially, the social behavior that causes solid waste and wastewater problems included activities related to food preparation and consumption. People are aware and accept the project. Dissemination of environmental knowledge and public relations must gear toward specific social target group that included commercial, industry, tourism, recreation and the mix. The latter group was mostly the agricultural community.

3.6 Dissemination of Research Results

The Royal LERD project is aimed to disseminate all research results of community wastewater treatment and garbage disposal to the people of Thailand in order to manage their home environment and also to human settlement, villages, urban areas, fresh-food markets, shopping centers, factories as the same as schools, colleges, medical centers, scientific laboratories, tourism facilities, government offices and public areas. The dissemination program has been planned on 5 manners as follows:

1) On-Site Studies

There have been recorded the academic visitors a half of million people that came to on-site visit at the Royal LERD project site on Laem Phak Bia sub-district, Ban Laem district, Phetchaburi province, Thailand as shown in Table2. The typicality of on-site visitors was composed of school boys and girls, school teachers, university professors, technocrats, researchers, environmentalists, local administrators, high ranking administrators, policy makers, general population, foreigners of all continents from Asia, North America, South America, Europe, Africa and Australia.

2) Short-Course Training

Actually, the community wastewater treatment and garbage disposal under the King's initiative nature-by-nature processes can be said as easy in practicing but they are very difficult to understand due to the scientific mechanisms. Therefore, short-course training program would be necessary to organize for some government and private sectors, local administration offices, schools and universities to know how to manage community wastewater treatment and garbage disposal. There were a lot of the said units that request to train the personnel for on-site and out-site programs as shown in Table2, approximately more than 7 programs per year.

3) Academic Servicing

Academic servicing would be the most important program for disseminating the in-depth knowledge on how community wastewater to be treated and garbage to be disposed in terms of scientific processing driving forces from nature-by-nature process. Such intention could be brought to ease the community wastewater treatment and garbage disposal and also to transfer knowledge to another people without doubts. However, there were academic servicing more than 5 programs per month onto industrial factories, schools, universities and production companies as shown in Table2. It would be noted that the Royal LERD project on community wastewater treatment and garbage disposal under the nature-by-nature processes have been known since 1990 among the practical men in the whole country.

4) Publishing Materials and VDO Dissemination

The Royal LERD project has prepared the publication of brochures, leaflets and articles that concern with knowledge on how to treat the community wastewater and to dispose the community garbage under the nature-by-nature processes. Moreover, the project team sent the papers to participate both the national and international conferences on oral presentation and poster sessions as the same as submitting the research papers to publish in both the national and international journals. In order to make clear on know-how to eliminate the community wastewater and garbage, the VDO cassettes have been made in both Thai and English versions for the visitors to learn before looking around the demonstration areas and also distributing as a gift to specified groups.

5) Radio Broadcasting and TV Telecasting

Generally speaking, the Royal LERD project for community wastewater treatment and garbage disposal under the nature-by-nature processes is one part of all 4,350 Royal projects that spread out around the country. They are in groups of agriculture (cropping, livestock, aquaculture), poverty elimination, water and irrigation, reforestation and headwater rehabilitation, public health and environmental protection. Undoubtedly, there must be at least one article to be telecasted on TV and broadcasted on radio every day.

4. Conclusion

The community wastewater treatment with small vertical flow constructed wetlands under the King's initiative on nature-by-nature processes has been conducted since 1990 by transferring wastewater from Phetchaburi municipal (BOD more than 1000 mg/L) through drainpipes (BOD about 450 mg/L) to Klongyang collection pond (BOD 230 mg/L) and then flowing along 18.5-km HPDE pipe by pumping before becoming the Influent (BOD about 70 mg/L) of growing *Typha* and *Cyperus* in the small VFCW units. Due to vertical flow through soil, sand and gravel media at 0, 25, 50, 75 and 100 m. length of constructed wetlands, the results found BOD gradually decreasing but COD increasing to meet the peak at the length of 50 m for all study units as used for serving in phytoremediation technique. In so far, the cutting periods were found 45 days for *Cyperus* and 90 days for *Typha*. The study was also paid more attention on the accumulation of macronutrients (N, P, K) and heavy metals (Pb, Cd, Hg) in stems more than rhizomes but opposite direction on micronutrients (Cu and Zn). In conclusion, the oxidation pond and vertical flow constructed wetland together with the phytoremediation technique would be characterized as the most valuable community wastewater treatment system.

Acknowledgments

We are indebted to His Majesty the King Bhumibol Adulyadej, the King of Thailand for his contribution in creating invaluable knowledge not only to Thai people but also to the people of the world. We are also in debt to Princess Sirindhorn, the chairperson of the Chaipattana Foundation, to conduct this project and supporting the Royal project extension to the people of Thailand and the world to have incalculable knowledge. We would like to give thanks to the staff of Eco-Science Community Research Group(ESCRG), Department Environmental Science, Faculty of Environment, Kasetsart University for facilitating the research personnel and laboratories. With appreciation, we are thankful to our colleagues of "The King's Royally Initiative Laem Phak Bia Research and Development Project at Laem Phak Bia Sub-District, Ban Laem District, Phetchaburi Province, Thailand" for their hard working since 1990.

References

- Ahn, C., & Mitsch, W. J. (2002). Evaluating the use of recycled coal combustion products in constructed wetland: An ecological-economic modeling approach. *Ecological Modelling*, 150, 117-140. [http://dx.doi.org/10.1016/S0304-3800\(01\)00477-X](http://dx.doi.org/10.1016/S0304-3800(01)00477-X)
- Arneth, S., & Stichlmair, J. (2001). Characteristics of thermosiphon reboilers. *International Journal of thermal Sciences*, 40(4), 385-391. [http://dx.doi.org/10.1016/S1290-0729\(01\)01231-5](http://dx.doi.org/10.1016/S1290-0729(01)01231-5)
- Bearman, R. J. (1957). The thermo-osmosis of rare gases through a rubber membrane. *Journal of Physical Chemistry*, 61(6), 708-714. <http://dx.doi.org/10.1021/j150552a002>
- Bialoweic, A., Albuquerque, A., & Randerson, P. F. (2014). The influence of evapotranspiration on vertical flow subsurface constructed wetland performance. *Ecological Engineering*, 67, 89-94. <http://dx.doi.org/10.1016/j.ecoleng.2014.03.032>
- Boyd, C. E., & Vickers, D. H. (1971). Variation in the elemental content of *Eichhonia crassipes*. *Hydrobiologia*, 38, 409-414. <http://dx.doi.org/10.1007/BF00036546>
- Buchel, H. B., & Grosse, W. (1990). Localization of the porous partition responsible for pressurized gas transport in *Alnus glutinosa* (L.) Gaertn. *Tree Physiology*, 6, 247-256. <http://dx.doi.org/10.1093/treephys/6.3.247>
- Chunkao, K., Nimpee, C., & Duangmal, K. (2012). The King's initiative using water hyacinth to remove heavy metals and plant nutrients from wastewater through Bueng Makkasan in Bangkok, Thailand. *Ecological Engineering*, 39, 40-52. <http://dx.doi.org/10.1016/j.ecoleng.2011.09.006>
- Dampin, N., Tarnchalanukit, W., Chunkao, K., & Maleewong, M. (2012). Fish growth model for Nile tilapia (*Oreochromis niloticus*) in wastewater oxidation pond, Thailand. *Procedia Environmental Sciences*, 13, 513-524. <http://dx.doi.org/10.1016/j.proenv.2012.01.042>
- Delgado, M., Bigeriego, M., & Guarniola, E. (1993). Uptake of Zn, Cr, and Cd by water hyacinths. *Water Research*, 27(2), 269-272. [http://dx.doi.org/10.1016/0043-1354\(93\)90085-V](http://dx.doi.org/10.1016/0043-1354(93)90085-V)
- Denbigh, K. G., & Raumann, G. (1952). The thermo-osmosis of gas through a membrane. II. Experimental *Proceeding of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 210(1103), 518-533. <http://dx.doi.org/10.1098/rspa.1952.0017>
- Gehlin, S., Hellstrom, G., & Nordell, B. (2003). The influence of the thermosiphon effect on the thermal response test. *Renewable Energy*, 28(14), 2239-2254. [http://dx.doi.org/10.1016/S0960-1481\(03\)00129-0](http://dx.doi.org/10.1016/S0960-1481(03)00129-0)

- Grosse, W. (1989). Thermo-osmosis air transport in aquatic plants affecting growth activities and oxygen diffusion to wet soils. In D. A. Hammer (Ed.), *Constructed Wetland for Wastewater Treatment* (pp.469-476). USA: Lewis Publisher, Inc.
- Grosse, W., & Bauch, C. (1991). Gas transfer in floating-leaved plants. *Plant Ecology*, 97(2), 185-192. <http://dx.doi.org/10.1007/BF00035391>
- Gupta, A. K., & Sinha, S. (2007). Phytoextraction capacity of the plants growing on tannery sludge dumping sites. *Bioresource Technology*, 98(9), 1788-1794. <http://dx.doi.org/10.1016/j.biortech.2006.06.028>
- Hadad, H. R., Maine, M. A., & Bonetto, C. A. (2006). Macrophyte growth in a plot-scale constructed wetland for industrial wastewater treatment. *Chemosphere*, 63(10), 1744-1753. <http://dx.doi.org/10.1016/j.chemosphere.2005.09.014>
- Huang, J., Wang, S. H., Yan, L., & Zhang, Q. S. (2010). Plant photosynthesis and its influence on removal efficiencies in constructed wetlands. *Ecological Engineering*, 3, 1037-1043. <http://dx.doi.org/10.1016/j.ecoleng.2010.04.016>
- Ivanov, Y., Watanabe, H., Hamabe, M., Kawahara, T., Sun, J., & Yamaguchi, S. (2012). Observation of the thermosiphon effect in the circulation of liquid nitrogen in HTS cable cooling system. *Physics Procedia*, 27, 368-371. <http://dx.doi.org/10.1016/j.phpro.2012.03.487>
- Jenssen, P. D., Mahlum, T., & Krogstad, T. (1994). Potential use of constructed wetlands for wastewater treatment in northern environment. *Water Science and Technology*, 28(10), 149-157.
- Juwarkar, A. S., Oke, B., Juwarkar, A., & Patnaik, S. M. (1995). Domestic wastewater treatment through constructed wetland in India. *Water Science and Technology*, 32(3), 291-294. [http://dx.doi.org/10.1016/0273-1223\(95\)00637-0](http://dx.doi.org/10.1016/0273-1223(95)00637-0)
- Li, S., Zhang, K., Zhou, S., Zhang, L., & Chen, Q. (2009). Use of dewatered municipal sludge on Canna growth in pot experiments with a barren clay soil. *Waste Management*, 29(6), 1870-1876. <http://dx.doi.org/10.1016/j.wasman.2008.12.007>
- Maine, M. A., Sune, N., Hadad, H., Sanchez, G., & Bonetto, C. (2006). Nutrient and metal removal in a constructed wetland for wastewater treatment from a metallurgic industry. *Ecological Engineering*, 26(4), 341-347. <http://dx.doi.org/10.1016/j.ecoleng.2005.12.004>
- Marin, J., & Ayele, J. (2003). Removal of some heavy metal cations from aqueous solutions by spruce sawdust. II. Adsorption-desorption through column experiments. *Environmental Technology*, 24(4), 491-502. PubMed ID: 12755450.
- Mevi-Schutz, J., & Grosse, W. (1988). The importance of water vapour for the circulating air flow through *Nelumbo nucifera*. *Journal of Experimental Botany*, 39(9), 1231-1236. <http://dx.doi.org/10.1093/jxb/39.9.1231>
- Mirmov, N. I., & Belyakova, I. G. (1982). Heat liberation during vapor condensation in a thermosiphon. *Journal of Engineering Physics*, 43(3), 970-974. <http://dx.doi.org/10.1007/BF00827237>
- Nivala, J., Wallac, S., Headley, T., Kassa, K., Brix, H., Afferden, M. V., & Müller, R. (2013). Oxygen transfer and consumption in subsurface flow treatment wetlands. *Ecological Engineering*, 61, 544-554. <http://dx.doi.org/10.1016/j.ecoleng.2012.08.028>
- Pattamapitooon, T., Siroroj, P., Pakkomg, P., & Chunkao, K. (2013). Nature of solar radiation as encouraged to produce an increment of dissolved oxygen and hydrogen peroxide in oxidation ponds for community wastewater treatment at H.M. the King's LERD Project site in Phetchabuti Province, Thailand. *Modern Applied Science*, 7(6), 26-41. <http://dx.doi.org/10.5539/mas.v7n6p26>
- Pedescoll, A., Sidrach-Cardona, R., Sánchez, J. C., & Bécáres, E. (2013). Evapotranspiration affecting redox conditions in horizontal constructed wetlands under Mediterranean climate: Influence of plant species. *Ecological Engineering*, 58, 335-343. <http://dx.doi.org/10.1016/j.ecoleng.2013.07.007>
- Phewnil, O., Chunkao, K., Pattamapitooon, T., Intaraksa, A., Chueawong, O., Chantrasoon, C., & Boonprakong, T. (2014). Choosing aquatic plant species for high wastewater treatment efficiency through small wetland. *Modern Applied Science*, 8(4), 187-194. <http://dx.doi.org/10.5539/mas.v8n4p187>
- Poommai, S., Chunkao, C., & Bualert, B. (2012). Variation of oxygen transfer along the rectangular weir crest distance of wastewater treatment pond. *Procedia Environmental Sciences*, 13, 498-512. <http://dx.doi.org/10.1016/j.proenv.2012.01.041>

- Poommai, S., Chunkao, C., Narauchid, D., Boonmang, S., & Nimpee, C. (2013). Determining the in-pipe anaerobic processing distance before draining to oxidation pond of municipal wastewater treatment. *International Journal of Environmental Science and Development*, 4(2), 157-162. <http://dx.doi.org/10.7763/IJESD.2013.V4.326>
- Pulford, I. D., & Watson, C. (2003). Phytoremediation of heavy metal-contaminated land by trees-a review. *Environment International*, 29(4), 529-540. [http://dx.doi.org/10.1016/S0160-4120\(02\)00152-6](http://dx.doi.org/10.1016/S0160-4120(02)00152-6)
- Rai, S., Hasan, S. H., Rupainwar, D. C., & Sharma, Y. C. (1994). Removal of cadmium from wastewater by water hyacinth. *International Journal of Environmental Studies*, 46, 251-256. <http://dx.doi.org/10.1080/00207239408710930>
- Reddy, K. R., Agami, M., & Tucker, J. C. (1990). Influence of phosphorus growth and nutrient storage by water hyacinth (*Eichhornia crassipes* (Mart.) Solms) plants. *Aquatic Botany*, 37(4), 355-365. [http://dx.doi.org/10.1016/0304-3770\(90\)90021-C](http://dx.doi.org/10.1016/0304-3770(90)90021-C)
- Semvimol, N., Chunkao, K., & Bualert, S. (2014). Gas extraction from sludge as acquired from oxidation ponds of community wastewater and cassava-factory wastewater treatment through nature-by-nature processes. *Modern Applied Science*, 8(2), 171-183. <http://dx.doi.org/10.5339/mas.v8n2p171>
- Sooknah, R. D., & Wilkie, A. C. (2004). Nutrient removal by floating aquatic macrophytes cultured in anaerobically digested flushed dairy manure wastewater. *Ecological Engineering*, 22(1), 27-42. <http://dx.doi.org/10.1016/j.ecoleng.2004.01.004>
- Srivastava, R. C., & Avasthi, P. K. (1975). Non-equilibrium thermodynamics of thermo-osmosis of water through kaolinite. *Journal of Hydrology*, 24(1-2), 111-120. [http://dx.doi.org/10.1016/0022-1694\(75\)90145-6](http://dx.doi.org/10.1016/0022-1694(75)90145-6)
- Stefanakis, A. I., & Tsihrintzis, V. A. (2012). Effects of loading, resting period, temperature, porous media, vegetation and aeration on performance of pilot-scale vertical flow constructed wetlands. *Chemical Engineering Journal*, 181-182, 416-430. <http://dx.doi.org/10.1016/j.cej.2011.11.108>
- Stottmeister, U., Wiessner, A., Kusch, P., Kappelmeyer, U., Kastner, M., Bederski, O., Muller, R. A., & Moormann, H. (2003). Effects of plants and microorganisms in constructed wetlands for wastewater treatment. *Biotechnology Advances*, 22, 93-117. <http://dx.doi.org/10.1016/j.biotechadv.2003.08.010>
- Supakata, N., & Chunkao, K. (2011). Thickness of moist sludge piling from community wastewater treatment through the Royal LERD technology for growing rice. *Journal of Agricultural Science*, 3(3), 93-100. <http://dx.doi.org/10.5539/jas.v3p93>
- Tanji, Y., Sakai, R., Miyanaga, K., & Unno, H. (2006). Estimation of the self-purification capacity of biofilm formed in domestic sewer pipes. *Biochemical Engineering Journal*, 31(1), 96-101. <http://dx.doi.org/10.1016/j.bej.2006.05.021>
- Tateyama, K., Egawa, H., & Yamamoto, H. (1977). Don't waste water- weeds. *Weed Resource Japan*, 22, 151-156.
- Thaipichitburapa, P., Meksumpun, C., & Meksumpun, S. (2010). Province-based self-remediation efficiency of the Tha Chin river basin, Thailand. *Water Science and Technology*, 62, 594-602. <http://dx.doi.org/10.2166/wst.2010.293>
- Tripathi, B. D., & Shukla, S. C. (1991). Biological treatment of wastewater by selected aquatic plants. *Environmental Pollution*, 69, 69-78.
- Wahla, I. H., & Kirkham, M. B. (2008). Heavy metal displacement in salt-water-irrigated soil during phytoremediation. *Environmental Pollution*, 155(2), 271-283. <http://dx.doi.org/10.1016/j.envpol.2007.11.020>
- Xia, H., & Ma, X. (2006). Phytoremediation of ethion by water hyacinth from water. *Bioresource Technology*, 97, 1050-1054. <http://dx.doi.org/10.1016/j.biotech.2005.04.039>
- Yang, Q., Tam, N. F., Wong, Y. S., Laun, T. G., Su, W. S., Lan, C. Y., Shin, P. K., & Cheung, S. G. (2008). Potential use of mangrove as constructed wetland for municipal sewage treatment in Futian, Shenzhen, China. *Marine Pollution Bulletin*, 57, 735-743. <http://dx.doi.org/10.1016/j.ecoleng.2009.05.017>
- Zaier, H., Ghnaya, T., Rejeb, B. K., Lakhdar, A., Rejeb, S., & Jemal, F. (2010). Effects of EDTA on phytoextraction of heavy metals (Zn, Mn, and Pb) from sludge-amended soil with *Brassica napus*. *Bioresource Technology*, 101(11), 3978-3983.
- Zhu, Y. L., Zayed, A. M., Qian, J. H., Souza, M., & Terry, N. (1999). Phytoaccumulation of trace elements by

wetland plants: II. Water Hyacinth. *Journal of Environmental Quality*, 28(1), 339-344.
<http://dx.doi.org/10.2134/jeq1999.00472425002800010042x>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

ADTC in Wind Turbines with Partial State and wind Speed Estimation

Ali Oudah¹, Izzeldin I. Mohd² & A. Hameed³

^{1,3} Faculty of Manufacturing Engineering, Universiti Malaysia Pahang, Pahang, Malaysia

² Faculty of Electrical & Electronics Engineering, Universiti Malaysia Pahang, Pahang, Malaysia

Correspondence: Ali Oudah, Faculty of Manufacturing Engineering, Universiti Malaysia Pahang, Pahang, Malaysia. E-mail: mark6ge@gmail.com

Received: June 13, 2014

Accepted: June 17, 2014

Online Published: September 26, 2014

doi:10.5539/mas.v8n5p247

URL: <http://dx.doi.org/10.5539/mas.v8n5p247>

The research is financed by (Universiti Malaysia Pahang)

Abstract

In wind turbines, although the construction, implementation, and analysis of Adaptive Disturbance Tracking Controller is simple, the main problem of using ADTC theory is the requirement of wind speed information. The idea of estimating the wind speed from rotor speed is discussed in another works. It is also observed that an accurate model of the wind turbine is needed to use the simple estimator due to a restriction in ADTC theory. If an accurate model of the wind turbine is available, linear control theories can be used to design the controllers since these theories are well established. The idea of using an adaptive controller actually came from the fact that an accurate model of the wind turbine is not available. In this paper the modification of the ADTC theory is discussed to incorporate the very simple reduced order wind turbine model to estimate the wind speed.

Keywords: wind energy, turbines, controllers, ADTC theory, speed estimation

1. Introduction

The idea behind the simple wind speed estimator is to use the disturbance generator model to design a wind disturbance estimator. A method to estimate the wind speed using rotor speed is proposed. Also, it is observed that the simple wind speed estimator alone does not full the conditions required for the implementation of the adaptive controller and hence the stability of controller with simple wind speed estimator cannot be guaranteed. The output feedback is used to make the plant look like minimum-phase to the adaptive controller. The wind turbine is assumed to be a linear, time - invariant, finite dimensional plant represented as:

$$\begin{cases} \dot{x} = A_p x_p + B_p u_p + \Gamma_p^1 u_D \\ y_p = C_p x_p; x_p(0) = x_0 \end{cases} \quad (1a)$$

And the reduced order model of wind turbine is expressed as:

$$\begin{cases} \dot{x} = A_m x_m + B_m u_p + \Gamma_m^1 u_D \\ y_m = C_m x_m; x_m(0) = x_0 \end{cases} \quad (1b)$$

Where, x_p is an N_p - dimensional plant state vector, u_p is M - dimensional the control input vector, y_p is P - dimensional the sensor output vector. x_m is m dimensional ($m < n$) lower order plant model state vector, y_m is the P_m dimensional model output vector. A_p , B_p , C_p are the state, input, output and disturbance matrix of plant with appropriate dimensions. A_m , B_m , C_m are the state, input, output and disturbance matrix of the lower order plant model with appropriate dimensions. u_D is an M_D - dimensional disturbance input vector, and will be assumed to come from the Disturbance Generator:

$$\begin{cases} \dot{u}_D = \theta z_D \\ \dot{z} = F z_D; z_D(0) = z_0 \end{cases} \quad (2)$$

Where, z_D is N_D - dimensional the disturbance state.

All matrices in equations (1b) and (2) have the appropriate compatible dimensions. Equation (2) can be rewritten in a form that is not a dynamical system, which is sometimes easier to use:

$$\begin{cases} u_D = \theta z_D \\ z_D = L\varphi_D \end{cases} \quad (3)$$

where, φ_D is a vector composed of the known basis functions for the solution of $u_D = \theta z_D$, i.e., φ_D are the basis functions that make up the known form of the disturbance, and L is a matrix of appropriate dimension. The method for tracking persistent disturbances used in this paper requires only the knowledge of the form of the disturbance, the amplitude of the disturbance does not need to be known, i.e. (L, θ) are unknown. In this chapter, a step disturbances of unknown amplitude that can be represented in the form of equation (3) as $\varphi_D = 1$ is considered, with (L, θ) unknown.

Now combining equations 1a and 1b we get a new augmented plant model:

$$\begin{cases} \dot{x}_m \\ \dot{z}_D \end{cases} = \begin{Bmatrix} A_m & \Gamma_m \theta \\ F & 0 \end{Bmatrix} \begin{Bmatrix} x_m \\ z_D \end{Bmatrix} + \begin{Bmatrix} B_m \\ 0 \end{Bmatrix} u \quad (4)$$

In equation (4) we augmented the lower order plant model with the model of disturbance generator. Using the augmented plant model in equation (4), a state estimator can be designed as:

$$\begin{cases} \dot{\hat{x}}_m \\ \dot{\hat{z}}_D \end{cases} = \begin{Bmatrix} A_m & \Gamma_m \theta \\ F & 0 \end{Bmatrix} \begin{cases} \hat{x}_m \\ \hat{z}_D \end{cases} + \begin{Bmatrix} B_m \\ 0 \end{Bmatrix} u + \begin{Bmatrix} K_x \\ K_D \end{Bmatrix} (y_p - \hat{y}_m) \quad (5)$$

The estimator equation (5) can also be broken down into wind disturbance estimator and state estimator as:

$$\hat{\hat{x}} = A_m \hat{x} + B_m u + \Gamma_m^1 \theta \hat{z} + K_x (y_p - \hat{y}) \quad (6)$$

$$\hat{\hat{z}} = F \hat{z} + K_D (y_p - \hat{y})$$

Using the wind disturbance state estimation, the estimated wind speed can be expressed as:

$$\hat{u} = \theta_D \quad (7)$$

Our control objective will be to cause the output of the plant (y_p) to asymptotically track some linear function of estimated disturbances of the form given by the disturbance estimator. We define the estimated output tracking error as:

$$\hat{e} \equiv y_p - Q \hat{u}_D \quad (8)$$

To achieve the desired control objective, we want $\hat{e}_y \xrightarrow{t \rightarrow \infty} 0$. Consider the plant given by equation (1a) with the disturbance generator given by equation (2) and respective disturbance and state estimator given by equation (5) and equation (6). Our control objective for this system will be accomplished by an Adaptive Control Law of the form:

$$u_p = G_e \hat{e}_y + G_D \varphi_D + G_x \hat{x}_m \quad (9)$$

Where G_e and G_D are adaptive gain matrices of the appropriate compatible dimensions, and G_x is the state feedback gain matrix. Now we specify the gain adaption laws, which will produce asymptotic tracking:

$$\begin{cases} \dot{G}_e = -\hat{e}_y \hat{e}_y^T \gamma_e \\ \dot{G}_D = -\hat{e}_y \varphi_D^T \gamma_D \end{cases} \quad (10)$$

Where γ_e and γ_D are arbitrary positive definite matrices. Our Adaptive Controller is specified by equation (9) with the above adaptive gain laws represented by equation (10).

2. Developing the Simple Reduced Order Wind Turbine Model

For the implementation of the state estimator, we require a nominal model of wind turbine. Commercially available software such as BLADED (gl-garradhassan.com) can be used to develop a complex model of wind turbine with many degrees of freedom. Also, a FORTRAN code, called FAST (Fatigue, Aerodynamics, Structure, and Turbulence), is available from National Wind Technology Center (NWTC) to model a wind turbine of

sufficient complexity (wind.nrel.gov). Although, these modeling tools have been used to develop a model of wind turbine to some accuracy, it is not always possible to use these tools to model an arbitrary turbine. The idea of Adaptive Disturbance Tracking Control was first motivated by the use of the least possible detailed information from the wind turbine. To further emphasize this idea, a simple and reduced order model of the wind turbine is developed that includes the drivetrain model of a wind turbine from available data. The drive - train model is the simplest model that can be used for the control development in Region II operation of wind turbine. A simple wind turbine model having drive - train mode is developed based on (A. D. Wright), (P. Novak, T. Ekelund, I. Jovik, and B. Schmidtbauer). The aerodynamic torque on a wind turbine is given by:

$$Q_A = \frac{1}{2} \rho A C_P w^2 \quad (11)$$

Let Q_G be the generator torque applied, Ω be the rotor speed and I_{tot} be the total rotational inertia of the turbine, including rotational inertia of generator, rotor, high - speed shaft, low - speed shaft, and gearbox. The generator speed is related to the rotor speed by gearbox ratio (GBR). Since, the power coefficient depends on blade pitch (β) and (λ), the TSR further depends on wind speed (w) and rotor speed (Ω), we can express the aerodynamic torque as a continuous function of wind speed, blade pitch angle and generator speed for simplicity as:

$$Q_A = Q_A(w, \beta, \Omega) \quad (12)$$

now, equation (12) can be extended using Taylors series expansion about its operating point as:

$$Q_A = Q_A(w_0, \beta_0, \Omega_0) + \frac{\partial Q_A}{\partial w} \partial w + \frac{\partial Q_A}{\partial \Omega} \partial \Omega + \frac{\partial Q_A}{\partial \beta} \partial \beta + H.O.T. \quad (13)$$

Where (w_0, β_0, Ω_0) are the operating point wind speed, blade pitch, and rotor speed respectively. ∂w , $\partial \Omega$, and $\partial \beta$ are the perturbation in wind speed, rotor speed and blade pitch respectively, and H.O.T. Are the higher order terms of the expansion.

Now the perturbation in aerodynamic torque can be expressed as:

$$\partial Q_A = Q_A(w, \beta, \Omega) - Q_A(w_0, \beta_0, \Omega_0) \quad (14)$$

In Region II the blade pitch is kept constant so we can omit this variable in equation (13). From equations (13) and (14), we can get:

$$\partial Q_A = \frac{\partial Q_A}{\partial w} \partial w + \frac{\partial Q_A}{\partial \Omega} \partial \Omega \quad (15)$$

in equation (15) we omit the $H.O.T$. Because we are interested in simplified linear model of wind turbine. In region II, the generator torque is used as control input. Based on this information, the equation of rotational motion of wind turbine can be expressed as:

$$I_{tot} \dot{\omega} = Q_A - Q_G \quad (16)$$

assuming generator rotational acceleration as zero at equilibrium, the perturbation of equation (15) can be expressed as:

$$I_{tot} \partial \dot{\Omega} = Q_A(w_0, \beta_0, \Omega_0) + \partial Q_A - (Q_{G0} + \partial Q_G) \quad (17)$$

where Q_{G0} is the operating point generator torque at operating point generator speed. In equilibrium, the operating point aerodynamic torque must balance the operating point generator torque to prevent the generator acceleration. So, equation (17) can be expressed as:

$$I_{tot} \partial \dot{\Omega} = \partial Q_A - \partial Q_G \partial \dot{\Omega} = \frac{\partial Q_A}{I_{tot}} \partial w \frac{\partial Q_A}{\partial w} + \frac{\partial Q_A}{I_{tot}} \partial \Omega - \frac{1}{I_{tot}} \partial Q_G \frac{\partial Q_A}{\partial \Omega} \quad (18)$$

Now assuming perturbed rotor speed as the state variable, perturbed wind speed and perturbed generator torque as inputs, and perturbed rotor speed as output, the state space representation of equation 18 can be written as:

$$\begin{cases} \dot{x} = A_m x_m + B_m u_p + \Gamma_m^1 u_D \\ y_m = C_m x_m; x_m(0) = x_0 \end{cases} \quad (19)$$

where, $x_m = \Omega_m$, $A_m = \frac{\partial Q_A}{I_{tot} \partial \Omega}$, $B_m = -\frac{1}{I_{tot}}$, and $\Gamma_m = \frac{\partial Q_A}{I_{tot} \partial w}$.

3. Stability and Convergence of ADTC with State Estimation and State Feedback

3.1 Problem formulation

Assume a linear, time invariant and finite dimensional plant of equation (1a) and disturbance generator of equation (2).

We define the estimator as:

$$\begin{cases} \dot{\hat{z}} = L_{21}y + L_{22}\hat{z} + L_{23}u \\ \text{where } \hat{z} \equiv \underbrace{z}_{TW} + e\check{z} \end{cases} \quad (20)$$

Where,

$$w \equiv \begin{Bmatrix} x \\ z_D \end{Bmatrix}$$

Now taking reference from equation (4), the state space equation for w can be written as:

$$\begin{cases} \dot{w} = \underbrace{\begin{bmatrix} A & \Gamma\theta \\ 0 & F \end{bmatrix}}_A w + \underbrace{\begin{bmatrix} B \\ 0 \end{bmatrix}}_B u \\ y = \begin{bmatrix} C & 0 \end{bmatrix} w \\ e_y = \underbrace{\begin{bmatrix} C & -Q\theta \end{bmatrix}}_C w \end{cases} \quad (21)$$

Define the ideal trajectories as:

$$\begin{cases} \dot{w}_* = \bar{A}w_* + \bar{B}u_* \\ y_* = [C \ 0]w_* \\ e_y^* = \bar{C}w_* = 0 \end{cases} \quad (22)$$

With,

$$\begin{cases} w_* \equiv \begin{bmatrix} x_* \\ z_D \end{bmatrix} = \begin{bmatrix} S_1 \\ I_D \end{bmatrix} Z_D \\ u_* \equiv S_2 z_D + G_z^* z_*; z_* \equiv Tw_* = T\bar{S}Z_D \\ \Rightarrow u_* \equiv \underbrace{(s_2 + G_z^* T \bar{S}_1)}_{S_2} z_D + G_z^* z_* \end{cases}$$

Let,

$$\begin{cases} \Delta w \equiv w - w_* = \text{Matrix} \\ \Delta u \equiv u - u_* \\ \Delta y \equiv y - y_* \\ \Delta e_y \equiv e_y - e_y^* = e_y \\ \Delta z \equiv z - z_* = Tw - Tw_* = T \Delta w \end{cases}$$

which combined with equation (21) results,

$$\begin{cases} \Delta w \equiv w - w_* = \text{Matrix} \\ \Delta u \equiv u - u_* \\ \Delta y \equiv y - y_* \\ \Delta e_y \equiv e_y - e_y^* = e_y \\ \Delta z \equiv z - z_* = Tw - Tw_* = T \Delta w \end{cases} \quad (23)$$

Now assume,

$$z_D = H_D z \text{ or } H_D T = [0 \ I_D]$$

This implies

$$\hat{z} = H_D \hat{z} = H_D T w + H_D e \quad (24)$$

$$\therefore \hat{z} = z_D + H_D e$$

Define:

$$\begin{cases} \hat{u}_D \equiv \theta \hat{z}_D \\ \hat{e}_y \equiv y - Q \hat{u}_D \end{cases}$$

Substituting values of u_D and z_D we obtain:

$$\begin{aligned} \hat{e}_y &\equiv y - Q \theta \hat{z} &&= e_y - Q \theta H_D e \\ &= \underbrace{y - Q \theta z_D}_{e_y} - Q \theta H_D e &&= (e_y - e_y^*) - Q \theta H_D e \\ &&&= \Delta e_y - Q \theta H_D e \\ &&&\Rightarrow \hat{e} = \bar{C} \Delta w - Q \theta H_D e \end{aligned} \quad (25)$$

Adaptive control law is:

$$u = G_e \hat{e} + G_D \phi_D + G_z \hat{z}$$

and we defined:

$$\Delta u \equiv u - u_*$$

substituting values of u and u_* we get

$$\begin{aligned} &= G_e^* \hat{e}_y + \underbrace{(G_D^* - S_2 L_D)}_{=0} \varphi_D + G_z^* \hat{z} - G_z^* z_* \end{aligned}$$

$$\begin{aligned} \Delta u &\equiv u - u_* = u - \bar{S}_2 z_D + \underbrace{\left[\begin{matrix} \Delta G_e & \Delta G_D & \Delta G_z \end{matrix} \right]}_{\Delta G} \underbrace{\begin{bmatrix} \hat{e}_y \\ \varphi_D \\ \hat{z} \end{bmatrix}} \\ &= u - (S_2 z_D + G_z^* z_*) \\ &= u - S_2 L_D \phi_D - G_z^* z_* \end{aligned} \quad (26)$$

$$= G_e^* (\bar{C} \Delta w - Q \theta H_D e) + G_z^* (T \Delta w + e) + \Delta \underbrace{\frac{G \eta}{h}}$$

$$= (G_e^* \bar{C} + G_z^* T) \Delta w + (G_z^* - G_e^* Q \theta H_D) e + h$$

Estimator error is expressed as:

$$\begin{aligned} \dot{e} &= \hat{z} - T\dot{w} \\ &= [L_{21} \overset{y}{[C \ 0]} \omega + L_{22} \overset{\hat{z}}{T\omega + e} + L_{23} u] - T[\bar{A}w + \bar{B}u] \\ &= (L_{21}[C \ 0] + L_{22}T - T\bar{A})w + L_{22}e + (L_{23} - T\bar{B})u \end{aligned} \tag{27}$$

In equation (27) choose $L_{23} \equiv T\bar{B}$ where $\exists T \ni T\bar{A} - L_{22}T = L_{21}[C \ 0]$ (uniquely solvable when $\sigma(\bar{A}) \cap \sigma(L_{22}) = \emptyset$) with L_{22} chosen appropriately stable. Then,

$$\dot{e} = L_{22}e \tag{28}$$

Let,

$$\begin{aligned} \text{Let } T &\equiv [T_1 \ T_2] \text{ then } H_D T \\ &= [0 \ I_D] \Rightarrow H_D T_1 = 0 \text{ and } H_D T_2 = I_D. \end{aligned}$$

Also,

$$\begin{aligned} L_{23} = T\bar{B} &= T_1 \bar{B} \text{ and } T \ni T\bar{A} - L_{22}T \\ &= L_{21}[C \ 0] \Rightarrow \begin{cases} T_1 \bar{A} - L_{22}T_1 = L_{21}C \\ T_2 \bar{A} - L_{22}T_2 + T_1 \Gamma \theta = 0 \end{cases} \end{aligned}$$

Closed loop analysis

Combining equations 21 and 28 we get:

$$\begin{cases} \begin{bmatrix} \Delta \dot{w} \\ \dot{e} \end{bmatrix} = \underbrace{\begin{bmatrix} \bar{A} + \bar{B}(G_z^* \bar{C} + G_z^* T) & \bar{B}(G_z^* - G_e^* Q \theta H_D) \\ 0 & L_{22} \end{bmatrix}}_{\bar{A}_C} \begin{bmatrix} \Delta w \\ e \end{bmatrix} + \underbrace{\begin{bmatrix} \bar{B} \\ 0 \end{bmatrix}}_{\bar{B}} h \\ \dot{e}_y = \underbrace{\begin{bmatrix} C & -Q \theta H_D \end{bmatrix}}_{\bar{C}} \begin{bmatrix} \Delta w \\ e \end{bmatrix} \end{cases} \tag{29}$$

where,

$$\bar{A}_C = \underbrace{\begin{bmatrix} \bar{A} + \bar{B}G_z^* T & \bar{B}G_z^* \\ 0 & L_{22} \end{bmatrix}}_{\bar{A}} + \underbrace{\begin{bmatrix} \bar{B} \\ 0 \end{bmatrix}}_{\bar{B}} G_e^* \underbrace{\begin{bmatrix} C & -Q \theta H_D \end{bmatrix}}_{\bar{C}}$$

Lemma 1

$(\bar{A}, \bar{B}, \bar{C})$ is ASPR $\Leftrightarrow (A, B, C)$ is ASPR

Proof of Lemma

from the expression of $\tilde{C}\bar{B}$, $\bar{C}\bar{B}$, and $\bar{C}B$ we have $\tilde{C}\bar{B} = \bar{C}\bar{B} = \bar{C}B$ now,

$$\begin{aligned} \tilde{P}(s) &\equiv \tilde{C}(sI - \bar{A})^{-1} \bar{B} \\ &= \underbrace{\begin{bmatrix} \bar{C} & -Q \theta H_D \end{bmatrix}}_{\bar{C}} \left(sI - \underbrace{\begin{bmatrix} \bar{A} + \bar{B}G_z^* T & \bar{B}G_z^* \\ 0 & L_{22} \end{bmatrix}}_{\bar{A}} \right)^{-1} \underbrace{\begin{bmatrix} \bar{B} \\ 0 \end{bmatrix}}_{\bar{B}} \end{aligned}$$

$$\begin{aligned}
 &= C(sI - (A + BG_z^*T))^{-1}B \\
 &= \underbrace{\begin{bmatrix} C & -Q\theta \end{bmatrix}}_{\tilde{C}} (sI - \begin{bmatrix} A + BG_z^*T_1 & \Gamma\theta + BG_z^*T_2 \\ 0 & F \end{bmatrix}) \cdot \underbrace{\begin{bmatrix} B \\ 0 \end{bmatrix}}_{\tilde{B}} \\
 &= \underbrace{\begin{bmatrix} C & -Q\theta \end{bmatrix}}_{\tilde{C}} \left(\begin{bmatrix} (sI - (A + BG_z^*T_1))^{-1} & W_{12} \\ 0 & (sI - F)^{-1} \end{bmatrix} \right) \underbrace{\begin{bmatrix} B \\ 0 \end{bmatrix}}_{\tilde{B}} \\
 &\tilde{P}(s) = C(sI - (A + BG_z^*T_1))^{-1}B \equiv P_1(s) \tag{30}
 \end{aligned}$$

using the expression from equation 30 we can use the relation: $(A + BG_z^*T_1, B, C)$ is ASPR $\Leftrightarrow CB > 0$ and $P_1(s)$ minimum phase.

To use this relation we have to prove that $P(s)$ minimum phase $\Leftrightarrow P_1(s)$ minimum phase. In another word, State Feedback does not change minimum phase of a transfer function. This is proved in following lemma.

Lemma 2

State Feedback does not change the minimum phase nature of the transfer function.

Proof of Lemma

Assume a square transfer function. The transmission zeros of (A, B, C) are

$$\begin{aligned}
 \lambda_* \ni \text{rank} \begin{Bmatrix} A - \lambda_*I & B \\ C & 0 \end{Bmatrix} < N + M \\
 \text{rank} \begin{Bmatrix} A - \lambda_*I & B \\ C & 0 \end{Bmatrix} &= \text{rank} \begin{pmatrix} A - \lambda_*I & B \\ C & 0 \end{pmatrix} \underbrace{\begin{pmatrix} I & 0 \\ G & I \end{pmatrix}}_{\text{Nonsingular}} \\
 &= \text{rank} \begin{Bmatrix} A + BG - \lambda_*I & B \\ C & 0 \end{Bmatrix} \\
 &\Rightarrow \text{Transmission zeros of } (A + BG, B, C)
 \end{aligned}$$

Are the same as those of (A, B, C) #

Now using Lemma 1 and Lemma.2 , $CB > 0$ and $P(s)$ minimum phase will guarantee that $\exists G_e^* \ni (\tilde{A} + \tilde{B}G_e^*\tilde{C}, \tilde{B}, \tilde{C})$ is SPR consequently, we have the following theorem:

4. Solvability of the Matrix Matching Conditions

We have the ideal trajectories:

$$\begin{cases} \dot{w} = \bar{A}w_* + \bar{B}u_* \\ e_y^* = \bar{C}w_* = 0 \end{cases} \tag{31}$$

with

$$\begin{cases} w_* \equiv \begin{bmatrix} x_* \\ z_D \end{bmatrix} = \underbrace{\begin{bmatrix} S_1 \\ I_D \end{bmatrix}}_{S_1} z_D \\ u_* \equiv S_2 z_D + G_z^*T\bar{S}_1 z_D = \underbrace{(S_2 + G_z^*T\bar{S}_1)}_{S_2} z_D \end{cases} \tag{32}$$

from equations (31) and (32) we get the following matrix matching conditions

$$\Leftrightarrow \begin{cases} \bar{S}_1 F = \bar{A}\bar{S}_1 + \bar{B}\bar{S}_2 = (\bar{A} + \bar{B}G_z^*T)\bar{S}_1 + \bar{B}S_2 \\ \bar{C}\bar{S} = 0 \end{cases} \tag{33a}$$

$$\Leftrightarrow \begin{cases} S_1 F = (A + BG_z^*)S_1 + BS_2 + \Gamma^1 \theta \\ CS_1 = Q\theta \end{cases} \quad (33b)$$

These Matrix Matching Equations can be uniquely solved for when CB is nonsingular (which it is when $B > 0$).

New Adaptive Controller:

$$\begin{cases} u = G_e \hat{e} + G_z \hat{z} + G_D \phi_D \\ \dot{\hat{z}} = L_{21} \hat{e}_y + L_{22} \hat{z} + L_{23} u \end{cases} \quad (34)$$

With

$$\begin{cases} \dot{G} = -\hat{e} \hat{e}^T \gamma_e; \gamma_e > 0 \\ \dot{G} = -\hat{e} \phi_D^T \gamma_D; \gamma_D > 0 \\ \dot{G} = -\hat{e} \hat{z}^T \gamma_z; \gamma_z > 0 \end{cases}$$

From equation 27 we have,

$$\begin{aligned} \hat{e} &\equiv y - Q\theta H_D \hat{z} \Rightarrow \Lambda z \\ &= L_{21} \hat{e} + L_{22} \hat{z} + \\ &= L_{21} y + \underbrace{(L_{22} - L_{21} Q\theta H_D)}_{\bar{L}_{22}} \hat{z} + L_{23} u \end{aligned} \quad (35)$$

Then

$$\begin{aligned} \hat{e}_y &= \hat{z} - T\bar{w} \begin{bmatrix} L_{21} y + \bar{L}_{22} \hat{z} + L_{23} u \\ [C\ 0]_w \end{bmatrix} - T[\bar{A}w + \bar{B}u] \\ &= (L_{21}[C\ 0] + \bar{L}_{22}T - T\bar{A})w + \bar{L}_{22}e = (L_{23} - T\bar{B})u \end{aligned} \quad (36)$$

Choosing $L_{23} \equiv T\bar{B}$ where $T \ni T\bar{A} - \bar{L}_{22}T = L_{21}[C\ 0]$ (uniquely solvable when $\sigma(\bar{A}) \cap \sigma(\bar{L}_{22}) = \emptyset$).

With \bar{L}_{22} chosen appropriately stable $\Rightarrow \dot{e} = \bar{L}_{22}e$

Also,

$$\begin{aligned} \Delta u &\equiv u - u_* = u - \bar{S}z_D \\ &= u - (S_2 z_D + G_z^* z_*) \\ &= u - S_2 L_D \phi_D - G_z^* z_* \\ &= G_e^* \hat{e}_y + \underbrace{(G_D^* - S_2 L_D)}_{=0} \phi_D + G_z^* \hat{z} - G_z^* z_* \\ &\quad + \underbrace{\begin{bmatrix} \Delta G_e & \Delta G_D & \Delta G_z \end{bmatrix}}_{\Delta G} \underbrace{\begin{bmatrix} \hat{e}_y \\ \phi_D \\ \hat{z} \end{bmatrix}}_{\eta} \end{aligned}$$

$$\begin{aligned}
 &= G_e^* e y + G_z^* z \cdot \Delta \hat{z} + \underbrace{\Delta G \eta}_h; \Delta \hat{z} \equiv \hat{z} - z_* \\
 &= G_e^* (\bar{C} \Delta w - Q \theta H_D e) + G_z^* (T \Delta w + e) + h \\
 &= (G_e^* \bar{C} + G_z^* T) \Delta w + (G_z^* - Q \theta H_D) e + h
 \end{aligned} \tag{37}$$

Since the expression for Δu is same in equations (2)6 and (37) are same, the adaptive controller also remains same in both conditions. In another word we can use the adaptive control law represented by equation (9) with gain adaption law of equation (10).

5. Controller Implementation

After The theory of adaptive disturbance tracking control with state estimation and state feedback is implemented on the National Renewable Energy Laboratory (NREL)’s 5 MW onshore wind turbine model that was used in previous sections. Since a nominal reduced order model is needed to implement the controller as discussed in previous sections, the wind turbine model is linearized with generator *DOF* (Degree of Freedom) only to get a one state turbine model. The one state linearized model is augmented with the disturbance generator model as in equation (2) to get the augmented model as in equation (4). Based on this augmented model an state estimator is designed using equation (5). This state estimator estimates the generator speed and wind speed. The schematic for implementation of adaptive disturbance tracking controller with state and wind speed estimator and state feedback is in figure (1). Figure (2) shows the Simulink implementation of state and wind speed estimation, and state

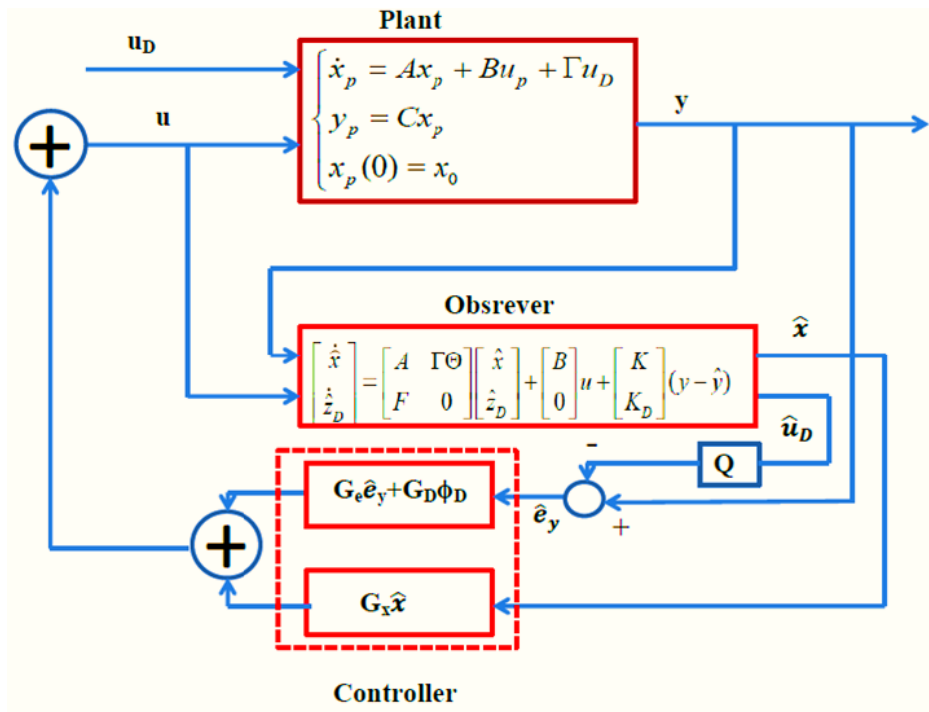


Figure 1. Figure Schematic for ADTC implementation with nominal plant model feedback and adaptive controller implementation

The values of γ_e and γ_D in equation 10 are set to 0.0001 and 15000 respectively by trial and error method to get the best simulation results. Both the step wind and turbulent wind profiles lying well under Region II operation region is used to evaluate the performance of the controller. The performance of adaptive controller is compared with the existing fixed gain baseline controller.

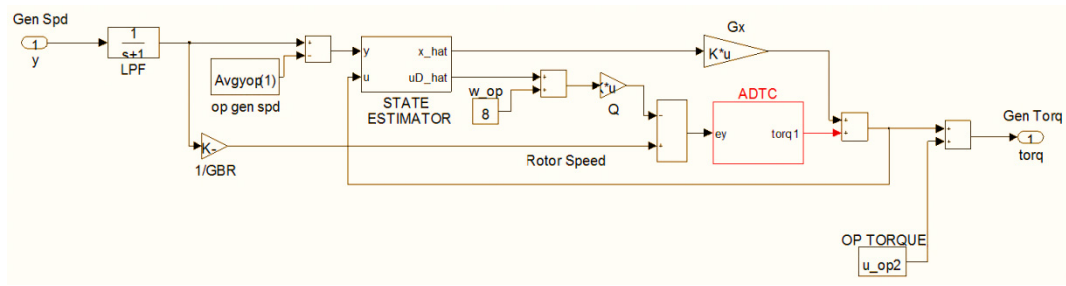


Figure 2. Simulink implementation of controller

6. Step Wind Profile

A step wind profile varying from 6 m/s to 10 m/s is used for the simulation which is well within Region II operation for the wind turbine under study. The step wind profile, estimated

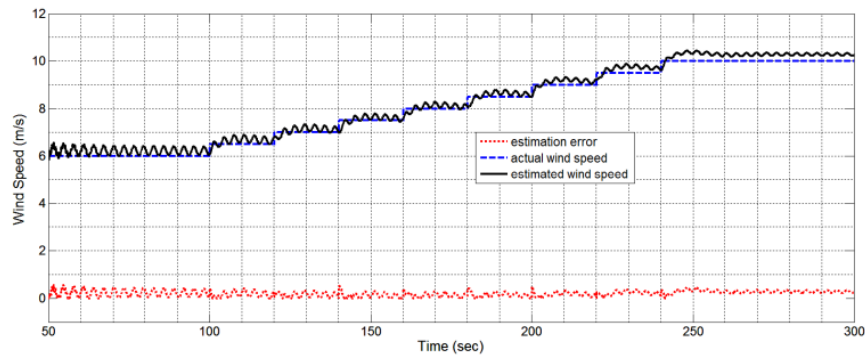


Figure 3. Step wind speed and estimated wind speed with estimation error

wind speed profile and estimation error is in figure (3). From the figure it can be observed the the state estimator does decent job in estimating the wind speed. Also, the estimation error is nearly zero.

In Region II control, *TSR* tracking is the most important controller function to ensure the maximum power capture. The *TSR* for step wind profile is in figure (4). For this particular wind turbine model, the optimum *TSR* is 7.55. From the figure we can observe that the adaptive controller has better *TSR* tracking compared to the baseline controller. The components of estimated output tracking error in equation (27) are in figure (5). In this case the value of tracking ratio (*Q*) is 1.12. The figure shows that the adaptive controller with state estimator and state feedback is performing decently to track the estimated disturbance.

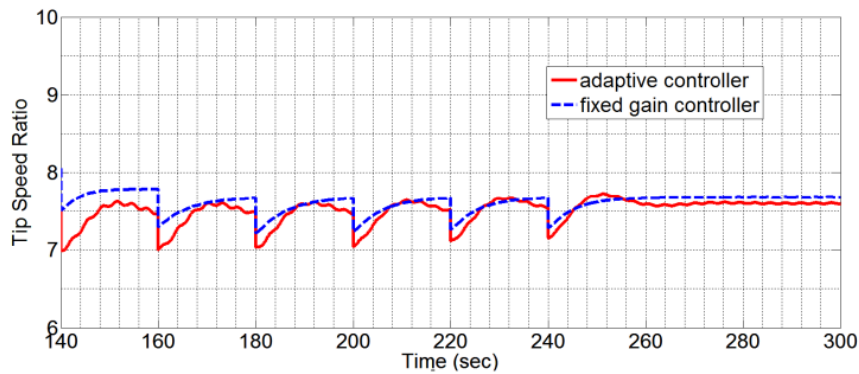


Figure 4. Tip Speed Ratio for step wind

The estimated tracking error of equation (27) is in figure (6). It is observed that the

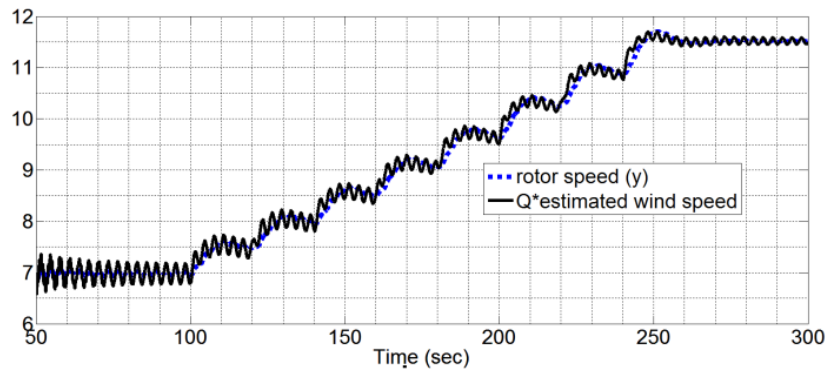


Figure 5. Rotor speed and Q *estimated wind speed

Tracking error is almost zero as desired. The amount of power captured in Region II operation is compared in figure (7).

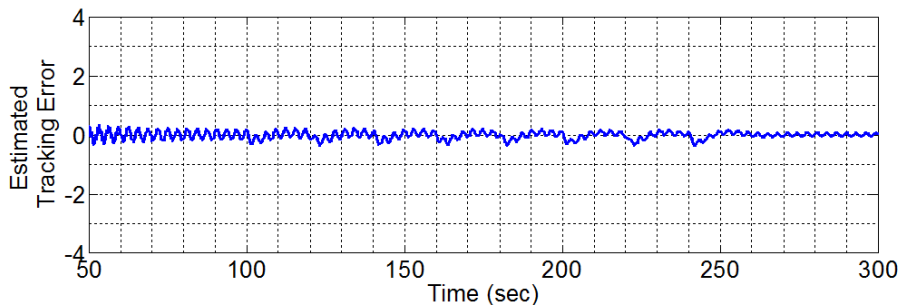


Figure 6. Estimated tracking error

It is difficult to compare during the transitioning period, the adaptive controller capturing more power during steady state. The plot of generator torque is in figure (8). The torque command for adaptive controller is more aggressive compared to the baseline controller. This aggressiveness is attributed to the tendency of adaptive controller to track the step wind profile. The generator speed is compared in figure (9). There is some discrepancy in generator speed in lower speed region; this is because different approaches taken by the controller to track the rotor speed. In case of adaptive controller, the wind speed is directly tracked where this is tracking in indirect in baseline controller. Also, the baseline controller uses another region called Region 1.5 in lower wind speed conditions whereas adaptive controller does not assume this region.

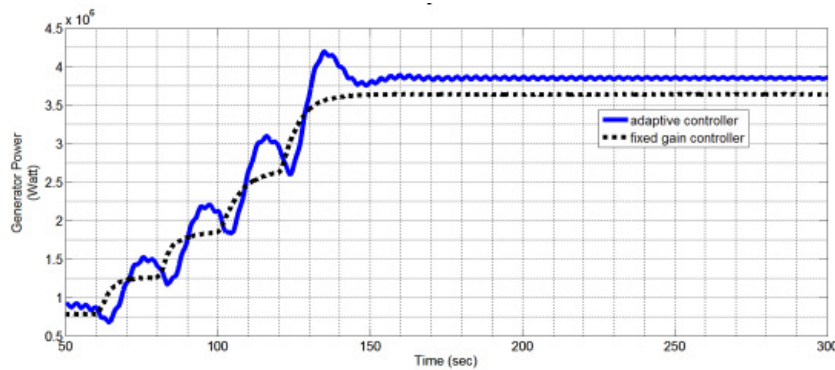


Figure 7. Generator power for Region II step wind

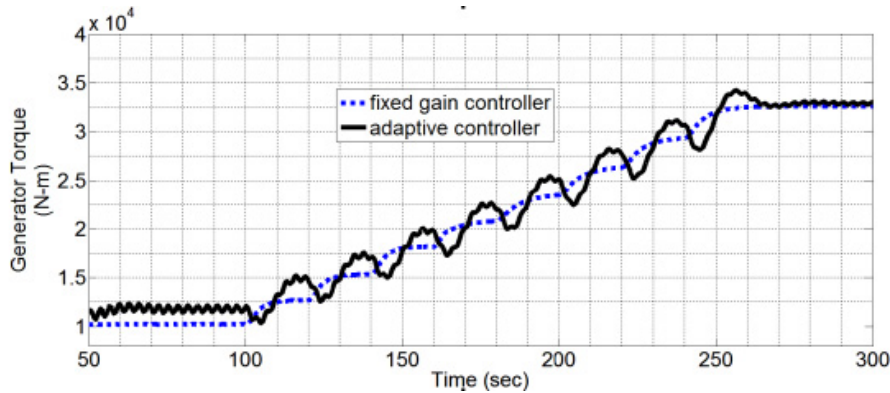


Figure 8. Torque command for step wind

7. Turbulent Wind

Identify Performance of the same controller for step wind profile described in previous section was also assessed using turbulent wind profile generated from the turbines. For this, the turbulent wind profile with mean wind speed of 7.5 m/s is chosen.

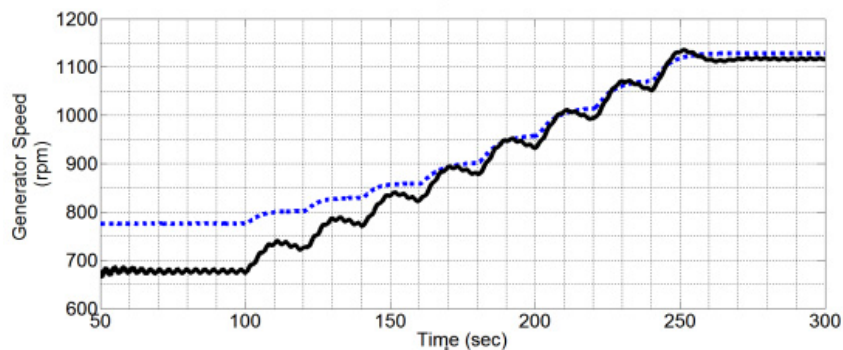


Figure 9. Generator speed for step wind

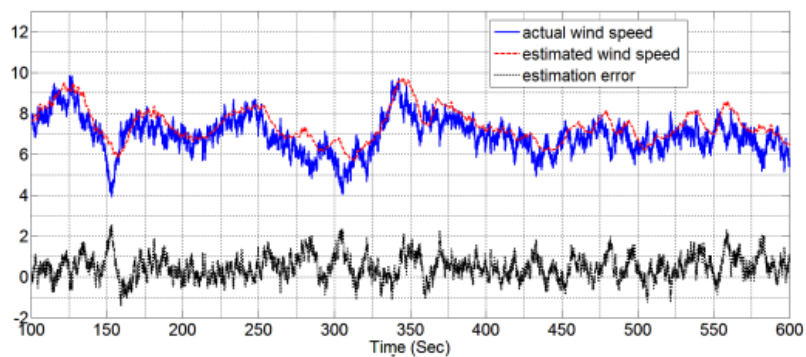


Figure 10. Turbulent wind speed, estimated wind speed, and estimation error

Figure (10) shows the turbulent wind profile used for simulation, estimation of wind speed using the disturbance estimator, and error in wind speed estimation. It can be observed that the estimated wind speed is close enough and also the estimation error is negligible. The performance of adaptive controller is compared with the baseline controller for turbulent wind profile. Figure (11) shows the comparison of TSR. In this figure the ideal TSR is the desired TSR for the wind turbine model used which is 7.55. Since it is difficult to observe the performance of controllers in turbulent wind case, the rms value of TSR is calculated to assess the performance with more

clarity. The rms value of TSR was found to be 7.84 for adaptive controller and 8.26 for baseline controller which concludes that the adaptive controller has better performance. The comparison of generator power is shown in figure (12). As in case of TSR, the rms value is calculated to have clarity in performance of controllers. The rms value ideal generator power is found to be 1.6973 MW. This ideal generator power is calculated using the ideal value of rotor speed required to track the desired TSR of 7.55. The rms value of generator power for adaptive controller is found to be 1.6619 MW whereas that of baseline controller is 1.279 MW which concludes that the adaptive controller is producing slightly more power compared to that of baseline controller.

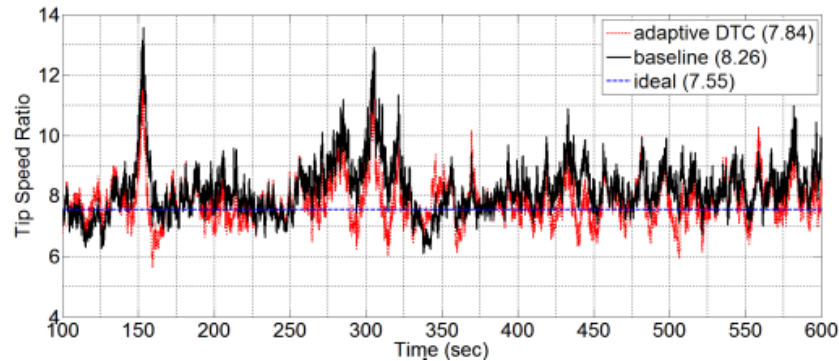


Figure 11. TSR comparison for turbulent wind

The estimated tracking error for turbulent wind is in figure 13. It can be observed that

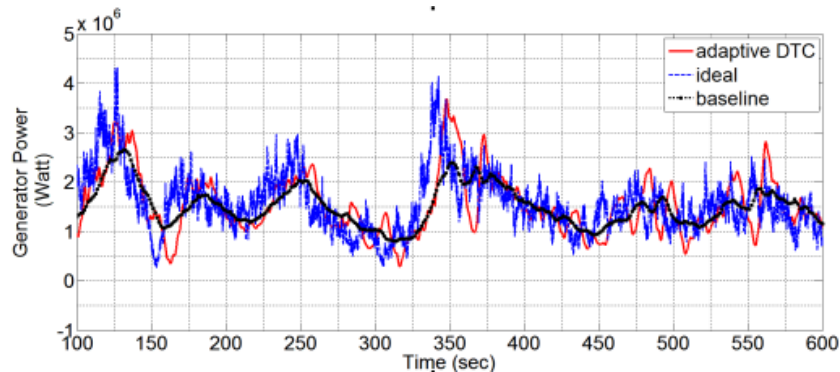


Figure 12. Generator power for turbulent wind

The estimated tracking error is negligible as expected from the theoretical analysis.

The generator speed profile is compared in figure (14). In this figure, the ideal generator speed is calculated with the desired TSR of 7.55 and the turbulent wind profile used for simulation using the relation $\lambda = \frac{\Omega}{w}$ and multiplying Ω with gear box ratio of 97 for this particular wind turbine model used. From the figure we can observe that the generator speed profile using adaptive controller is closely following the ideal generator speed profile compared to the baseline controller. The generator torque command produced by the controllers are in figure (15). As expected, the adaptive controller has more aggressive torque compared to that of baseline controller. This is because the adaptive controller tries to follow the varying turbulent wind profile.

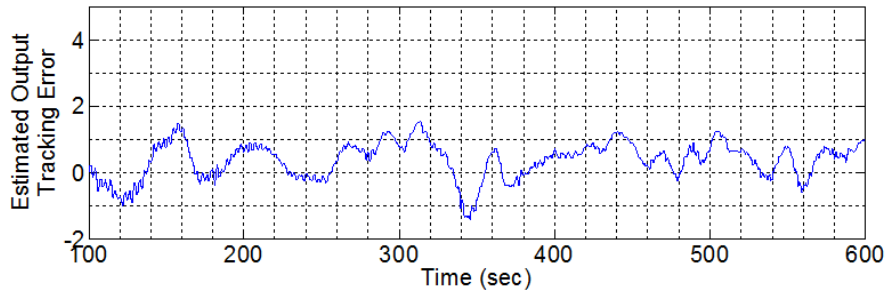


Figure 13. Estimated tracking error for turbulent wind

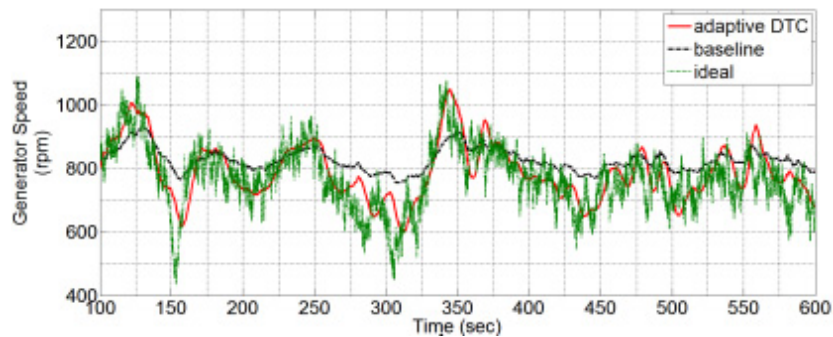


Figure 14. Generator speed for turbulent wind

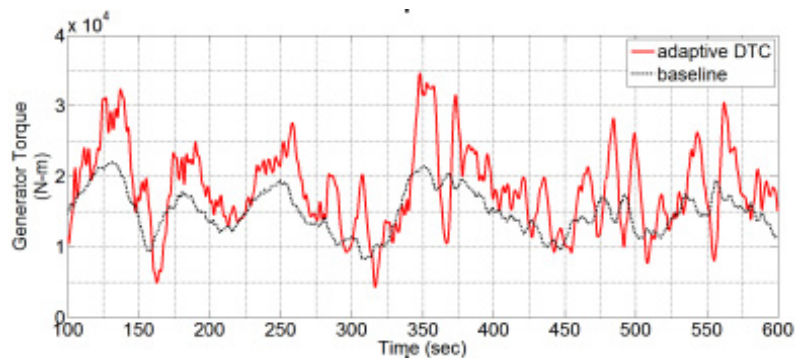


Figure 15. Generator torque for turbulent wind

8. Conclusions

The drawback of the simple wind estimator with output feedback is addressed in this chapter. The structure of wind speed estimator is slightly changed to incorporate a nominal reduced order plant model which does not have to be accurate. The nominal plant model is then combined with the disturbance generator model to design a state estimator where the wind speed is treated as one of the estimated state. It was also emphasized that the nominal model can be very simple having only one state. It is also observed that the estimated state must be used in feedback loop to ensure the stability and convergence of the controller. From the simulation results we observed that the proposed estimator closely estimates the wind speed. Also, the proposed control algorithm with adaptive control accompanied with state feedback has slightly better performance than the baseline controller while maximizing the power captures. Although the state feedback is used just to fulfill the stability requirements, this can also be used to achieve other advanced control objectives such as damping the tower motions, and minimizing the drive - train torsion. Also, the state feedback does not necessarily have to be fixed gain; it can also be adaptive.

Acknowledgments

The authors' deepest gratitude and thanks should go to Universiti Malaysia Pahang (UMP) for the full support during conducting this research work.

References

- Novak, P., Ekelund, T., Jovik, I., & Schmidtbauer, B. (1995). Modeling and Control of Variable-Speed Wind Turbine Drive System Dynamics. *Control Systems, IEEE, 15*, 28-38.
- Wright, A. D. (2004). *Modern Control Design for Flexible Wind Turbines*. NREL/TP-500-35816, Tech. Rep.
<http://www.g1-garradhassan.com/en/GHBladed.php>
<http://wind.nrel.gov/designcodes/simulators/fast/>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

Hardware-Software Complex for Laser Scanning in Color Sorting Machines

Elnur Megraliyevich Babishov¹, Gennady Vladimirovich Pakhomov¹, Vladimir Alekseevich Shulgin¹, Evgeny Yur'evich Buslov¹ & Dmitry Anatol'evich Minakov¹

¹ Voronezh State University, Universitetskaya pl., 1, 394006 Voronezh, Russia

Correspondence: Dmitry Anatol'evich Minakov, Voronezh State University, Universitetskaya pl., 1, 394006 Voronezh, Russia. E-mail: minakov_d_a@mail.ru

Received: July 13, 2014 Accepted: August 1, 2014 Online Published: September 26, 2014

doi:10.5539/mas.v8n5p262

URL: <http://dx.doi.org/10.5539/mas.v8n5p262>

Abstract

The paper concerns the problem of grain mixture analysis based on processing images synthesized during line-by-line scanning of each object in color sorting systems. The paper presents a hardware-software complex for sorting objects in the real time. The hardware part of the complex consists of two blocks: light source and device for reception and processing of images. The light source is a laser, passed through optical fiber and linearly expanded across the entire width of the photoseparator's chute. Linear laser scan produces significant intensity of illumination. It is sufficient for working on the transmission of radiation through objects. The software part of the complex also consists of two blocks: the thresholding algorithm and an automated program for finding the optimal parameters of sorting on the basis of that algorithm. The algorithm calculates the number of connected defective pixels with arbitrary shape. Automated program works on the basis of pre-formed images of the objects of two classes: good and defective. As a result, the program displays in tabular form the most optimal sorting parameters. The program shows the dependence of the loss of good product from the missouts of the defective objects. The customer gets a clear choice of the most suitable sorting results. This complex was tested for sorting of unshelled rice seeds via transmission. It is shown that the complex allows to effectively detect hidden seed defects: red pigmentation, immaturity, fungal diseases, and others.

Keywords: laser sorting, object recognition, image processing

1. Introduction

The technologies of sorting elements of grain mixture (EGM) used in today's color sorting machines (Satake et al., 1998) are based on the following principle: constant material flow is sorted basing on the analysis of one, two or three spectra of reflected electromagnetic radiation. The objects are illuminated either by broadband light from a single source, or by several light sources of the respective spectral range - usually LED or other type lamps. To register the data, the CCD or CMOS image sensors can be used. They allow to quickly identify defective objects and immediately remove them from the material flow, usually by means of an air-valve system.

In such sorting systems, thresholding algorithms (see, for the example (Gonzalez and Woods, 2002; William, 2001; Sahoo et al., 1988; Lee et al., 1990; Cseke and Fasekas, 1990)) are usually sufficient. As an accept/reject criteria these algorithms often use such properties of the EGM as the reflected radiation intensity, the number of defective pixels, lateral dimensions, etc. Such systems, however, cannot be applied in any case. It is, therefore, necessary to develop new hardware systems and elaborate optimal sorting algorithm parameters.

As for the hardware, new approaches suggest using laser illumination systems. As lasers produce narrow beams of coherent light of a single wavelength, it is possible to analyse the objects not only by the amount of light scattered from the surface, but also by the transmitted light after short-term exposures, and by their luminescence spectra (Algazinov et al., 2013; Algazinov et al., 2014).

Using fibre optics to calculate various source laser light emission will further add to the benefits of laser systems, as it allows for flexible programme control over the laser illumination spectra.

Another way to improve the EGM sorting system is to automate the process of choosing the optimal sorting parameters. The choice is based on the analysis of a large number of object images, synthesised by the sorting equipment.

Thus, the aim of this paper is to describe a hardware-software complex for color sorting, basing on laser illumination, which can be used to solve the problems that the existing sorting equipment is not able to cope with. During the test the complex was applied to sort shelled rice grains.

2. Materials and Methods

The initial registration of data as well as the image processing and analysis of the EGM was performed using the constructed hardware-software complex with laser illumination system, integrated into an original color sorting machine F5.1 (Chuiko et al., 2009; Babishov et al., 2013). The flow chart of the complex is shown in Fig.1.

Fig. 1 shows that the complex includes a video module (VM) connected to a personal computer (PC). The VM main element is a line-scan image sensor combined with a computing unit based on a dual-core processor.

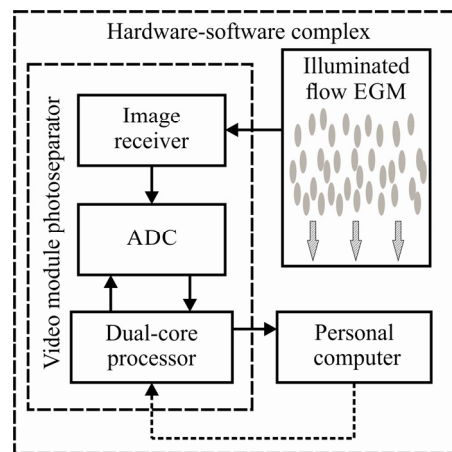


Figure 1. Flow chart of the hardware-software complex

The image sensor used in the complex is a Hamamatsu CMOS linear image sensor of high sensitivity S11108 (Hamamatsu, 2013). Its photosensitive area consisting of 2048 pixels, each with a pixel size of $14 \times 14 \mu\text{m}$. When all the 2048 pixels are active, maximum line rate of the sensor is 4.67 kHz.

The output data is then digitalized by the analog-to-digital converter (ADC) and passes to a special signal processing processor - dual-core Analog Devices ADSP-BF561 Blackfin processor (AD, 2009). The integrated development environment Analog Devices VisualDSP++ 5.0 with Update 9.1 (AD, 2007) was chosen for software development of VM with BF561 processor.

The software for the dual-core processor functions in real time and includes the following elements: connection to the PC, linear image sensor control, linear sensor image analysis and processing, connection with the other modules of the sorting machine.

During the sorting process, the VM registers transmitted laser light line-by-line, converts received video data into digital format, processes the lines (frames) of the image using special algorithms and determines whether the objects are pure or should be removed from the material flow.

The connection interface between the PC and the VM is the industrial interface of RS485 standard. The communication protocol is Modbus RTU protocol, with the operator's computer as the master device and the video module as the slave device.

The described complex creates a two-dimensional image of each object of the material flow. Linear monochrome sensor is used to form the lines (frames) of every single object coming in. Processor BF561 accumulates in SDRAM memory all images. After accumulating enough images of various object classes, they are transferred to PC, in which images can be further processed by the software described below.

Illumination by a narrow laser light beam was provided using cylindrical lenses (Fig. 2). The lens LJ1695L2 and LK1982L2 (Thorlabs) was chosen for this purpose. The choice of these lenses was determined by photoseparator geometry and required parameters of the laser line. In this case, the formed laser line had length 400 mm and width 2 mm. Laser diodes ATC-S1-0-FS-665-5-F200 and ATC-C5000-200-AMF-808-5-F200 (ZAO Semiconductor Devices) were chosen as the illumination sources. The first one has wavelength 650 nm with optical power 1 Watt, and the second one is 808 nm with optical power 4 Watts. These wavelengths were the

most suitable for our special task (shelled rice sorting).

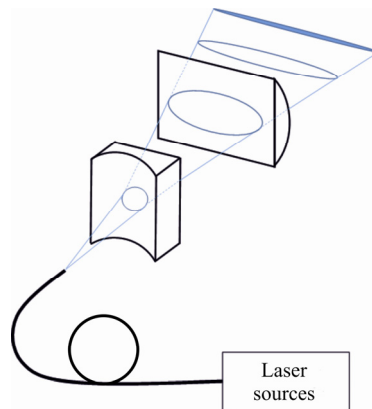


Figure 2. Optical lenses for laser light linear sweeping

The radiation summation from different laser sources is one of the main advantages of the proposed lighting system. We use optical fiber from Thorlabs with diameter 200 μm and numerical aperture 0.22. The radiation from each laser source using focusing optics was introduced into common fiber (Shulgin et al., 2014). The benefits of such a system are obvious, as it provides flexible computer control over the laser illumination spectra. The system also allows to place lasers and the optical circuit in any place.

To apply laser light linear sweep effectively, the resulting intensity received by the linear sensor, was distributed evenly along the width of the chute with the help of special software. This allows to apply uniform sorting parameters for the whole width of the chute. Apart from the programmed distribution of the laser light sweep, it is also necessary to set correctly the background light to detect the objects edges effectively.

To analyse EGM and sort them into classes according to their biological features, various spectrum analysis methods were applied - reflectance spectroscopy, transmitted light spectroscopy, luminescence spectroscopy (Schmidt, 2007). Reflectance, luminescence and transmitted light spectrum, were registered using experimental equipment based upon the fiber optic spectrometer USB4000-VIS-NIR (Ocean Optics). The detailed description of the equipment is given in the following papers (Sarycheva et al., 2012; Algazinov et al., 2013; Algazinov et al., 2014).

Spectral analysis for color sorting machine application forms two object classes: pure and defective.

Pure and defective objects recognition algorithm is an essential element of the VM software. Thresholding algorithms seem to meet all the requirements (Gonzalez and Woods, 2002; Sauvola and Pietikainen, 2000; Bichsel, 1998).

After an image frame is processed, a so called binary image is created, where pixels can be denoted only 0 or 1 depending on the pixel type - pure and defective. Further processing is performed upon the binary image. The next stage of the image frame processing within the sorting process includes finding defective areas (formed by the defective pixels) and determining their size. The material is thus sorted according to the size of the defective area on the object.

We have tested the most uniform variant of the sorting algorithm, when the defected area can be any shape. We can also call this algorithm the number of connected defective pixels. If the number of such pixels exceeds the set limit, then the object is rejected and removed from the flow.

The algorithm is thus based on two sorting criteria: the intensity threshold and the number of connected defective pixels. These criteria are then sent using Modbus protocol from the PC to the VM processor.

Before the sorting, it is, of course, necessary to adjust the sorting machine parameters to fit the objectives. In our work, we use automated search for each algorithm parameters based on the large number of object images of two classes: pure and defective.

Automated search for best sorting parameters presupposes determining best intensity threshold values and the number of defective pixels as a dependency of lost pure objects and accepted defective objects. This task is performed separately by the PC basing on the image analysis of a large number of objects of each class. We

should also point out, that image frames of pure and defective objects are stored separately.

3. Results

Lets assume that there is a model sample of pure objects' images N and defective objects' images M . The search algorithm works as follows. First, the threshold T and the number of connected pixels C is set. All the images are then binarized according to the threshold T . With the given parameters $\{T, C\}$ let the pure object images N_I be recognized as defective, and the defective object images M_I be classified as pure. Ratio $Q=N_I/N$ is the pure objects loss, and the ratio $P=M_I/M$ is the defective objects missout. Thus, Q and P are the functions of $\{T, C\}$, i.e. $Q=Q(T, C)$, $P=P(T, C)$.

Apart from the object images the PC receives also the information about the background level of the captured objects. Let T_{bkg} be the background level of the sample objects. If the defective objects are brighter (lighter) than pure ones, the threshold T varies in the range $[T_{bkg}, T_{max}]$; if the defective objects are darker, $[0, T_{bkg}]$, where $T_{max} = 255$ for 8-bit grayscale images.

For each T from the interval above, missouts $Q(T, C)$, and losses $P(T, C)$ are calculated as functions of the number of connected defective pixels C . For each object image for the given threshold T the maximum number of connected pixels is determined with the brightness interval $(T, T_{max}]$ (for light defects) or $[0, T)$ (for dark defects). The values are added to an array and sorted in descending order. Pure and defective objects are analysed separately. The defective pixels array for pure objects is then presented as $\{C_1, C_2, \dots, C_N\}$, with $C_i > C_{i+1}$. Similar array is formed for defective objects $\{C_1, C_2, \dots, C_M\}$. Then for any set number of defective pixels C the number of object images from the initial sample that will be recognized as defective, can be easily determined. Indeed, let C lay within the interval $C_k > C > C_{k+1}$, then with the set $\{T, C\}$ k object images (either pure or defective) will be recognized as defective. Thus, there is no need to analyse the whole sample every time C changes.

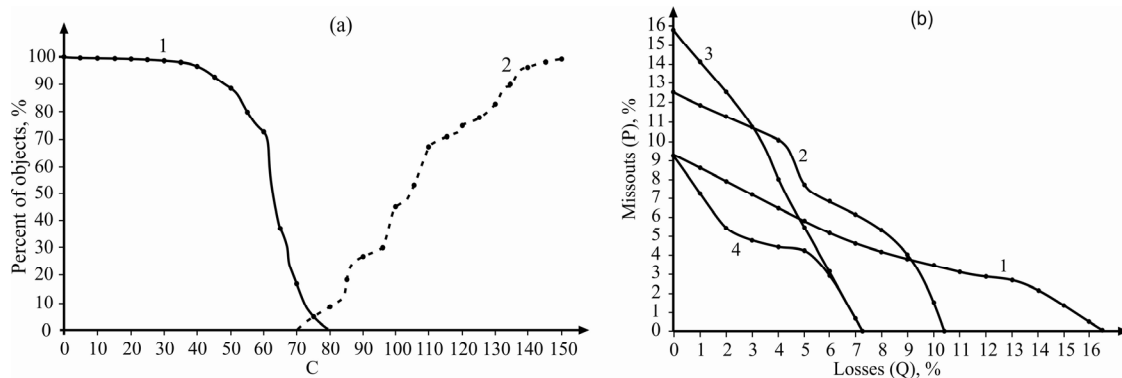


Figure 3. (a) - characteristic graph of losses (curve 1) and missouts (curve 2) with the set threshold T and various number of connected defective pixels C . (b) - typical dependency $P(Q)$ with various threshold T (for example, curve 1 – $T=60$, curve 2 – $T=80$, curve 3 – $T=100$; threshold gradation is equal 256). The dependency $P_{min}(Q)$ is also demonstrated (curve 4)

Characteristic for such threshold T , dependency of pure product loss $Q(T, C)$ and defective objects missout $P(T, C)$ presented in form of function of the number of defective pixels C is shown in Fig.3(a). The higher the number of defective pixels C with the given threshold T the less losses Q and the more missouts P there will be. The losses and missouts' graphs will have an overlap region. If the graphs do not overlap, then, with the set threshold T the objects are completely parted in two classes. In this case, values $\{T, C\}$ appear, where C is some number of defective pixels from the graphs' partitioning region, and the analysis is thus completed. Complete partitioning, however, is hardly ever possible with any threshold T , and losses and missouts' graphs always have an overlap region (Fig.3(a)).

Lets examine, with the given threshold T , the overlap region of losses' curve Q and missouts' curve P . In the overlap region there is a parametric dependence $P(Q)$, because for any number of defective pixels C form the overlap region $P(C)$, $Q(C)$ are determined. Subsequently applying all C values from the overlap region of the losses and missouts' curves (Fig.3(a)), it is possible to form a dependency $P(Q)$. Typical dependencies $P(Q)$ for various thresholds T are shown in Figure 3(b). Function $P(Q)$ will always monotonely decrease with growing Q , because with constant T , the higher are the losses Q of pure objects, the lower the missouts of defective objects

P .

This results in a set of dependencies $P_T(Q)$ for each threshold value T .

Each curve $P_T(Q)$ crosses the X-axis at some value $Q=Q_{max}(T)$ (Fig.3(b)). Then the minimum value from the set $Q_{max}(T)$ is determined, defined as Q_{max}^{min} . Value Q_{max}^{min} is the minimum level of pure product losses at which the number of defective objects' missouts P equals 0. The only way to reduce the number of losses to even lower values than Q_{max}^{min} is to introduce certain permissible level of defective objects' missouts P . To optimize the process, it is necessary for any given value of losses $Q_i < Q_{max}^{min}$ to find the minimal value of missouts $P_{min}(Q_i)$.

To determine the minimum missouts values $P_{min}(Q_i)$ we need to find the minimum value from functions $P_T(Q)$ with constant $Q=Q_i$ and all possible values of the threshold T . Thus, dependencies $P_{min}(Q_i)$, $T_{opt}(Q_i)$ are obtained, where $T_{opt}(Q_i)$ is the value of index T , i.e. the value of the sorting threshold, at which $P_T(Q_i)$ has minimal value $P_{min}(Q_i)$. Q_i acquires values from the range $[0, Q_{max}^{min}]$.

In actual process, the range $[0, Q_{max}^{min}]$ is discretized with a step-size ΔQ . It is usually enough to select $\Delta Q=0.01$, which fits the step-size of 1% loss.

From the loss function $Q(T,C)$ (Fig.3(a)) with $T=T_{opt}$ and a given value Q_i the number of defective pixels $C_{opt}(Q_i)$ is found. Therefore, $\{T_{opt}(Q_i), C_{opt}(Q_i)\}$ are the threshold values and the number of connected defective pixels, which yield the losses Q_i and missouts $P_{min}(Q_i)$.

For the sorting process it is vital to determine not the amount of defective objects missouts P_{min} at the given value of losses Q_i , but the quantity of dockage in the sorted grain mixture as compared to the initial grain mixture at the given value of losses Q_i . Let us show, how this difference can be determined by the found parameters Q_i u $P_i=P_{min}(Q_i)$.

Assume that the initial EGM batch consists of N_r defective and N_a pure objects. Lets now introduce parameter $\gamma=N_r/N_a$. Then the quantity of dockage in the initial batch will be $\beta=N_r/(N_r+N_a)=\gamma/(1+\gamma)$.

After the sorting, with losses Q_i and missouts P_i , the sorted grain mixture will contain $M_r = P_i N_r$ defective and $M_a = (1-Q_i) N_a$ pure objects. Lets introduce $m = P_i / (1-Q_i)$. Then for the sorted material $\theta = m\gamma$, the quantity of dockage will be $\lambda = m\gamma / (1+m\gamma)$.

As we can see, the less is m with constant $[\gamma]$, the less is the quantity of dockage $[\lambda]$ in the sorted material. To determine how much the quantity of dockage will reduce after the sorting, we need to know not only the value of m , but the value of $[\gamma]$ as well, i.e. the quantity of dockage at the initial stage.

As the quantity of dockage in the initial material is usually small, the value of $[\gamma]$ is not big. Then $\beta \approx \gamma$. As losses P_i and missouts Q_i are usually under 0.5, we may assume that $m \leq 1$, an then $m\gamma$ is also small and $\lambda \approx m\gamma \approx m\beta$. Thus, when the quantity of dockage in the raw material is small ($\leq 10\%$) index $m = P_i / (1-Q_i)$ shows, how does this quantity change after the sorting.

By the set of values Q_i , $P_{min}(Q_i)$ we determine $m(Q_i)=P_{min}/(1-Q_i)$ and get the following set of optimal sorting parameters: Q_i , $m(Q_i)$, $T_{opt}(Q_i)$, $C_{opt}(Q_i)$. These parameters are shown in a table with the step-size $\Delta Q_i=0.01$, starting with $Q_i=0$. Each of the following set of parameters is put into the table provided that the growth of losses goes together with better sorting quality, i.e. if $m(Q_i)$ reduces as compared to the previous value. From the table we can choose the best parameters as to the number of losses and the exit product purity depending on the certain objective.

Objects under examination: pure objects - shelled white-grain rice, provided by State Scientific Institution All-Russian Rice Research Institute, Krasnodar, Russia. Defective objects - grains with hidden structural defects: contaminated grains, unripe grains, cracked grains, chalky grains, and red rice grains. Pure objects sometimes look completely similar to the defective ones (Fig.4) This can be explained by the fact, that, for example, a contaminated grain is damaged from the inside, and the defect does not show on the hull.



Figure 4. Unshelled "Ametist" rice grains and defective grains (from left to right): pure grains, grains contaminated by fungi, red rice grains, unripe grains

The reflectance spectra analysis has demonstrated that it is very difficult to separate pure rice grains within the visible spectrum, even if more than one spectrum is analysed.

The transmittance spectra analyses yielded much better results. The transmitted light intensity for the defective grains in the visible spectrum is over 20% lower than for the pure ones. The best contrast ratio was registered, when the wavelength was 660 nm.

The spectral analysis data, therefore, demonstrates that shelled rice grain sorting can be successfully performed only by analysing the transmittance spectrum. The hardware-software complex described above allows to do this using laser light sources. As a result of the complex operation, several hundred images of pure and defective objects were obtained. The transmittance spectra analysis algorithm also allows to eliminate grains with very small defective areas (for example, slightly contaminated grains).

The results of the automated search for the best sorting parameters algorithm described above are shown in Table 1.

Table 1. Best sorting parameters for shelled rice

Losses Q , %	Dockage reduction m	Threshold T_{opt}	The number of defective pixels C_{opt}
0.00	0.495	8640	96.0
1.00	0.129	4480	22.8
2.00	0.084	4480	19.0
3.00	0.050	4800	20.8
4.00	0.045	4800	18.0
5.00	0.040	5120	18.8
...
32.00	0.000	8320	21.0

The first column of Table 1 contains the percentage of pure grain loss compared to the initial state of the batch. The second column shows dockage reduction index, i.e. dependence of the resulting and initial quantity of dockage. The third and the fourth columns contain the determined best sorting parameters of the threshold and the number of defective connected pixels. We should note, that when the transmittance method is used, defective objects will be darker than the pure ones. Hence, the pixels with brightness lower than the threshold values will be recognised as defective.

The Table demonstrates the flexibility of the automated search algorithm, as the consumer can choose any sorting parameters he finds appropriate. For example, if the consumer considers the 5% loss after the first sorting appropriate, then the expected percentage of the rejected defective objects is 96%, which demonstrates high efficiency of the sorting process. However, most of the lost objects can be regained during re-sorting of the rejected material. The final loss percentage will thus be minimum.

4. Discussion

Spectral analysis showed that unshelled rice seeds with hidden defects can not be sorting by reflecting method with high accuracy. The best result of separation can be achieved by registering the radiation passed through the seeds, since in this case there is maximum contrast between the good seeds and bad. So, the color sorting machines which used only reflecting method in visible and near infra-red spectral ranges are ineffective for solving such problem.

At the other point, some works (Grundas et al., 1999; Arkhipov, 2003; Hill et al., 1996) demonstrate that seeds with hidden defects can be recognition by using x-ray radiography. However, at present time these x-ray systems have some restrictions. First of all, such systems can not perform sorting in real-time scale. They can only reject all seeds in samples. Moreover, x-ray devices are complicated and respectively expensive. Also, the special care must be apply during working with x-ray devices. Our complex has no such disadvantages. Indeed, sorting in real-time scale gives us possibility do not reject all seeds, but reject only defective one. Therefore, we can significantly improve initial quality of seed batch. Also, laser sources do not require such special conditions for

health safety as x-ray systems.

It should be noted that the using of laser sources for separation tasks is not new. The interesting schemes of laser sorting are proposed in some works (Ruymen et al., 2005; Calcoen et. al., 2010; Adams et. al., 2009). In these works, the laser beam is deployed to the entire width of the scanning area by the use of a special fast-rotating mirror. Reflected from the object radiation again falls on the same mirror, and then it is directed by different optical devices in the receiving detector. Obviously, high performance sorting requires that the laser beam ran across the scan area with a frequency of several thousand times per second, and respectively, high rotational speeds of the mirror is needed. In addition, the photodetectors are unlikely to be able to provide high resolution for sorting small objects (1-5 mm).

The above disadvantages are partially removed in the scheme described in the work (Dirix et al., 2010). In that scheme, the rotating mirror is only intended to sweep the laser beam and does not redirect the reflected light to the photodetectors. Photodetectors are located in a separate block. This fact allows to simplify the optical system of the device. In addition, CCD image sensors can be used instead of simple photodetectors. As a result, the resolution can significantly improve. However, this scheme also requires a high speed rotation of the mirror prism.

Most of the above mentioned solutions which used laser illumination can be implemented only in conveyor separators. At the other hand, there are many color sorting machines in the market, which have inclined chutes for product feeding and can not use such laser systems, because they are too bulky. Almost all world leaders (Buhler Sortex, Satake, ASM and other) have color sorting machines with inclined tray which use only lamps or LEDs as lighting. We propose hardware-software complex with fiber laser system, which can be organically fit in color sorting machine with inclined chutes. In our experiments, we use color sorting machine from Voronezhselmash (www.vselmash.ru).

The radiation summation from different laser sources in one optical fiber is one of the main advantages of the proposed lighting system, which allows to sort in a few wavelengths. The angle between input laser beam and fiber is an important factor of the proposed lighting system. In work (Shulgin et al., 2014), it is shown that this angle must be non-zero. In this case, there are only ring mods from optical fiber. At the same time, axial mods are not excited and, accordingly, speckle structure with low spatial frequencies is absent. This fact ensures the stability in time of the output light along the width of the scan line.

Work (Upadhyaya, 2007) describes the issue of the presence of red rice as a weed for white rice in North America. The same problem is present in the Krasnodarsky region of Russia (Vasilieva et al., 2013). At the other hand, in work (Kawube et al., 2005), chemical and manual sorting method for rice seeds is described in order to increase the yields, because there is a lot of seeds with diseases. All these facts show that our system may be perspective and demanded for the solution of such problems.

Thresholding algorithms are considered as one of the easiest and popular in real-time systems (see, for example, Palumbo et al., 1986; Sahoo et al., 1988; Lee et al., 1990; Glasbey, 1993; Trier et al., 1995; Sezgin et al., 2004). These works consider techniques with fixed threshold. At the same time, there are more complicated methods (Parker, 1991; Yang et al., 1994; Shen et al., 1997; Sauvola et al., 2000; Yang et al., 2000; Bradley et al., 2007), which is called adaptive thresholding. The adaptive thresholding technique is more effective in situation when the illumination varies spatially. Such situations are likely in color sorting machines. In our present work, we use fixed thresholding method, which shows good results. The adaptive thresholding techniques are not required in our work, because any spatial irregularity of illumination was eliminated by the software correction coefficients. These coefficients remain unchanged during sorting because illumination is stable in time.

There are much more effective algorithms for object recognition: neuro-network algorithms (Gonzalez and Woods, 2002; Principe et al., 2000; Osowski, 2000), support vector machine algorithms (Scholkopf, 1998; Burges, 1998). These methods works as algorithms with training. The complexity of these algorithms imposes certain conditions on their application in machine vision systems with real-time response. Indeed, these algorithms require appropriate hardware with much more high-performance computing resources. At the same time, our work represents special method for automated searching of the optimal parameters for certain thresholding algorithm. The flexibility of this method is that the user can choose a suitable quality of the final product as a function of losses of good objects. From a practical point of view, our approach can be considered as algorithm with training for two classes of objects (good and defective).

5. Conclusion

A new hardware-software complex with fibre optic laser illumination system was suggested. Its main feature is

the laser light sweep with a stream of high density. Such system allows to sort objects according to their structural properties illuminated by transmitted light.

The software for the automated search for best sorting algorithm parameters was created. The said algorithm consists of two parameters: threshold and the number of connected defective pixels. The search results are presented in form of a specialised table, using which the consumer can consider various sorting options and choose the most suitable.

Despite of successful work, the proposed complex has some limitations. It is known that the intensity of the transmitted light considerably depends on the thickness of the object. Therefore, the complex will have some difficulty in sorting product, in which there is considerable variation in thickness of objects. Indeed, if accepted objects vary in thickness, so that the intensity of the transmitted through them radiation overlaps with the intensity of the defective object, there will be difficulties in recognition. We will receive a lot of defect missouts or large losses of good product. Successful sorting of the unshelled rice is due to the absence of the great variation in thickness.

Thus, the modernization of the complex is needed for taking into account the thickness of the objects. Modernization must be for the hardware part of the complex and the software one. Upgrading the hardware in the first place is to add another laser source. This is possible by entering into a single optical fiber laser sources from two different spectral ranges. Working with two sources will allow to exclude the influence of object's thickness. Also, two video-modules (with filters) are needed to introduce in the complex for receiving signals from both sources.

Software must also be modernized, because we will need to calculate the relationship of the radiation transmittances of the two spectral ranges for different types of objects (good and bad). In this regard, the sorting algorithm and automated program for finding of optimal sorting parameters will be upgraded.

These improvements are the subject of our further research.

Acknowledgement

This work was supported by the program of the Ministry of Education and Science of the Russian Federation "Cooperation of educational institutions and industrial enterprises" (the Decree of the Russian Government # 218 dated 09 April 2010 - series 3, project # 02.G25.31.0002)

Reference

- Adams D., Peelaers, B., & Dirix, B. (2009). Detection system for use in a sorting apparatus, a method for determining drift in the detection system and a sorting apparatus comprising such detection system. Patent US #7557922 B2.
- Algazinov, E. K., Dryuchenko, M. A., Minakov, D. A., Sirota, A. A., & Shulgin, V. A. (2013). Elements of grain mixture recognition method based on the spectral characteristics analysis in real-time color sorting systems. *VSU Bulletin, 2*(System analysis and information technologies): 9-19.
- Algazinov, E. K., Dryuchenko, M. A., Minakov, D. A., Sirota, A. A., & Shulgin, V. A. (2014). Methods of measuring the spectral characteristics and identifying the components of grain mixtures in real-time separation systems. *Measurement Technique, 1*, 36-41. <http://dx.doi.org/10.1007/s11018-014-0406-3>.
- Analog, D. (2007). VISUAL DSP++ 5.0. User Guide. Rev. 3.0, pp: 442.
- Analog, D. (2009). Blackfin Embedded Symmetric Multiprocessor. ADSP-BF561. Rev. E, pp: 64.
- Arkhipov, M. V. (2003). Radiographic technology of the Control and monitoring of Roentgen-negative attributes in seed browing. 2nd International Workshop "Applied Physics in Life Science", Prague, pp: 4-5.
- Babishov, E. M., Goldfarb, V. A., Minakov, D. A., Pahomov, G. V., Strygin, V. D., Chuiko, G. V., ... Shulgin, V. A. (2013). Laser sorter machine. Patent RU #2489215 C1. Statement. #22.
- Bichsel, M. (1998). Analyzing a Scene's Picture Set under Varying Lighting. *Computer Vision and Image Understanding, 71*, 271-280. <http://dx.doi.org/10.1006/cviu.1997.0627>
- Bradley, D., & Roth, G. (2007). Adaptive Thresholding Using the Integral Image. *Journal of Graphics Tools, 12*, 13-21. <http://dx.doi.org/10.1080/2151237X.2007.10129236>
- Burges, C. J. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery, 2*, 121-167. <http://dx.doi.org/10.1023/A:1009715923555>
- Calcoen, J., Janssens, C., & Erinkveld, J. J. W. (2010). Apparatus and method for sorting articles. Patent US

#7842896 B1.

- Chuiko, G. V., Strygin, V. D., Pahomov, G. V., Chyurikov, A. A., Shulgin, V. A., Ageev, A. A., & Shkirya, Y. I. (2009). Grain color sorter apparatus. Patent RU #83436 U1. Statement. #16.
- Cseke, I., & FASEKAS, Z. (1990). Comments on gray-level thresholding of images using a correlation criterion. *Pattern Recognition Letters*, *11*, 709-710. [http://dx.doi.org/10.1016/0167-8655\(90\)90105-B](http://dx.doi.org/10.1016/0167-8655(90)90105-B)
- Dirix, B., Adams, D., & Beeck, P. Op De. (2010). Method and system for use in inspecting and/or removing unsuitable objects from a stream of products and a sorting apparatus implementing the same. Patent US Application Publication #2010/0046826 A1.
- Glasbey, C. A. (1993). An analysis of histogram-based thresholding algorithms. *Graphical Models and Image Processing*, *55*, 532-537. <http://dx.doi.org/10.1006/cgip.1993.1040>
- Gonzalez, R. C., & Woods, R. E. (2002). Digital Image Processing. Second Edition. New Jersey: Prentice-Hall, Inc., pp: 793. ISBN: 0-201-18075-8.
- Grundas, S., Velikanov, L., & Arkhipov, V. (1999). Importance of wheat grain orientation for the detection of internal mechanical damage by the X-ray method. *Int. Agrophysics*, *13*, 355-361.
- Hamamatsu, 2013. CMOS linear image sensor. S11108, pp: 9.
- Hill, M. J., & Hill, K. A. (1996). X-ray radiography for rapid assessment of seed quality. Yield and quality in herbade seed production, Third international herbade seed conference. Haale. Germany, pp: 125.
- Kawube, G., Kanobe, C., Edema, R., Tusiime, G., Mudingotto, P. J., & Adipala, E. (2005). Efficacy of manual seed sorting methods in reduction of transmission of rice and cowpea seed-borne diseases. *African Crop Science Conference Proceedings*, *7*, 1363-1367.
- Lee, S. U., Chung, S. Y., & Park, R. H. (1990). A comparative performance study of several global thresholding techniques for segmentation. *Computer Vision, Graphics and Image Processing*, *52*, 171-190. [http://dx.doi.org/10.1016/0734-189X\(90\)90053-X](http://dx.doi.org/10.1016/0734-189X(90)90053-X)
- Osowski, S. (2000). Sieci Neuronowe Do Przetwarzania Informacji. Warszawa: Oficyna Wydawnicza Politechniki Warszawskiej. ISBN: 83-7207-187-X.
- Palumbo, P. W., Swaminathan, P., & Srihari, S. N. (1986). Document image binarization: Evaluation of algorithms. *SPIE*, *697*, 278-285. <http://dx.doi.org/10.1117/12.976229>
- Parker J. R. (1991). Gray level thresholding in badly illuminated images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *13*, 813-819. <http://dx.doi.org/10.1109/34.85672>
- Principe, J. C., Euliano, N. R., & Lefebvre W. C., (2000). Neural and adaptive systems: Fundamentals through simulations. Wiley, pp: 672. ISBN: 978-0-471-35167-2.
- Ruymen, M. J., & Berghmans, P. C. (2005). Apparatus and method for scanning products with a light beam to detect and remove impurities or irregularities in a conveyed stream of the product. Patent US #6864970 B1.
- Sahoo, P. K., Soltani, S., Wong, A. K., & Chen, Y. C. (1988). A survey of thresholding techniques. *Computer Vision. Graphics and Image Processing*, *41*, 233-260. [http://dx.doi.org/10.1016/0734-189X\(88\)90022-9](http://dx.doi.org/10.1016/0734-189X(88)90022-9)
- Sarycheva, I. N., Yanushevich, O. O., Shulgin, V. A., Minakov, D. A., & Kashkarov, V. M. (2012). Fiber optic device for fluorescence detection. Patent RU #2464549 C1. Statement. #29.
- Satake, S., Ito, T., & Ikeda, N. (1998). Color sorting apparatus for grains. Patent US #5,779,058.
- Sauvola, J., & Pietikainen, M. (2000). Adaptive document image binarization. *Pattern Recognition*, *33*, 225-236. [http://dx.doi.org/10.1016/S0031-3203\(99\)00055-2](http://dx.doi.org/10.1016/S0031-3203(99)00055-2)
- Schmidt, W. (2007). Optical spectroscopy in chemistry and life sciences. Moscow: Technosphaera, pp: 363. ISBN: 978-5-94836-140-6.
- Scholkopf, B., Burger, C. J. C., & Smola, A. J. (1998). Advances in Kernel Methods: Support Vector Learning. Cambridge, USA: MIT Press, pp: 376.
- Sezgin, A. E., & Sankur, B. (2004). Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, *13*, 146-168. <http://dx.doi.org/10.1117/1.1631315>
- Shen, D., & Ip, H. H. S. (1997). Hopfield neural network for adaptive image segmentation: An active surface paradigm. *Pattern Recognition Letters*, *18*, 37-48. [http://dx.doi.org/10.1016/s0167-8655\(96\)00117-1](http://dx.doi.org/10.1016/s0167-8655(96)00117-1)

- Shulgin, V. A., Babishov, E. M., Goldfarb, V. A., Minakov, D. A., Pahomov, G. V., Strygin, V. D., Chuiko, G. V., & Chyurikov, A. A. (2014). Fibre-optic laser-based sorting machine. Patent RU #2521215 C1. Statement. #18.
- Trier, O. D., & Jain, A. K. (1995). Goal-directed evaluation of binarization methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *17*, 1191-1201. <http://dx.doi.org/10.1109/34.476511>
- Upadhyaya, N. M. (2007). Rice Functional Genomics. Challenges, Progress and Prospects. New York: Springer Science+Bussness Media, LLC, pp: 499. ISBN: 0-387-48903-7.
- Vasilieva, N. K., & Kovryakova, E. A. (2013). Condition and improve the efficiency of rice production areas in the Kuban. *Terra economicus*, *11*, 205-208.
- William, K. P. (2001). *Digital Image Processing* (3rd ed). New York: John Wiley & Sons, Inc., pp: 758. ISBN: 0-471-22132-5.
- Yang, J. D., Chen, Y. S., & Hsu, W. H. (1994). Adaptive thresholding algorithm and its hardware implementation. *Pattern Recognition Letters*, *15*, 141-150. [http://dx.doi.org/10.1016/0167-8655\(94\)90043-4](http://dx.doi.org/10.1016/0167-8655(94)90043-4)
- Yang, Y., & Yan, H. (2000). An adaptive logical method for binarization of degraded document images. *Pattern Recognition*, *33*, 787-807. [http://dx.doi.org/10.1016/S0031-3203\(99\)00094-1](http://dx.doi.org/10.1016/S0031-3203(99)00094-1)

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

Proper Insulated Materials for Temperature Accumulation in Box Technology to Catalyze the Organic Digestion Processing on Community Garbage Disposal

Chulabut Chanthasoon¹ & Kasem Chunkao¹

¹Department of Environmental Science, Faculty of Environment, Kasetsart University, Bangkok, Thailand

Correspondence: Chulabut Chanthasoon, Department of Environmental Science, Faculty of Environment, Kasetsart University, Bangkok 10900, Thailand. Tel: 66-2579-2116. E-mail: chulabuut@gmail.com

Received: July 16, 2014

Accepted: July 28, 2014

Online Published: September 26, 2014

doi:10.5539/mas.v8n5p272

URL: <http://dx.doi.org/10.5539/mas.v8n5p272>

The research is financed by The King's Royally Initiated Laem Phak Bia Environmental Research and Development Project, Chaipattana Foundation.

Abstract

The research is aimed to determine the appropriate insulated materials and the thickness of insulated boxes for keeping in-box temperature around 70 °C. The experiments were conducted by taking amount of 500 g river gravels (number 3) to heat in Hot air oven at 100 °C for 24 hours before putting in the 0.027 m³ (30x30x30 cm) insulated boxes. The in-box (T1) and out-box (T2) ambient air temperature were recorded by automatic data locker from beginning to the end of experiments. The relationship between time (t) and in-box temperature by graphical techniques in order to select the appropriate insulated material under the criterions of The King's Royally initiative nature-by-nature process, simplicity technology, and low expense (or local materials using for constructing technology). Also, the relations between Q (heat conduction) of temperature differences (in-box and outside box) in varying time (t) as same as the insulated material thickness was evaluated from graphical products of fixing T1 (equivalent to 70 °C) and Q of varying ambient air temperature. The results found that rice straw as the appropriate insulated material tether with the minimum rice-straw insulated thickness of 6 cm in which the in-box temperature could be kept long enough for psychrophiles, mesophiles, thermophiles, and hyper-thermophile to complete the digestion of carbohydrates, proteins, celluloses, hemicelluloses, and fibers. Moreover, the research result was also pointed out the values of ambient air temperature (from -10 °C to 70 °C) is inversely related to thickness of the rice-straw insulated boxes, by taking the minimum thickness of 6 cm for T2 equivalent to 30 °C.

Keywords: insulated materials, heat conduction, box technology

1. Introduction

Thailand has been confronted on community and industrial solid waste disposal as well as infectious garbage in all parts of the country, especially populated cities, populated communities, and industrial estates due to the population growth and tourism promotion. In consequence, such solid wastes were over the carrying capacity because of sanitary landfills which located in good land for agriculture, particularly paddy areas and non-steep slope areas. It is realized that the failure of unqualified local management agencies, and people participation could be the reason why handling solid wastes did not success as appeared at the present time. Expectedly, the garbage disposal techniques are mostly used the machineries for sanitary landfill works in order to obtain the disposed products in terms of organic fertilizer, fermented objects, exchanging toxicants to non-toxic materials, sterilization, eliminating bad smells, disease spreading control, and treating leachate. Unfortunately, there has been very less sanitary landfill that could be achieved the target according to the failure of local administration and working experiences. It is no doubted that the landfill problems have been found scatteringly in all parts of the country, not only waste contaminants but also forming the visual pollution to the residents and tourists. In facts, every administer of the local administration offices really need to solve those problems but the success seems far distance to get the achievement due to politically discontinued measures as well as the discipline of residents and visitors concerning with garbage disposal management plan. However, the most important issue

would be placed on changing behaviors of stakeholders for their lifestyle in relation to garbage disposal management rather than issuing the obligations which is King Bhumibol's motto for community garbage disposal in Thailand. The King also stressed that it has to be taken in longer time, but it would be worthwhile in the future time (LERD 1999, 2000, 2011).

Accordance with community garbage composition is normally comprised of organic, recycling, and hazardous materials which are existed their own particularity. In principles, the recycled wastes can be reused by direct modifying for appliances, and also through industrializing processes for new products (Ahmed, 2004; Okot-Okumu & Nyenje, 2011; Phetchaburi Municipal, 2011; and Thapa, 1998). In addition, both the recycled and hazardous wastes have been paid more attention on how to get rid of them in terms of promoting the extension programs and declared obligations. In part of organic matter point of view, the landfill technique is applicably used for all over the world instead of using for energy (Geng et al., 2010; Chang & Chang, 1998; Fernandez-Rodriguez et al., 2014). As stated before, developing countries are still in using landfill technique for community garbage disposal, but land constraint is still existed in populated cities and communities (JICA, 1981, 1982, 1991; Bhuiyan, 2010). Followings found toxic contamination not only landfills but also the surrounding areas (Finnveden et al, 1995; Hao et al, 2008; Koroneos & Nanaki, 2012; Keenen et al., 1984). This is the reason why King of Thailand has to pay attention on using the concrete box technology for community garbage disposal that can be applicable in house gardens, towns, big or small cities, flatlands, uplands, highlands, arid areas, and islands (LERD, 1999, 2000, 2011, 2012).

LERD (1999, 2000, 2011, 2012) emphasized that the King's Royally Initiated Laem Phak Bia Environmental Research and Development Project, Laem Phak Bia sub-district, Ban Laem District, Phetchaburi province, Thailand has been announced to become the learning center of research and development on community garbage disposal and wastewater treatment. The said research and development projects, especially community garbage disposal, have to follow the King's words on applicable nature-by-nature process, simplicity technology, and useable local materials for constructing technology or very low expense. For serving the King's intention, the research and development on community garbage disposal project has been conducted since 1990 together with the environmental education promotion and academic servicing at the Royal LERD's Learning Center itself and also in all parts of the country. (LERD, 1991, 2000, 2011) However, the simple technology was used for making organic wastes to become inorganic materials (garbage compost) through aerobically bacterial organic digestion process by using Royal Concrete Box technology (RCB technology). The strength of RCB technology is that it can be constructed in limited area in small size and even concrete floor of populated city, So, the RCB technology is applicable everywhere, even in the populated areas and concrete ground floors. Unexpectedly, when it was constructed in highland areas (elevation above 700 mMSL) in Chiangrai province which is normally cool climate. It was functioned to digest only carbohydrate (at temperature 30-40 °C) and proteins (at temperature 40-50 °C) of the organic home garbage but no digesting cellulose and hemicellulose at temperature (> 50 °C) and fiber as well. Actually, RCB technology could release temperature above 70 °C as found during the bacterial organic digestion processing in which the carbohydrates, proteins, celluloses, hemicellulose and fiber were become to inorganic materials (LERD, 1999, 2000, 2011, 2012; Metcalf & Eddy, 1972; Nopparatanaporn, 1992; Steger et al., 2007; Ugwuanyi et al., 2005; Wilson, 1977).

Naturally, either aerobic process or anaerobic process, the heat has exactly to liberated during the bacterial organic digestion processes in landfills, garbage piles, and litter bins as stated by Keenen et al. (1984), Botkin & Keller (2010), Atchley & Clark (1979), Ugwuanyi et al. (2005), Tchobanoglous et al. (1993), Wilson (1977), Steger et al. (2007), Metcalf & Eddy (1972), Nopparatanaporn (1992), Fernandez-Rodriguez et al. (2014), Geng et al. (2010), and Appels et al. (2010). The liberated heat quantity and producing time are relied on the content and proportion of carbohydrates, proteins, celluloses, hemicelluloses, and fibers in their composition of community solid wastes (Ariunbaatar et al., 2014; Finnveden et al., 1995; Chang & Chang, 1998; Koroneos & Nanaki, 2012; Swati & Joseph, 2008; Chunkao, 1998). In parallel conditions, the liberated heat plays a significant role in activating on both the aerobes and anaerobes for specific functions to their processes in organic digestion for converting the fresh dead plants and animals to become the inorganic materials as nutrient sort for plant growth (Prabuddham, 1985; Botkin & Keller, 2010; Atchley & Clark, 1979). However, elongation of liberated heat is very necessary to complete the maturing organic compost through the natural processes of bacterial organic digestion (Ariunbaatar et al., 2014; Appels et al., 2010; Swati & Joseph, 2008; Atchley & Clark, 1979; Ugwuanyi et al., 2005; Steger et al., 2007; Fernandez-Rodriguez et al., 2014), in other words, the prolongation of heat liberation during bacterial organic digestion processes from the beginning to the end is expected to keep in-box temperature equivalent to ambient air up to more 70 °C which provides the conditions for completely fermentation of all kinds of organic wastes.

The aforesaid description brought to take part for elongating liberated heat from bacterial organic digestion processing in order to achieve the completion of community solid wastes, especially in parts of carbohydrates, celluloses, hemicelluloses, and fibers. Hence, using insulated boxes are needed to prolong liberated heat from organic waste digestion to enhance the digesting activities of psychrothermopiles, mesothermopiles, thermopiles, and hyper thermopiles one way or another. So far as stated before, an attention has to focus on liberated-heat concrete boxes due to rapid re-radiation to transfer through the 3 cm thick concrete walls to the cooler air outside boxes that making difficult to apply the concrete box technology in highlands or the areas with low temperature all-year periods. For research experiences of LERD (1999, 2000, 2011, 2012) from northern highland of Thailand (Doi Tung in Chiangrai province, elevation above 1200 mMSL), the temperature inside concrete boxes has been measured below 50 °C and found the existence of high-pertained cellulose and fiber of dead plants still fresh after fermenting 30 days due to the concrete boxes not being all-the-time keeping the liberated heat during bacterial organic digestion processing as stated by Botkin & Keller (2010), Metcalf & Eddy (1972), Tchobanoglous et al. (1993), Atchley & Clark (1979), and Appels et al. (2010). If the RCB technology was covered with insulated materials, the heat liberation was expected to elongate for making the shorter solid retention time (SRT). In consequence, the bacterial organic digestion process could be fully functioned on carbohydrates by psychrophilic bacteria, proteins by mesophilic bacteria, celluloses /hemicellulose by thermophilic bacteria and fiber by thermophilic bacteria/ hyperthermophilbacteria (Swati & Joseph, 2008; Zupancic & Ros, 2003; Chunkao, 1998; Shahriari et al., 2013; Juteau, 2006).

As mentioned before, this study is aimed to find the insulated materials to reduce heat transfer between objects of differing temperature in thermal contact in which it can be achieved with specially engineering method concerning with suitable shape and insulation capacity (thermal conductivity) of materials. However, the thermal conductivity plays role in heat conductive transfer in which the Fourier's law of heat transfer have been applied Fourier's law of heat conduction by Holman (2009), Sellers (1965), Chunkao (1979), Gates (1965, 1980), and Hartman (1994) in the content of "the law of heat conduction states that the time rate of heat transfer through a material is proportional to the negative gradient in temperature and to the area at right angle to that gradient, through which the heat flows". It can be illustrated in equation (1) and showed the hypothetical views in Figure1:

$$Q/t = -kA(T_1 - T_2) / X \quad (1)$$

where:

- Q = heat transfer, cal/cm
- k = thermal conductivity, cal/cm²/sec/°C
- A = surface area, cm²
- T1 = hot temperature, °C
- T2 = cold temperature, °C
- X = thickness of insulated material, cm
- t = time taken, sec
- q = Q/t, cal/cm/sec

According to vary X (thickness) of materialized surface area, and q = Q/t, then equation (1) can be rewritten as:

$$q = -kA(T_1 - T_2) / X \quad (2)$$

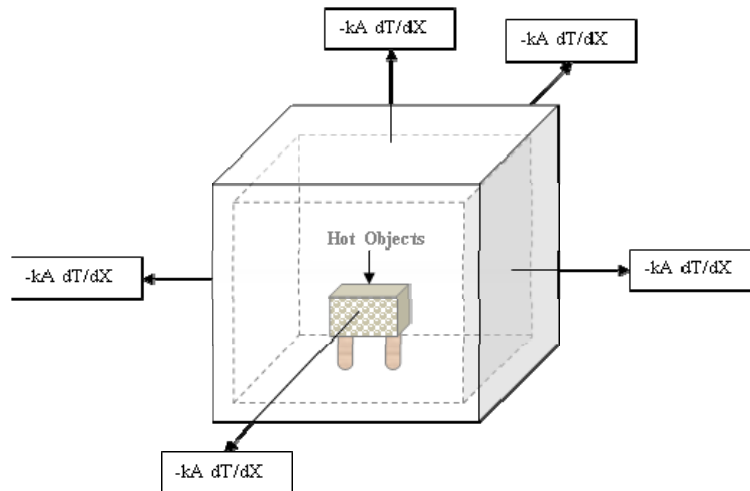


Figure 1. Hypothetical illustration Fourier's law of transferring heat conduction from higher-temperature to cooler-temperature objects at right angle to the temperature gradient

T_1 is hot temperature due to heat liberation during the bacterial digestion processes for cellulose, hemicellulose and fiber in which the thermophilic bacteria required temperature inside the boxes above $70\text{ }^\circ\text{C}$ as followed the research results from LERD (1999, 2000, 2011, 2012), Ariunbaatar et al. (2014), Appels et al. (2010), Steger et al. (2007), and Shahriari et al. (2013). Thus, T_1 could be fixed at temperature of $70\text{ }^\circ\text{C}$. Then, T_1 in equation (2) can be replaced by $70\text{ }^\circ\text{C}$ which will be presented as:

$$X = -kA(70 - T_2) / q \quad (3)$$

Actually X (thickness of insulated material) is used for this experiment on 2.5 cm for reduction of heat transfer (q) from T_1 point to the T_2 point through thermal conductivity with cross sectional area A . So, the equation (3) can be expressed below:

$$Q/t = -(kA)(T_1 - T_2) / X \quad (4)$$

In case of expressing the heat transfer per unit area, then the equation (4) becoming to equation (5) as following:

$$Q = -(kA(T_1 - T_2)/X)(t) \quad (5)$$

If t is fixed, then T_2 is varied from place to place, the same as fixed T_2 value and taking time (t) is varied. Then after, the relationships between Q and T_2 can do the same as Q and t relation. Because of aforesaid statement, those relations can be graphically constructed in which the proper t (time taken) and proper T_2 (temperature of applicable location) can be calculated in order to determine the thickness of specific insulated materials.

The proper thickness of insulated materials for reduction heat transfer from organic digested boxes plays vital role in keeping all-time high temperature inside boxes for accelerating the psychrophilic, mesophilic, and thermophilic bacteria activities to digest carbohydrates, proteins, cellulose, hemicellulose and fibers in which they are in part of home garbage. Besides, the levels of released heat in boxes are belonged to the important environment in catalyzing the bacteria activation on the organic-sort digestion under aerobic and anaerobic processes, such as psychrophiles (cold-loving bacteria) for below $20\text{ }^\circ\text{C}$, mesophiles for 20 to $45\text{ }^\circ\text{C}$, and thermophiles (hot-loving bacteria) for 45 to $70\text{ }^\circ\text{C}$, hyperthermophiles for more 70 to $122\text{ }^\circ\text{C}$. However, the environmental scientists are agreed on the bacterial organic digestion process as the basic heat releasing that are depended on the sorts of organic matter, for example, up to 40 to $50\text{ }^\circ\text{C}$ from carbohydrate group, 50 to $60\text{ }^\circ\text{C}$ from proteins, and above $60\text{ }^\circ\text{C}$ from cellulose, hemicellulose, and fibers. In facts, the composition of Thai municipality and household solid wastes is generally varied on proportion of organic matters, recycles, and hazardous wastes. It is remarkable that the organic matters are comprised of various proportions of carbohydrates, proteins, celluloses, hemicelluloses, and fibers. In other words, the ratio of carbohydrate: protein: cellulose/hemicellulose/fiber plays a significant role in heat releasing from organic wastes during digestion processing by providing gradually higher temperature until no organic matters left in technological boxes. Keeping higher temperature in boxes is depended on how thick of insulated materials as thermal conductivity to reduce heat transferring to the next one.

Not only thickness of insulated materials is influenced on heat transfer from hot temperature from inside to

outside boxes but also the density and its specific gravity are involved with the pathway of transferring heat as liberated during the bacterial organic digestion process. In conclusion, the Q values are directly varied to insulated material density and specific gravity but inversely to the thickness of insulated materials and ambient air temperature (outside technological boxes). In facts, the outside box temperature is supposed to be the temperature at any concerned areas in which the box technology for organic waste disposal. Only thickness of insulated materials plays vital role in control the inside box temperature to enhance the thermophilic bacterial activities for digesting cellulose, hemicellulose, and/or fibers which are in part of vegetables. In other words, the application of insulated-material box technology is really needed the proper temperature to enhance psychrophiles, mesophiles, thermophiles and hyperthermophiles for digesting carbohydrates, proteins, celluloses, hemicellulose, and fibers to become mature-garbage compost as the final products of fermenting organic solid wastes.

This study is aimed to determine the thickness of insulated materials for reduction of heat transfer from bacterial organic digestion process in order to keep the enough hot temperature up to higher than 70 °C which is proper heat to activate the thermophilic bacteria to release enzyme for digesting cellulose, hemicellulose, and fiber as in part of municipality and household garbage. Also, the solid retention time (SRT) is to show how short of time for doing digestion of organic solid wastes that turning to become garbage compost (organic compost).

2. Method

2.1 Experimental Area

The experimental area was taken the zinc-roof cottage inside the Royal LERD project site at Laem Phak Bia subdistrict, Ban Laem district, Phetchaburi province, in southern part of Thailand about 120 km distance from Bangkok. The zinc-roof cottage was covered with open space about 100 m² together without walls of all-four sides in order to provide all-direction wind blowing as shown in Figure 2.

2.2 Choosing Insulated Materials

This research project is belonged to The King's Royally Initiative Laem Phak Bia Environmental Research and Development Project (Royal LERD Project) which is qualified nature-by-nature process, simplicity technology, and either intentionally using domestic materials for creating technology or low expense. For serving the Royal LERD project criteria, the six-type appropriate insulated materials were chosen, that is, earth soil, rice husk, rice straw, wood, cement, plastic PVC, and foam (see Figure 2).

2.3 Insulated Sheet Making

The research methodology is required the cube-shaped experimental units with 30 cm in equal sizes of width, length, and height and 2.5 cm thickness (0.027 cubic meters) for molding the insulated sheets which are used for box walls, beds, and covers. According to the six-malleable insulated walls were needed to assemble for one cube-shaped experimental unit and 21 boxes required (7 treatments and 3 replications) for each insulated materials (earth soil, rice husk, rice straw, wood, cement, plastic PVC, and foam) by mixing with some cementing agents to insulated sheets for assembling the cube-shaped experimental unit before laying down on the smooth floor under the zinc-roof cottage in order to air drying (see Figure 2).

2.4 Determination of Insulated Material Density and Specific Gravity

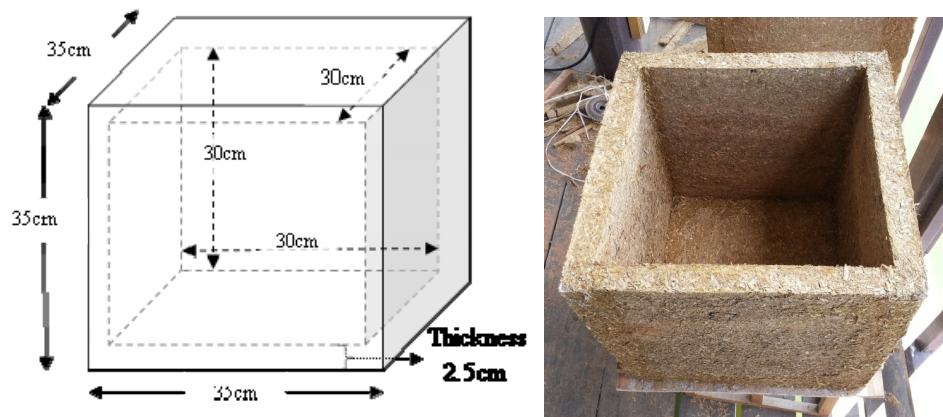
Each of insulated-material molding sheets had to determine the density, specific gravity, and thermal conductivity in which they played vital role in amount and transferring rate of heat conduction from hot objects or heat liberation during the bacterial organic digestion processing.

2.5 Heating Insulated Boxes

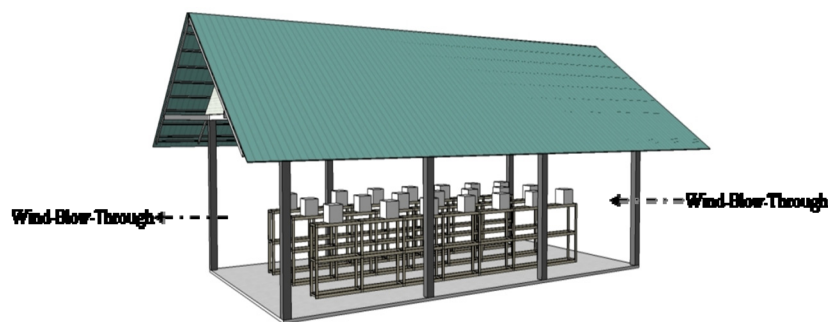
Since this research purpose is aimed to select the type of insulated materials and its appropriate thickness of insulated-material molding sheet, the 0.5 kg river gravels (number 3) at 100 °C by oven heating were used as hot objects to put in every experimental unit before closing the thermometer-inserted cover for measuring in-box temperature. Also, the outside experimental-insulated box temperature was measured by hanging the thermometer above the cottage floor about 1.50 m height.

2.6 In-Box and Outside-Box Temperature Measurement

Two automatic thermometers (Data Logger Testo 0572.1764 Waterproof Immersion Probe T/C Type K) was inserted through the insulated box cover for measuring inside insulated box temperature, and another one that hanging above cottage floor about 1.50 m height for measuring ambient air temperature.



(a) Insulated box of cube shape with the size of 0.027 cubic meter



Open Ventilation Zinc-Roof Cottage

(b) The experimental sites for laying down 21 boxes experimental unit

Figure 2. Insulated box of cube shape with the size of 0.027 cubic meter (30 x 30 x 30 cm) and 2.5cm thickness as constructed by seven- insulated materials of plastic PVC, cement, wood, foam, rice husk that mixing with another cementing agents in different proportions and laying down on the smooth floor under the zinc-roof cottage in order to air drying

2.7 Calculation of Heat Conduction

By taking in-box and outside box temperature as measured continuous period from beginning to the end of experiments, the relation graphs between in-box temperature (T1) and time (t), heat conduction (Q) and time (t) by fixing hot T1 equivalent to 70 °C, Q and t of the seven-type insulated materials by fixing T1 equivalent to 70 °C, and finally between Q and t of selecting insulated material by fixing T1 equivalent to 70 °C and varying selected insulated thickness from 1.0, 1.5, 2.0 to 9.0 cm with 0.5 cm interval.

2.8 Graphical Techniques for Evaluating Insulated Materials and Thickness

The relations as described in number 2.7 above were plotted graphs into 4 cases: firstly, the relation between T1 and t of seven-type insulated materials by fixing box thickness equivalent to 2.5 cm, and T1 equivalent to 70 °C, secondly, Q and t relation by fixing box thickness equivalent to 2.5 cm and T1 equivalent to 70 °C, thirdly, for selected-insulated-material heat conduction Q and t by fixing box thickness equivalent to 2.5 cm and T1 equivalent to 70 °C, and finally, Q and t by fixing T1 equivalent to 70 °C and varying selected insulated thickness from 1.0 to 9.0 cm with 0.5cm interval.

Intentionally, the appropriate thickness of selected insulated materials for making the insulated sheets that becoming the experimented box walls can be presumably find out which plays the significant role in keeping the liberated heat from bacterial organic digestion process inside the boxes, and also it can elongate enough working time of psychrophiles, mesophiles, thermophiles, and hyper-thermophiles to digest the organic wastes in parts of carbohydrates, proteins, fats and oils, cellulose, hemicellulose, and fiber to become the inorganic materials in

form of the organic compost.

3. Results and Discussion

Since, this research is focused on determine the kind of insulated material and its composition to mold in sheet form in order to construct the squared-insulated box technology (INBOT) to keep the hot-temperature for thermophilic bacteria activities in digesting cellulose-fiber organic wastes. Nevertheless, the insulated-sheet thickness is another purpose of this study on shortening the solid retention time (SRT) through the King's Royally initiative nature-by-nature process at the hand of bacterial organic digestion process. To achieve the aforesaid statement, the followings will be presented in details on both the appropriate insulators and its thickness that can be applicable in every climatic condition in Thailand. The research results will be presented in the followings.

3.1 Molding Insulated-Material Sheets

As stated before, there were 7 insulated materials that laminating to the insulated sheets for assembling the insulated boxes in order to keep the liberated heat from bacterial organic digestion process. Actually, some insulated materials were required another materials together with cementing agents in order to mold the insulated-laminated sheets. The component of mixed insulated-laminated sheets was shown in Table 1 in which they applied to calculate the averaged thermal conductivity (k) by weighing system as illustrated in Table 2. At the same time, the density and specific gravity were determined by laboratory techniques as Table 3.

Table 1. Insulated materials and mixing components for molding insulated- laminated sheets (30 cm width, 30 cm length, and 2.5 cm thickness) to assemble the cube-box (30 x 30 x 30 cubic centimeters) as the experimental units

Insulated Materials	Ratio of Mixing Materials	Component
Rice Straw	Rice straw : cassava starch solution = 1:1	Rice straw + cassava starch solution
Rice Husk	Rice husk : cassava starch solution = 1:1	Rice husk + cassava starch solution
Earth Soil	Clay : rice husk : sand = 3:1:1	Clay + rice husk + sand
Cement	cement Portland : Sand = 1:3	cement Portland + Sand
Plastic, PVC	-	PVC + epoxy
Wood	-	Wood + Latex glue
Foam	-	Foam + Latex glue

Note. Concentration of cassava starch solution is 25% of cassava starch.

Table 2. Calculation by weighing system of averaged thermal conductivity values of earth soil, rice husk, rice straw, cement, foam, plastic PVC, wood, as used for molding insulated sheets to assemble the cube boxes (30 x30 x 30 cm³) with 2.5 cm thickness

Insulated material	Averaging by Weighing System									Average k
	k1	w1 (Kg)	k1w1	k2	w2 (Kg)	k2w2	k3	w3 (Kg)	k3w3	
Rice Husk	Rice Husk 0.00009	0.745	0.0001	CS Solution 0.00071599	0.745	0.000533	-	-	-	0.00040
Rice Straw	Rice Straw 0.0002	0.85	0.00018	CS Solution 0.00071599	0.85	0.000609	-	-	-	0.00047
Earth Soil	Earth Soil 0.0026	3.24	0.0085	Rice Husk 0.00009	1.08	0.0001	Sand 0.001	1.08	0.000696	0.00172
Cement	Cement 0.0007	1.4	0.000969	Sand 0.000644391	4.2	0.002706	-	-	-	0.00066
Plastic, PVC	-	-	-	-	-	-	-	-	-	0.00021
Wood	-	-	-	-	-	-	-	-	-	0.00029
Foam	-	-	-	-	-	-	-	-	-	0.00007
Concrete	-	-	-	-	-	-	-	-	-	0.00239

Table 3. Averaged thermal conductivity, density, and specific gravity of insulated materials and some components for applying to calculate the heat conduction transferring

No.	Materials	Density (g/cc)	Specific Gravity	Thermal conductivity, k,(cal/sec)/(cm ² °C/cm)
1	Rice Husk	0.64	0.641	0.00040
2	Rice Straw	0.50	0.501	0.00047
3	Earth Soil	1.63	1.643	0.00172
4	Cement	1.84	1.855	0.00066
5	Plastic,PVC	1.51	1.522	0.00021
6	Wood	0.48	0.483	0.00029
7	Foam	0.01	0.01	0.00007
8	Cassava Starch Solution	-	-	0.00072
9	Sand	-	-	0.00064
10	Concrete	-	-	0.00239

3.2 Selection of Insulated Materials

The measured inside insulated-box temperature (T1) and ambient air temperature (T2) of each insulated-material box were conducted in relation to measuring time (t) and heat conduction (Q) in order to use for making decision of selecting the appropriate insulated materials as shown in Figures 3 and 4.

The results found that the outside box temperature was not influenced to uprise the inside box temperature but they were up to the powering radiated heat from 100 °C river gravel in which they were gradually released until all-box temperature equivalent to outside box temperature after radiating for 265 minutes. Among those temperature tendency, foam insulated boxes could be most sensitive changing in relation to outside box temperature by rapid increasing at the beginning and fast decreasing after radiating about 90 minutes, and sinking to the lowest level of temperature decreasing at night time. This phenomena is brought to point out that the foam insulated box technology would be used for keeping heat in box if the outside temperature was cool at night time. In spite of rice husk and rice straw can be used for insulated box technology, but the plastic PVC, cement, earth soil, and wood insulated boxes would be appropriate technology for keeping high temperature for longer period in boxes as well. If the convenience is concerned with the local materials available, the rice straw should be compensated by local grasses because of its grow everywhere in the whole country. Neither more effective insulated-box technology nor less appropriate ones, their box temperature lines as belonged to insulated materials (foam, rice husk, rice straw, plastic PVC, cement, earth soil, and wood) technology have met together at about 265 minutes after putting the hot river gravels in insulated boxes as shown in Figures1 and 3.

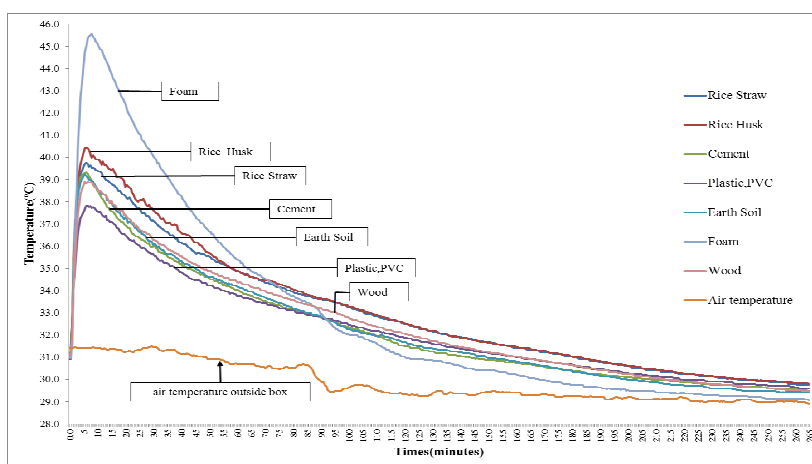


Figure 3. Tendency of air temperature outside boxes (inside of the insulated zinc-roof hut) and inside box temperature from beginning and climbing up to the maximum of insulators as made by foam, rice husk, rice straw, cement, earth soil, plastic PVC, and wood as obtained from the experiment

For making sure of selecting the insulated materials for assembling the box technology to dispose the community

solid wastes, the heat transfer from hot river gravels was applied to make decision. In experimental point of view, the insulated materials which use for box making were determined their density and specific heat capacity as well as thermal conductivity in order to induce heat flow from 100 °C hot river gravels to transfer to outside boxes. The weighting system was determined the averaged values of insulated-material, specific gravity, density, specific gravity, and thermal conductivity as illustrated in Table 3. The results found the thermal conductivity of foam, plastic PVC, wood, rice husk, rice straw, cement, and earth soil as 7×10^{-5} , 22.1×10^{-4} , 29×10^{-5} , 4×10^{-4} , 4.7×10^{-4} , and 6.6×10^{-4} cal/cm²/sec, respectively as shown in Table 3. Based on Fourier's law of heat transfer in equation (5), the thermal conductivity is related directly to quantity of heat transfer, which means that the more the thermal conductivity are the heat transfer. In the same manner, the density and specific gravity of insulated materials showed the same trends of thermal conductivity as illustrated in Table 4 and Figure 4. It is observed that the maximum in-box temperature found 45.5 °C for foam, 40.3 °C for rice husk, 39.8 °C for rice straw, 39.3° C for cement, 38.9 °C for earth soil, 38.8 °C for wood, and 37.9 °C for plastic PVC (see Figure 3). This would be reasoned that the radiation transfer from a handful of hot river gravels (100 °C) was not forceful enough to homogeneously spread out to nearly empty boxes as shown in Figure 1.

The results pointed out that the outside box temperature was not influenced to uprise the inside box temperature but they were up to forcefulness of heat radiating from 0.5 kg river gravels at 100 °C in which they were gradually liberated until equilibrium temperature between inside and outside box temperature after radiating for 265 minutes (Figure 3). Among those temperature tendency, foam insulated boxes could be most sensitive changing in relation to outside box temperature by rapid increasing at the beginning and fast decreasing after radiating about 70 minutes, and sinking to the lowest level of temperature decreasing at night time. This phenomena is brought to express that the foam insulated box technology would be good enough to use for keeping heat in box at daytime, only the outside box temperature was cool at night time. Besides, the rice husk and rice straw can be better useable for insulated box technology, the plastic PVC, cement, earth soil, and wood insulated boxes might be appropriate to assemble the insulated-material technology for keeping high temperature for longer period in boxes as well. Neither more effective insulated-box technology nor less appropriate ones, their box temperature lines as belonged to insulated materials (foam, rice husk, rice straw, plastic PVC, cement, earth soil, and wood) technology have met together at about 265 minutes after pouring the hot river gravels as shown in Figure 3. In other words, every studied insulated material could be appropriate to use for insulated-material technology.

When the relationships between Q and time t in equation (5) was calculated under real time measurement of in-box temperature (T1) and ambient air temperature (T2) by fixing the values of X (thickness) equivalent to 2.5 cm, the relationships between Q and t by graphical techniques of foam, rice straw, rice husk, cement, plastic PVC, earth soil, foam, and wood were constructed as illustrated in Figure 4 and heat transfer values as shown in Table 4.

Table 4. Heat transfer of insulated boxes as used for evaluating the appropriate insulated materials

Times (minutes)	Q, heat transfer (cal/cm)						
	Rice Straw	Rice Husk	Cement	Plastic,PVC	Earth Soil	Foam	Wood
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	803.0	731.5	1000.6	279.7	2653.6	197.3	447.2
20	1417.9	1284.5	1614.2	480.3	4513.6	322.6	773.8
30	1723.7	1598.4	1925.8	574.6	5133.6	373.0	904.8
40	2029.4	1889.3	2303.4	674.9	6249.6	410.9	1106.6
50	2318.4	2059.2	2572.4	752.4	6882.0	412.8	1227.2
60	2540.2	2177.3	2832.0	839.0	7812.0	406.1	1385.3
70	2704.8	2378.9	2973.6	925.7	8246.0	403.2	1514.2
80	2768.6	2442.2	2945.3	936.3	7936.0	372.5	1514.2
90	3326.4	2903.0	3653.3	1190.2	9709.2	393.1	1853.3
100	3662.4	3139.2	3870.4	1292.0	9796.0	345.6	1976.0
110	3696.0	3199.7	3790.2	1320.9	9957.2	332.6	1944.8
120	3830.4	3317.8	3681.6	1386.2	10564.8	293.8	2046.7
130	3669.1	3145.0	3497.5	1323.9	9833.2	280.8	1919.8
140	3622.1	3024.0	3370.1	1276.8	9721.6	248.6	1834.6

150	3326.4	2894.4	3044.4	1208.4	8556.0	216.0	1684.8
160	3279.4	2857.0	3020.8	1167.4	8332.8	192.0	1630.7
170	3255.8	2888.6	2968.9	1162.8	8010.4	155.0	1591.2
180	3265.9	2799.4	2888.6	1176.5	7812.0	138.2	1572.5
190	3064.3	2681.3	2780.1	1097.4	6832.4	118.6	1462.2
200	2822.4	2419.2	2548.8	1003.2	5704.0	96.0	1289.6
210	2751.8	2419.2	2577.1	989.5	5728.8	90.7	1223.0
220	2365.4	2090.9	2180.6	836.0	4637.6	73.9	1006.7
230	2550.2	2252.2	2388.3	909.0	4848.4	88.3	1100.3
240	2257.9	2004.5	2039.0	802.6	4166.4	57.6	948.5
250	2268.0	2016.0	2124.0	836.0	4340.0	48.0	988.0
260	2009.3	1797.1	1963.5	711.4	3868.8	37.4	811.2

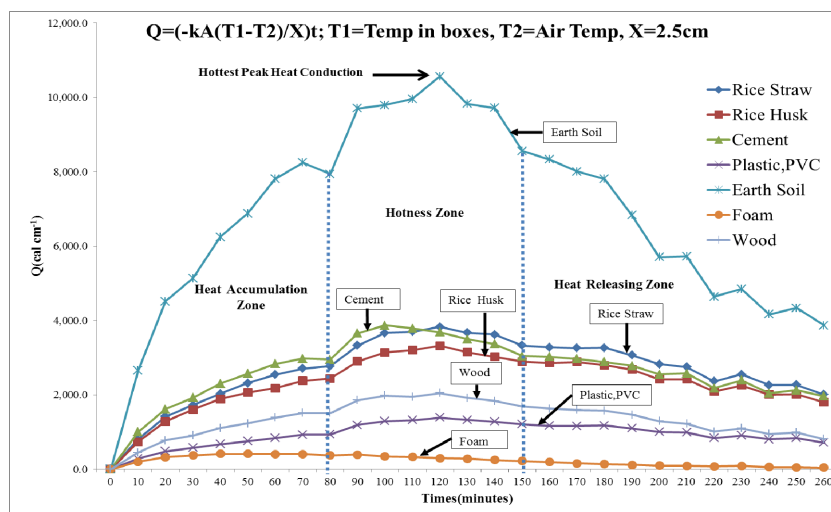


Figure 4. Amount of heat transfer through Fourier's law of heat transfer under the fixed values of insulated box thickness (2.5 cm), inside box temperature, and averaged k-values of each insulated materials, surface area ($A = 900$ squared cm), and varying with time (minutes)

The relationship between amount of heat transfer (Q) and measuring time (t) which was constructed (as shown in Figure 4) can be gathered into 3 groups of insulated groups; they are low (wood, plastic PVC, and foam), medium (cement, rice husk, and rice straw), and high (earth soil) heat transferring by conduction process. Accordance with more minerals in composition, the highest heat conduction was placed on earth soil that should be ignored in using for assembling the insulated boxes. While the lowest heat conduction is identified in wood, plastic PVC, and foam but it might be risky to create the environmental problems even they are so convenient to get in everywhere. Finally, the medium heat conduction group is the most appropriate insulated materials (cement, rice husk, rice straw) to be selected for molding insulated-material sheets for assembling the insulated boxes. If the convenience is concerned with the local materials available, the rice straw should be the best one because it can be compensated by local grasses as grow everywhere in the whole country.

It is obvious that the graphical techniques in Figure 4 can be divided into 3 zones; first zone is called as "heat accumulation zone" beginning with rising limb in order to accumulate the heat storage for about 80 minutes; second zone called as "hotness zone" it starts transferring heat from "heat accumulation zone" with the energetic conduction before entering "hotness zone" between 80 - 150 minutes (70 minutes) for maintaining the maximum hottest period to digest some more cause textural organic wastes. Evidently, the heat was transferred from the hotness zone throughout the 'heat releasing zone' until the ending of bacterial organic digestion process around 265 minutes (4 hours and 25 minutes) and plus (see Figure 3). However, the period from the beginning of organic digestion process to hottest-peak heat conduction at the 120 minutes (2 hours) before gradually heat liberating more than twice. In reality, the hottest peak heat conduction was placed on around the fifth days (in-box temperature at 77°C) and heat liberating period up to 25 days for converting the organic wastes to

become the mature organic compost (Figures 4 and 5) through the bacterial organic digestion process (LERD 1999, 2000, 2011, 2012; Shahriari et al., 2013; Appels et al., 2010; Steger et al., 2007; Fernandez-Rodriguez et al., 2014; Zupancic & Ros, 2003; Metcalf & Eddy, 1972; Ariunbaatar et al., 2014; JICA, 1982; Atchley & Clark, 1979; Ugwuanyi et al., 2005; Juteau, 2006).

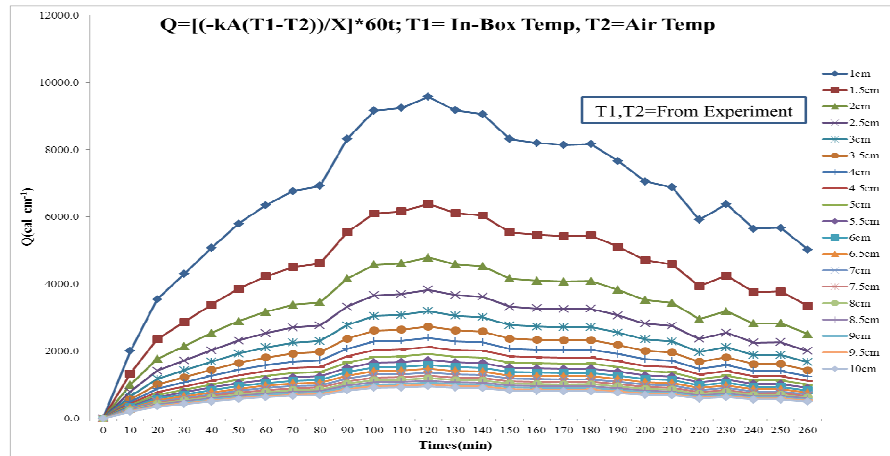


Figure 5. Systematic appearances of transferring heat in various times of "heat accumulation zone", "hotness zone" and "heat releasing zone" in relation to various rice-straw insulated thicknesses

3.3 Box Insulated Thickness Determination

As finalized before and shown in Figure 3, the rice straw was selected to as appropriate insulated materials for assembling the insulated box technology under the fixed thickness of 2.5 cm because of its minimum thickness to make sheet insulators, the amount of heat transfer (Q) in relation to another thickness sizes were calculated as illustrated in Figure 5. It was obviously seen that no matter the insulated thickness of any sizes, the systematic appearances of "heat accumulation zone", "hotness zone", and "heat releasing zone" were somewhat the same as illustrated in Figure 4. In other words, the thickness sizes of rice straw insulators cannot change in their systematic appearances even its thickness changed in which the Figure 5 can be useable for selecting the appropriate thickness. In contrary, it was so difficulty to make decision what thickness of rice-straw insulated box should be the best one. Roughly, the minimum rice-straw insulated box should be placed at 6 cm due to the Q-t line looked steadier rather than Q-t lines 1 cm to 5.5 cm, but this result could not make sure to the decision makers. Therefore, the Figure 6 was constructed in order to emphasize the aforesaid decision making in which the Q-thickness lines of each ambient-air-temperature line by fixing 70 °C as the in-box temperature seemed agreeable at the 6-cm thickness but it was still confused to the decision makers due to the existence of less sloping characters.

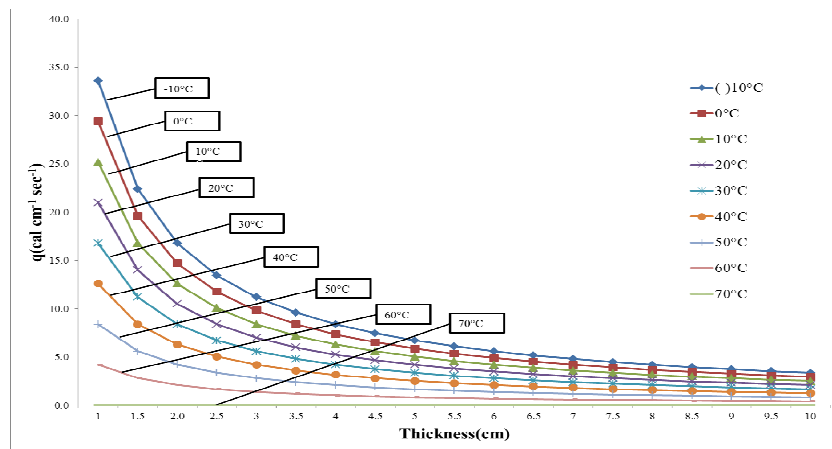


Figure 6. Determining appropriate thickness of rice straw as insulated material to assemble the insulated box technology for elongating solid retention time (SRT) to enhance the well-fermented community garbage to become the fruitful compost

Finally, the relationship between heat conduction (Q) and rice-straw insulated thickness through ONE-WAY ANOVA statistical method which uses for its stratification as shown in Table 5. The similarity of numbers in small frames was indicated the coordination between the rice-straw insulated thickness (Thickness) and ambient air temperature (T2), for example, if T equivalent to -10 °C, Thickness equivalent to 9.5 cm to 10 cm; 30 °C of T going hand in hand with Thickness of minimum 6 cm to maximum 10 cm. therefore, the Table5 is applicable to give the ambient air temperature and reading downwards to the T2-Thickness frames before reading the thickness column in order to get the appropriate rice-straw insulated box for bacterial organic digestion process by controlling in-box temperature at 70 °C. By this method, the means how to find out the rice-straw appropriate insulated box can exactly be applicable not only one ambient-air-temperature condition but also to extend to anywhere with different local temperature to coordinate with the appropriate thickness. However, the minimum thickness in Thailand should be placed on 6 cm. expectedly, this minimum rice-straw insulated box can be applicable in all parts of the country, either lowland, flatlands, mountainous lands, and islands.

Table 5. Calculated frame values between heat conduction (Q) and rice-straw insulated thickness by fixing in-box temperature at 70 °C through ONE-WAY ANOVA statistical method (level of significant 95%) for finding the appropriate thickness of insulated boxes

Thickness (cm)	Air Temperature, T2 (°C)								
	-10°C	0°C	10°C	20°C	30°C	40°C	50°C	60°C	70°C
1.0	33.6 ^r	29.4 ^s	25.2 ^r	21.0 ^q	16.8 ^p	12.6 ^q	8.4 ^p	4.2 ^l	0 ^a
1.5	22.4 ^q	19.6 ^r	16.8 ^q	14.0 ^p	11.2 ^o	8.4 ^p	5.6 ^o	2.8 ^k	0 ^a
2.0	16.8 ^p	14.7 ^q	12.6 ^p	10.5 ^o	8.4 ⁿ	6.3 ^o	4.2 ⁿ	2.1 ^j	0 ^a
2.5	13.4 ^o	11.8 ^p	10.1 ^o	8.4 ⁿ	6.7 ^m	5.0 ⁿ	3.4 ^m	1.7 ⁱ	0 ^a
3.0	11.2 ⁿ	9.8 ^o	8.4 ⁿ	7.0 ^m	5.6 ^l	4.2 ^m	2.8 ^l	1.4 ^h	0 ^a
3.5	9.6 ^m	8.4 ⁿ	7.2 ^m	6.0 ^l	4.8 ^k	3.6 ^l	2.4 ^k	1.2 ^g	0 ^a
4.0	8.4 ^l	7.4 ^m	6.3 ^l	5.3 ^k	4.2 ^j	3.2 ^k	2.1 ^j	1.1 ^f	0 ^a
4.5	7.5 ^k	6.5 ^l	5.6 ^k	4.7 ^j	3.7 ⁱ	2.8 ^j	1.9 ⁱ	0.9^e	0 ^a
5.0	6.7 ^j	5.9 ^k	5.0 ^j	4.2 ⁱ	3.4 ^h	2.5 ⁱ	1.7 ^h	0.8^{de}	0 ^a
5.5	6.1 ⁱ	5.3 ^j	4.6 ⁱ	3.8 ^h	3.1 ^g	2.3 ^h	1.5 ^g	0.8^{de}	0 ^a
6.0	5.6 ^h	4.9 ⁱ	4.2 ^h	3.5 ^g	2.8^f	2.1 ^g	1.4 ^{fg}	0.7^{cd}	0 ^a
6.5	5.2 ^g	4.5 ^h	3.9 ^g	3.2 ^f	2.6^{ef}	1.9 ^f	1.3 ^{ef}	0.6^{bc}	0 ^a
7.0	4.8 ^f	4.2 ^g	3.6 ^f	3.0 ^e	2.4^{de}	1.8 ^{ef}	1.2 ^{de}	0.6^{bc}	0 ^a
7.5	4.5 ^e	3.9 ^f	3.4 ^e	2.8 ^d	2.2^{cd}	1.7 ^{de}	1.1 ^{cd}	0.6^{bc}	0 ^a
8.0	4.2 ^d	3.7 ^e	3.2 ^d	2.6 ^c	2.1^{bcd}	1.6 ^{cd}	1.1 ^{cd}	0.5^{ab}	0 ^a
8.5	4.0 ^c	3.5 ^d	3.0 ^c	2.5 ^c	2^{abc}	1.5 ^{bc}	1.0 ^{bc}	0.5^{ab}	0 ^a
9.0	3.7 ^b	3.3 ^c	2.8^a	2.3^b	1.9^{ab}	1.4 ^{ab}	0.9 ^{ab}	0.5^{ab}	0 ^a
9.5	3.5^a	3.1 ^b	2.7^a	2.2^{ab}	1.8^a	1.3 ^a	0.9 ^{ab}	0.4^a	0 ^a
10.0	3.4^a	2.9 ^a	2.5 ^b	2.1^a	1.7^a	1.3 ^a	0.8 ^a	0.4^a	0 ^a

3.4 Applicability of Rice-Straw Insulated Box Technology

Referring to the previous study of The King's Royally Initiated Laem Phak Bia Environmental Research and Development (Royal LERD) Project in Phetchaburi Province, Thailand through nature-by-nature process, the concrete box technology has been mainly used for community garbage disposal and also for innovating the new applicable technique in some constraint areas as shown in Figure7. The matured compost was obtained after fermenting 30 days (SRT equivalent to 30 days) for separating organic wastes and 90 days (SRT equivalent to 90 days) for unseparated community (municipal) garbage. It was pointed out that the separation of organic wastes from the whole proportion of community solid wastes could make the shorter SRT from 90 days to become 30 days. Consequently, the compost texture was composed of cause texture of some solid component which could be identified mostly as fibers but they were excellently useable for fertilizing to plant growth after cutting in tiny

pieces.

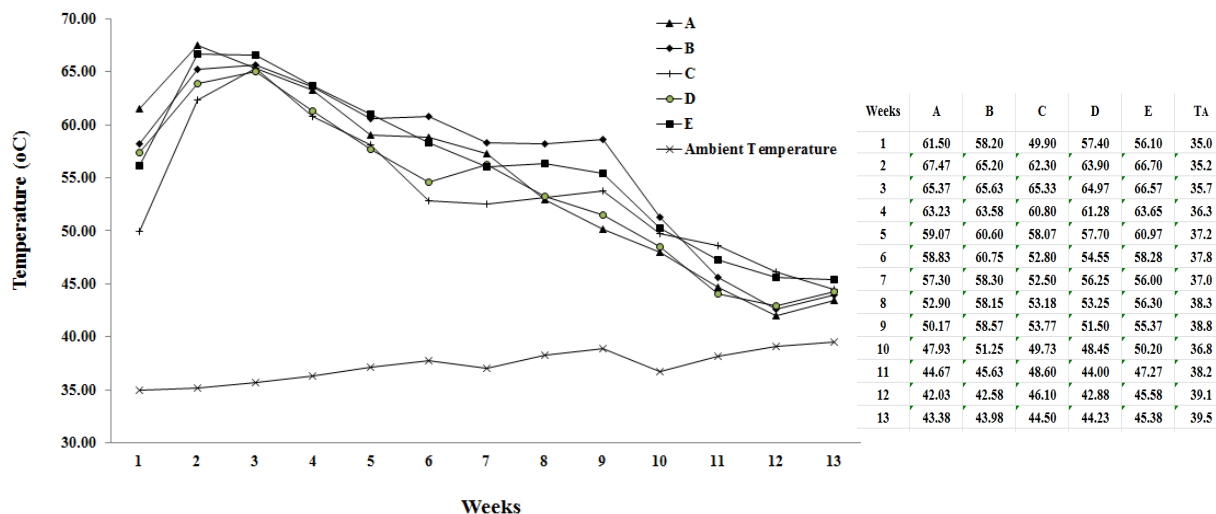


Figure 7. Consecutive temperature as measured during fermenting community garbage inside box with the size of 1.5 m height, 2.0 m width, and 3.0 m length for 2 tons garbage accommodation accordance with fermenting time 30 days for only separated organic wastes and 90days for unseparated community garbage

Note. Treatment A: The unseparated community garbage pile left to be decomposed naturally.

Treatment B: The pile with glass, metal and plastic scraps sorted out, then spread over with compost starter, phosphate supplement, urea and loose soil.

Treatment C: The pile with glass, metal and plastic scraps sorted out, then mixed with compost starter, phosphate supplement, urea and loose soil.

Treatment D: The unseparated community garbage pile spread over with compost starter, phosphate supplement, urea and loose soil.

Treatment E: The unseparated community garbage pile mixed with compost starter, phosphate supplement, urea and loose soil.

In practical point of view, there was 30% hollow in each concrete box (3-cm thickness) and without insulating materials for reduction of transferring heat (as liberated from bacterial organic digestion process) which used for assembling the concrete box technology. Nevertheless, the temperature was still risen in rapid rate in the heat accumulation zone, and also it still met the peak temperature inside hotness zone close to the averaged values of 70 °C in the hotness zone. In reality, there were some experimental concrete boxes found the peak temperature up to 69 °C during fermenting the Phetchaburi municipal solid wastes. Besides, the heat transfer was taken in account with quite low rate in the heat releasing zone (see Figures 7 and 8) due to the outside concrete box above 35 °C that caused low net re-radiation rate from naturally heat liberation from bacterial organic digestion process inside experimental boxes. Although the peak temperature was more or less 70 °C, the hotness zone in Figure 7 found quite narrow which would be shorter period of heat storage inside concrete boxes. If the 4-side concrete box were strict by 2.5-cm thickness of rice-straw insulated sheets (k equivalent to 4.7×10^{-4} cal/cm²/°C/sec) the in-box temperature should be less than one fifth of heat conduction from bacterial organic digestion process in the concrete boxes (k equivalent to 2.3×10^{-3} cal/cm²/°C/sec). It could be concluded that the rice straw plays vital role in keeping heat as liberated from the digestion process of organic wastes rather than without rice straw insulated sticking in concrete box walls.

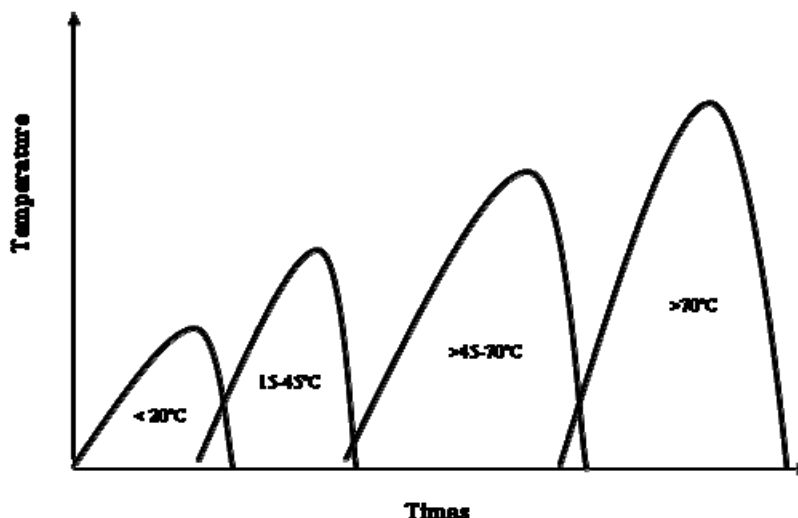


Figure 8. Hypothetically consecutive temperature rising during bacterial organic digestion processes on carbohydrate protein celluloses hemicellulose and fibers

Theoretically speaking, the organic wastes of Phetchaburi municipal garbage have evidently converted to become inorganic-material garbage compost through aerobically bacterial organic digestion process for 30 day SRT (LERD, 1999, 2000, 2011, 2012). Then after, the garbage compost was produced by constructing the 9 m³ concrete box (1.5 m deep, 2 m wide, 3 m long) for 5 treatments 3 replications on slope of 1:1,000; compacting before paving gravel at the level above 2 cm diameter PVC pipe for leachate draining, then overtopping on 2 cm height of sand; putting 3 alternated layers of two level of 670 kg organic wastes along with overtopping 5 cm height of clayed soils (called electron acceptors), and the top layer of 670 kg organic wastes with overtopping 15 cm of clayed soils. Amount of water which is plenty of dissolved oxygen (DO) about 60 liters will be showered every seven days in order to bring down the electron acceptors (such as O₂, Fe³⁺, Mn²⁺, etc.) to catch with the electrons as broke from bacterial organic digestion process to form the inorganic materials in terms of chemical compounds (Prabuddham, 1985) as illustrated in Figure 7.

This would be stressed that the releasing heat from bacterial organic digestion process was kept longer period because of the smaller temperature gradient between inside-outside concrete box which caused low heat transferring by re-radiation. In contrary, concrete box may become heat acceptor whenever the surrounding temperature is higher than inside concrete box. Naturally, it is inevitable to stop transferring the releasing heat from bacterial organic digestion process in box of CB technology due to contact between inner and outer surfaces. Unavoidably, the hot temperature inner concrete box has been gradually decreased due heat conduction transfer to outer surface until temperature of both sides was in balancing condition. Exactly, the insulated-covered materials on concrete box surfaces are not preferable to use for reduction of heat transfer to outer ambient air due to high thermal conductivity of concrete boxes. There are a lot of suggestions to use insulated materials for constructing the insulated box technology instead of single concrete box technology.

Undoubtedly, if the inner concrete box was assembly insulated by the effective materials, no matters rice straw and others (rice husk, wood, plastic PVC, earth soil, and cement together with another cementing agent) might be enhanced to elongate the hotness zone including peak temperature increasing more than 70 °C, and also high rate of bacterial organic digestion process. The expectation of better quality of community garbage compost would be obtained one way or another.

4. Conclusion

The research is aimed to select the insulated materials in which they were taken in pre-selection of foam, earth soil, rice husk, rice straw, plastic PVC, and cement. Those insulated materials were laminated in frame with the size of 30 cm width, 30 cm length and 2.5 cm thickness (22.5 cm³/sheet) in order to use for assembling the insulated-cube box (30x30x30 cm³ or 0.027 m³). Then after, the density, specific gravity, and averaged thermal conductivity were determined under the laboratory basis. For serving research needs, the amount of 500 g river gravels (number 3) was heated in Hot air oven at 100 °C for 24 hours before putting in the insulated boxes. The in-box (T1) and out-box (T2) ambient air temperature were recorded by automatic data logger from beginning to

the end of experiments. The relationship between time (t) and in-box temperature by graphical techniques in order to select the appropriate insulated material under the criteria of The King's Royally initiative nature-by-nature process, simplicity technology, and low expense (or local materials using for constructing technology). Also, the relations between Q (heat conduction) of temperature differences (in-box and outside box) in varying t (time) as the same as the insulated material thickness was evaluated from graphical products of fixing T (equivalent to 70 °C) and Q of varying ambient air temperature.

The results found that the rice straw looked better appropriate insulated material rather than rice husk, earth soil, cement, foam, and plastic PVC. Nevertheless, the minimum thickness could be placed on 6 cm in which the in-box temperature could be elongated period for psychrophiles, mesophiles, thermophiles, and hyper-thermophile to complete the digestion of carbohydrates, proteins, celluloses, hemicelluloses, and fibers. So far, the relation between the ambient air temperature and rice-straw insulated thickness that brought to direct how to obtain the insulated thickness in case of ambient air temperature to be identified its level.

Acknowledgement

This research received financial supports from The King's Royally Initiated Laem Phak Bia Environmental Research and Development Project, Chaipattana Foundation, Thailand.

References

- Ahmed, S. A., & Ali, M. (2004). Partnerships for solid waste management in developing countries: Linking theories to realities. *Habitat International*, 28(3), 467-479. [http://dx.doi.org/10.1016/S0197-3975\(03\)00044-4](http://dx.doi.org/10.1016/S0197-3975(03)00044-4)
- Appels, L., Degreve, J., Van der Bruggen, B., Impe, J. V., & Dewil, R. (2010). Influence of low temperature thermal pre-treatment on sludge solubilisation, heavy metal release and anaerobic digestion. *Bioresource Technology*, 101(15), 5743-5748. <http://dx.doi.org/10.1016/j.biortech.2010.02.068>
- Ariunbaatar, J., Panico, A., Esposito, G., Pirozzi, F., & Lens, P. N. L. (2014). Pretreatment methods to enhance anaerobic digestion of organic solid waste. *Applied energy*, 123, 143-156. <http://dx.doi.org/10.1016/j.apenergy.2014.02.035>
- Atchley, S. H., & Clark, J. B. (1979). Variability of temperature, pH, and moisture in an aerobic composting process. *Applied and Environmental Microbiology*, 38(6), 1040-1044. <http://aem.asm.org/content/38/6/1040>
- Bhuiyan, S. H. (2010). A crisis in governance: Urban solid waste management in Bangladesh. *Habitat International*, 34(1), 125-133. <http://dx.doi.org/10.1016/j.habitatint.2009.08.002>
- Botkin, D. B., & Keller, E. A. (2010). *Environmental Science: Earth as a Living Planet* (8th ed.). New York, NY: John Wiley and Sons, Inc.
- Chang, Y. H., & Chang, N. (1998). Optimization analysis for the development of short-term solid waste management strategies using presorting process prior to incinerators. *Resources, Conservation and Recycling*, 24(1), 7-32. [http://dx.doi.org/10.1016/S0921-3449\(98\)00036-6](http://dx.doi.org/10.1016/S0921-3449(98)00036-6)
- Chunkao, K. (1979). *Micrometeorology*. Thailand: Department of Conservation, Faculty of Forestry, Kasetsart University, Bangkok, 147 p.
- Chunkao, K. (1998). *Environmental Science*. Thailand: Kasetsart University Publishing, 784 p.
- Fernandez-Rodriguez, J., Perez, M., & Romero, M. P. (2014). Dry thermophilic anaerobic digestion of the organic fraction of municipal solid wastes: Solid retention time optimization. *Chemical Engineering Journal*, 251, 435-440. <http://dx.doi.org/10.1016/j.cej.2014.04.067>
- Finnveden, G., Albertsson, A., Berendson, J., Eriksson, E., Hoglund, L. O., Karlsson, S., & Sundqvist, J. (1995). Solid waste treatment within the framework of life-cycle assessment. *Journal of Cleaner Production*, 3(4), 189-199. [http://dx.doi.org/10.1016/0959-6526\(95\)00081-X](http://dx.doi.org/10.1016/0959-6526(95)00081-X)
- Gates, D. M. (1965). *Energy Exchange in the biosphere*. New York, NY: Harber & Brothers publisher.
- Gates, D. M. (1980). *Biophysical Ecology*. Mineola, NY: Springer-Verlag New York, Inc.
- Geng, Y., Tsuyoshi, F., & Chen, X. (2010). Evaluation of innovative municipal solid waste management through urban symbiosis: a case study of Kawasaki. *Journal of Cleaner Production*, 18(10-11), 993-1000. <http://dx.doi.org/10.1016/j.jclepro.2010.03.003>
- Hao, Y., Wu, W., Wu, S., Sun, H., & Chen, Y. (2008). Municipal solid waste decomposition under over-saturated condition in comparison with leachate recirculation. *Process Biochemistry*, 43(1), 108-112.

- [http://dx.doi.org/ 10.1016/j.procbio.2007.10.004](http://dx.doi.org/10.1016/j.procbio.2007.10.004)
- Hartman, D. L. (1994). *Global Physical Climatology*. San Diego, CA: Academic Press.
- Holman, J. P. (2009). *Heat Transfer* (10th ed.). New York, NY: McGraw-Hill.
- JICA. (1981). *The Bangkok Solid Waste Management Study in Thailand: Interim Report*, Bangkok, Thailand: Japan International Cooperation Agency. 126 p.
- JICA. (1982). *The Bangkok Solid Waste Management Study in Thailand: Interim Report*, Bangkok, Thailand: Japan International Cooperation Agency. 145 p.
- JICA. (1991). *The Study on Bangkok Solid Waste Management*. Bangkok, Thailand: Japan International Cooperation Agency. 205 p.
- Juteau, P. (2006). Review of the use of aerobic thermophilic bioprocesses for the treatment of swine waste. *Livestock Science*, 102(3), 187-196. [http://dx.doi.org/ 10.1016/j.livsci.2006.03.016](http://dx.doi.org/10.1016/j.livsci.2006.03.016)
- Keenen, J. D., Steiner, R. L., & Fungaroli, A. A. (1984). Landfill leachate treatment. *Journal of Water Pollution Control Federation*, 56(1), 27-33.
- Koroneos, C. J., & Nanaki, E. A. (2012). Integrated solid waste management and energy production - a life cycle assessment approach: the case study of the city of Thessaloniki. *Journal of Cleaner Production*, 27, 141-150. [http://dx.doi.org/ 10.1016/j.jclepro.2012.01.010](http://dx.doi.org/10.1016/j.jclepro.2012.01.010)
- LERD. (1999). *Economized Technology for Community Garbage Disposal and Wastewater Treatment by Aquatic Plants*. Phetchaburi, Thailand: The King's Royally Initiated Laem Phak Bia Environmental Research and Development Project, 420 p.
- LERD. (2000). *Sciences for Garbage Disposal and Wastewater Treatment Towards The King's Initiative Project*. Technical Paper for 10-year LERD Seminar, Organized by Chaipatana Foundation, during 24-25 August 2000 at Kasetsart University, Bangkok, Thailand.
- LERD. (2011). *LERD Annual Report*, Thailand: The King's Royally Initiated Laem Phak Bia Environmental Research and Development Project, (Serial numbers in the years of 2005, 2006, 2007, 2008, 2009, 2010, 2012).
- LERD. (2012). *Technology for wastewater treatment and organic waste composting under nature-by-nature process*. Thailand: The King's Royally Initiated Laem Phak Bia Environmental Research and Development Project, 64 p.
- Metcalf & Eddy Inc. (1972). *Wastewater engineering: collection, treatment, disposal*. New York, NY: McGraw-Hill Publishing Company Ltd., 782 p.
- Nopparatanaporn, N. (1992). *Microorganisms in Wastewater*. Bangkok, Thailand : Department of Microbiology, Faculty of Science, Kasetsart University Publishing, 412 p.
- Okot-Okumu, J., & Nyenje, R. (2011). Municipal solid waste management under decentralization in Uganda. *Habitat International*, 35(4), 537-543. [http://dx.doi.org/ 10.1016/j.habitatint.2011.03.003](http://dx.doi.org/10.1016/j.habitatint.2011.03.003)
- Phetchaburi Municipal. (2011). *Annual Report of Phetchaburi Municipal year 1992-2011*. Phetchaburi Thailand, 25 p.
- Prabuddham, P. (1985). *Soil Chemistry*. Bangkok, Thailand : Department of Soil Science, Faculty of Agriculture, Kasetsart University, 502 p.
- Sellers, W. D. (1965). *Physical Climatology*. Chicago: The University of Chicago Press, 272 p.
- Shahriari, H., Warith, M., Hamoda, M., & Kennedy, K. (2013). Evaluation of single vs. staged mesophilic anaerobic digestion of kitchen waste with and without microwave pretreatment. *Journal of Environmental Management*, 125, 74-84. [http://dx.doi.org/ 10.1016/j.jenvman.2013.03.042](http://dx.doi.org/10.1016/j.jenvman.2013.03.042)
- Steger, K., Jarvis, A., Vasara, T., Romantschuk, M., & Sundh, I. (2007). Effects of differing temperature management on development of Actinobacteria populations during composting. *Research in Microbiology*, 158(7), 617-624. [http://dx.doi.org/ 10.1016/j.resmic.2007.05.006](http://dx.doi.org/10.1016/j.resmic.2007.05.006)
- Swati, M., & Joseph, K. (2008). Settlement analysis of fresh and partially stabilized municipal solid waste in simulated controlled dumps and bioreactor landfills. *Waste Management*, 28(8), 1355-1363. [http://dx.doi.org/ 10.1016/j.wasman.2007.06.011](http://dx.doi.org/10.1016/j.wasman.2007.06.011)
- Tchobanoglous, G., Theisen, H., & Vigil, S. A. (1993). *Integrated Solid Waste Management, Engineering*

Principles and Management Issues. New York, NY: McGraw-Hill higher education Publisher.

- Thapa, G. B. (1998). Lessons Learned from Solid Waste Management in Kathmandu, Nepal. *Habitat International*, 22(2), 97-114. [http://dx.doi.org/10.1016/S0197-3975\(97\)00030-1](http://dx.doi.org/10.1016/S0197-3975(97)00030-1)
- Ugwuanyi, J. O., Harvey, L. M., & McNeil, B. (2005). Effect of digestion temperature and pH on treatment efficiency and evolution of volatile fatty acids during thermophilic aerobic digestion model high strength agricultural waste. *Bioresource Technology*, 96(6), 707-719. <http://dx.doi.org/10.1016/j.biortech.2004.06.027>
- Wilson, D. G. (1977). *Handbook of solid waste management*. Michigan: Van Nostrand Reinhold Co., 752 p.
- Zupancic, G. D., & Ros, M. (2003). Heat and energy requirements in thermophilic anaerobic sludge digestion. *Renewable Energy*, 28(14), 2255-2267. [http://dx.doi.org/10.1016/S0960-1481\(03\)00134-4](http://dx.doi.org/10.1016/S0960-1481(03)00134-4)

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

Assessment of Toilet's Indoor Air Quality in Relation to Asthmatic People

Ismail Abdul Rahman¹, Jouvan Chandra Pratama Putra¹ & Ade Asmi²

¹ Department of Construction and Architecture, Faculty of Civil and Environmental Engineering, Universiti Tun Hussein Onn Malaysia, Malaysia.

² Civil Engineering Department, Bakrie University, Jakarta, Indonesia

Correspondence: Jouvan Chandra Pratama Putra, Department of Construction and Architecture, Faculty of Civil and Environmental Engineering, Universiti Tun Hussein Onn Malaysia, Malaysia. E-mail: jouvanchandra@gmail.com

Received: June 17, 2014

Accepted: June 19, 2014

Online Published: September 27, 2014

doi:10.5539/mas.v8n5p289

URL: <http://dx.doi.org/10.5539/mas.v8n5p289>

Abstract

This paper presents an assessment on toilet's indoor air quality on asthmatic people. The assessment was carried out by measuring temperature, relative humidity, air velocity and SO₂ in the 4 selected toilets which are naturally ventilated. The measurements used two equipments that are *Yes Plus LGA meter and Anemometer*. It was done in early morning and break hours for 10 minutes within three (3) days consecutively. This study found that SO₂ concentration in toilet 1 exceeded the threshold limit value of SO₂ for 10 minutes exposure which means this toilet potentially trigger asthmatic symptom for those asthmatic people who frequently used this toilet. High concentration of SO₂ was due to inadequacy of air velocity in this toilet. Inadequacy of air velocity means less ventilation and thus, increased the humidity which can stimulate the growth of fungi/mold in the toilet that affects asthmatic people.

Keywords: asthma, indoor air quality, so₂, toilet

1. Introduction

Indoor Air Quality (IAQ) is a terminology of air quality within and around buildings which pertinent with health and comfort of indoor occupant (Muhamad-darus et al., 2011; Murdoch & Lloyd, 2010). The duration of time that people spent indoor and the risk to health posed by indoor air are the most issue that encourage researchers to conduct the study on indoor air (Frontczak & Wargocki, 2011; Lee, 1997; Massey et al., 2009). In addition, several findings of IAQ studies from previous researchers indicated that coarse particulate matter (PM₁₀) which is less than 10 micrometers in diameter (Brunekreef & Forsberg, 2005; Donaldson et al., 2000; Karakatsani et al., 2012; Zhang, 2005), fine particulate matter (PM_{2.5}) which is less than 2.5 micrometers in diameter (Brunekreef & Forsberg, 2005; Karakatsani et al., 2012) and Sulfur Dioxide (SO₂) (Awbi, 1991; Guo, 2012; Luttinger & Wilson, 2003; Restrepo, 2012; Tseng et al., 2012) in particular concentration could lead adverse effect for human inhalation such as asthmatic symptoms. However, the exposure of SO₂ on human inhalation for asthmatic people in a concise time at particular concentration gives more significant effect to trigger the asthmatic symptoms as compared to PM₁₀ and PM_{2.5} (World Health Organization, 1987).

Asthma is a heterogeneous disorder of the conducting airways involving chronic inflammation, declining function and tissue remodeling (Murdoch & Lloyd, 2010). For asthmatic individuals, the inflammation causes wheezing, breathlessness, chest tightness and cough particularly at night and/or early morning (Bousquet et al., 2000). There are various researches on the effect of SO₂ exposure to asthmatic people. Koren (1995) had exposed asthmatic people within 20 minutes of SO₂ at 0.25 ppm and found it affect to the bronchoconstriction function which finally lead to asthmatic symptom. In other situation, a laboratory experiment was conducted which involves 7 asthmatic people exposed to 0.5 ppm of SO₂ for 10 minutes and resulted to 3 people having wheezing and breathlessness (Sheppard et al., 1980). Furthermore, Balmes et al., (1987) had carried out a study on the effect of exposure to a humidified air for 5 minutes and SO₂ in the range of 0.5 – 1 ppm for 1,3 and 5 minutes on 2 females and 6 males of non-smoking asthmatic adult. They found that bronchoconstriction function increase the Specific Airway Resistance (SRaw) above baseline with respect to the exposure time and

concentration. Additionally, the other asthmatic triggers are smoke and humidity (Al-anzi & Salman, 2011). Smoke is a mixture of smoke from the burning end of a cigarette, pipe or cigar and it exhaled by the smoker that is often found in homes, toilets and car where smoking is allowed. Besides, humidity can affect to the growth of mold and commonly found in the toilet, kitchen and the basement (Al-anzi & Salman, 2011)

Based on the findings of previous studies which had been described above, it can be concluded that the exposure of SO₂ in concise time is the main problem that could have effect on human inhalation. Subsequently, toilet in university is one of public facilities which frequently used either by staff or student in concise time. Thus, it motivates to assess its IAQ level in order to compare with the Threshold Limit Value (TLV).

2. Method

A. Site Description

This study was conducted on 4 selected toilets at Faculty of Civil and Environmental Engineering of Universiti Tun Hussein Onn Malaysia. The selected toilets (2 for gents and 2 for women) are naturally ventilated through opening of windows. Table 1 describes the characteristics of selected toilets.

Table 1. Characteristics of the selected toilets

Toilet Number	Toilet Type	Number of Users During Measurement (person)						Area (m ²)	Height (m)
		Morning			Noon Time				
		1	2	3	1	2	3		
1	Male	6	4	2	4	2	1	10.12	2.8
2	Female	4	3	2	4	2	3	10.12	2.8
3	Male	1	4	1	1	5	3	10.12	2.8
4	Female	3	5	1	1	2	3	10.12	2.8

B. Sample Collection and Analysis

The measurements were conducted twice a day (in early morning and noon during break hours) for 10 minutes within 3 consecutive days. The measured parameters are air temperature, relative humidity and SO₂ concentration which are measured by using *Yes Plus LGA Meter* and air velocity by using *Anemometer*. These equipments were placed vertically at a height of 1.2 m from the floor. During the measurement periods, the selected toilets function as usual with several users as in table 1. All the data gathered are analysed using descriptive statistical approach which includes mean, maximum and minimum values calculated using Microsoft Excel and presented into graph and bar chart.

3. Results and Discussion

The parameters of IAQ considered in this study are temperature, relative humidity, air velocity and SO₂ concentration. The measured values of all the parameters and the analysis on the data are presented in this section

3.1 Temperature and Relative Humidity

The temperature and humidity are measured during 10 minutes duration in 3 days continuously in all the 4 toilets involved. The average values of temperature are presented as in figures 1 and 2 while relative humidity is presented as in figures 3 and 4. Figure 1 shows the highest recorded temperature in the morning is 30 °C in the toilet 1 or toilet 2 at 2nd day of and also in the toilet 1, toilet 2, or toilet 4 at 3rd day of the measurement. For the lowest recorded temperature is 29 °C at 1st day of the morning measurement in toilet 1, toilet 2, toilet 3 and toilet 4. Subsequently, for temperatures at noon time, the highest recorded value is 31 °C at the toilet 4 at 2nd day of noon time measurement and at the toilet 2 at 3rd day of the measurement. For the lowest recorded temperature at noon time is 28.5 °C which is at the toilet 1 at 1st day of the measurement.

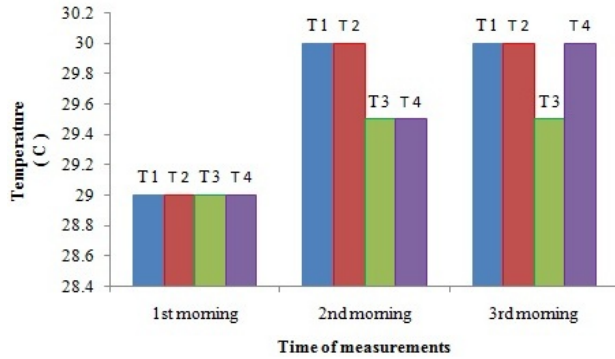


Figure 1. Measured morning temperature (toilet)

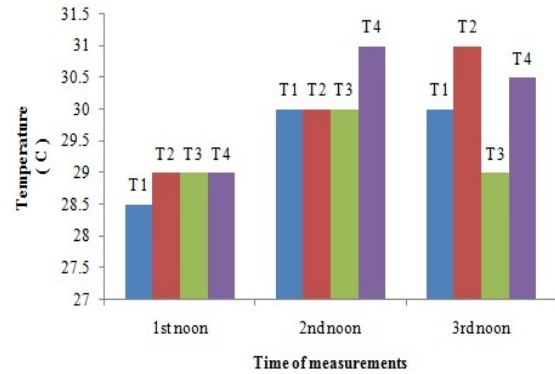


Figure 2. Measured noon temperature (toilet)

For the morning measurement as in figure 3, the highest average relative humidity measured is at 1st day of toilet 1 or toilet 2 with the value of 94.5%. For the lowest case, it was achieved at the 3rd day with the value of 82.5% in toilet 4. For the noon measurement as in figure 4, the highest value is at the 1st day with the value of 91.5% for toilet 1 and for the lowest is at 3rd day with the value of 74.5% for toilet 2.

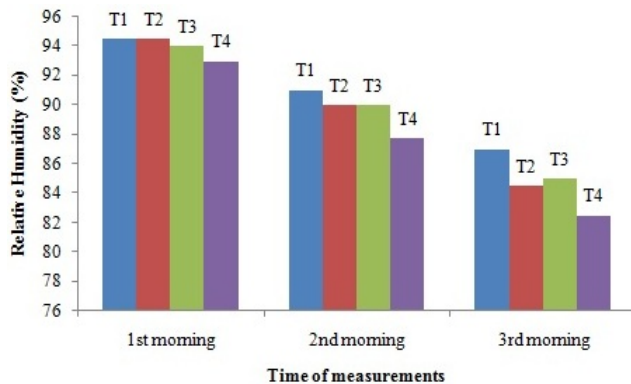


Figure 3. Measured morning relative humidity (toilet)

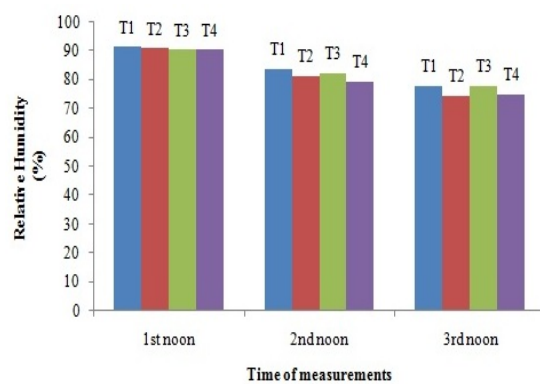


Figure 4. Measured noon relative humidity (toilet)

Al-anzi and Salman (2011) had stated that indoor humidity is one of the factors that contributes to asthmatic symptoms to asthmatic people. Since most of the toilets are experiencing high humidity thus this will trigger asthmatic symptoms for those asthmatic people who had frequently used this toilet.

3.2 Ventilation Rate

The air velocity was also measured for 10 minutes duration in all the 4 toilets in the morning and at noon time within 3 days period. The measured values are tabulated as in table 2.

Table 2. Measured Air Velocity in the Toilets

Toilet	Time of Measurement					
	Morning			Noon Time		
	Day 1	Day 2	Day 3	Day 1	Day 2	Day 3
1	0.091 m/s	0.069 m/s	0.051 m/s	0.09 m/s	0.2 m/s	0.09 m/s
2	0.132 m/s	0.099 m/s	0.065 m/s	0.11 m/s	0.16 m/s	0.09 m/s
3	0.122 m/s	0.103 m/s	0.083 m/s	0.08 m/s	0.14 m/s	0.26 m/s
4	0.091 m/s	0.098 m/s	0.105 m/s	0.1 m/s	0.19 m/s	0.05 m/s

For morning measurement, the table shows that the highest air velocity is at toilet 2 with the value of 0.132 m/s

on the 1st day of the measurement and for the lowest value is in toilet 1 with 0.051 m/s on the 3rd day of the measurement. For noon situation, the highest value is in toilet 3 with 0.26 m/s on the 3rd day and the lowest value is 0.05 m/s in toilet 4 also on the 3rd day of the measurement. These air velocity values are used to calculate airflow and thus, to determine the ventilation rate experienced by all the toilets. The airflow is calculated using the following formula:

$$Q = V \times A \tag{1}$$

Where:
 Q = Airflow (m³/s)
 V = Air Velocity (m/s)
 A = Area (m²)

The calculated airflows of the toilets are presented as in figures 5 and 6. For morning condition in figure 5, the highest calculated airflow is at toilet 2 with the value of 1.337 m³/s on the 1st day of measurement and the lowest is at toilet 1 on the 3rd day of measurement with the value of 0.463 m³/s. For noon time session, the highest of airflow is at toilet 1 with the value of 1.97 m³/s on the 2nd day of measurement and the lowest is 0.05 m³/s at toilet 4 on the 3rd day of measurement.

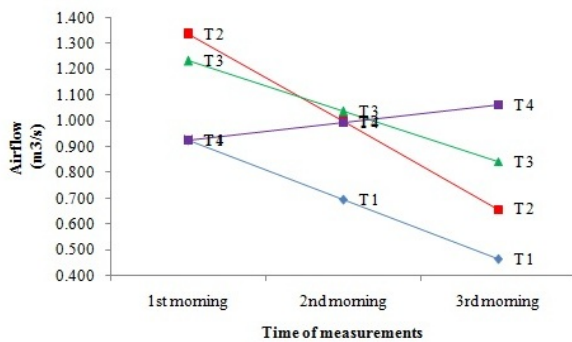


Figure 5. Calculated morning airflow

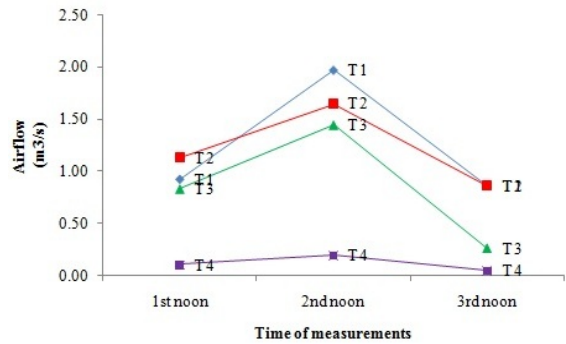


Figure 6. Calculated noon airflow

Subsequently, the flow rate is converted into ventilation rate to determine the adequacy of indoor air distributed for comparing with CIBSE standard, where the minimum ventilation rate for toilet is 629 l/s/p (CIBSE, 2001). The ventilation rate is calculated using the following equation:

$$V_r = Q / person \tag{2}$$

Where:
 Vr = Ventilation rate (l/s/p)
 Q = Airflow (l/s)

The calculated ventilation rates for all 4 toilets in morning and at noon time are presented as in figures 7 and 8.

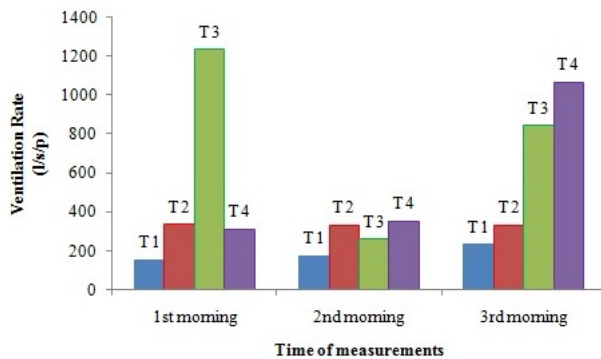


Figure 7. Morning ventilation rate

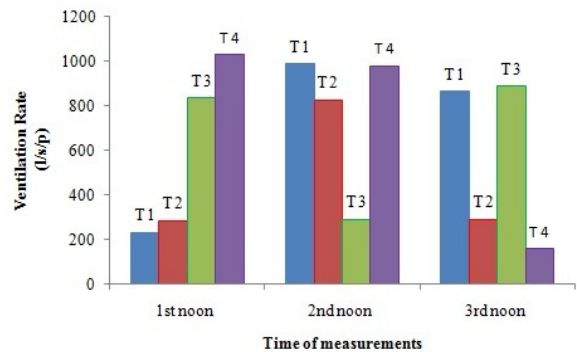


Figure 8. Noon ventilation rate

Figure 7 for morning session, indicates the highest ventilation rate is 1233.83 l/s/p in toilet 3 and for the lowest value is 154.23 l/s/p at toilet 1 on the same 1st day of measurement. In the afternoon, the highest value is 1028.19 l/s/p at toilet 4 on the 1st day of measurement and the lowest is 157.66 l/s/p at toilet 4 on the 3rd day of measurement. These results indicated that few of toilets are experiencing ventilation rate that are lower than the standard requirement of CIBSE (2001)

3.3 SO₂ Concentration

The SO₂ concentrations in all the 4 toilets were measured for 10 minutes duration in the morning and at noon time in 3 consecutive days. The measured SO₂ concentrations are as in figures 9 and 10. The measured SO₂ from all the toilets are compared to the TLV of SO₂ exposure which is 0.4 ppm (Protection, 2004; The National Advisory Committee for Acute Exposure Guideline Levels for Hazardous, 2008).

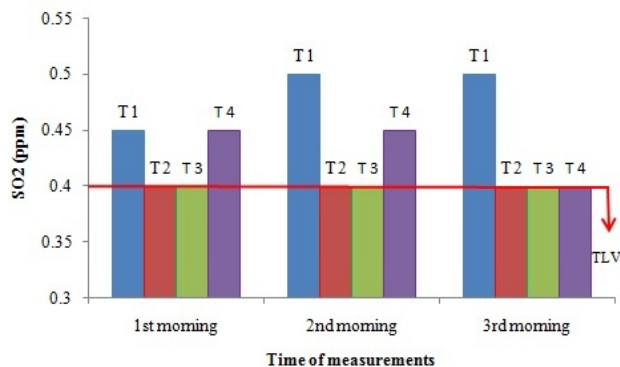


Figure 9. Measured morning SO₂

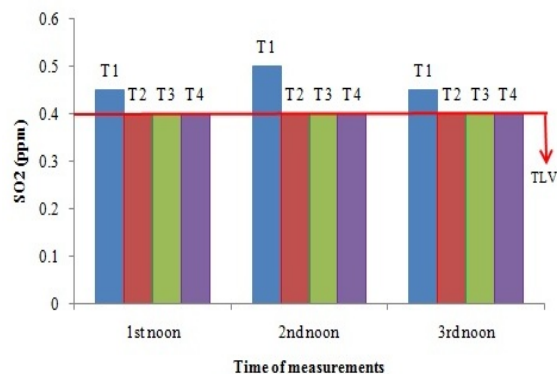


Figure 10. Measured noon SO₂

Figure 9 for morning measurement, SO₂ concentrations in toilet 1 and 4 exceeded the TLV of SO₂ with the value recorded up to 0.5 ppm. However, for noon time, only toilet 1 experience SO₂ concentrations exceeding the TLV up to 0.5 ppm. This situation is due to the inadequacy of air velocity/ventilation to dilute the indoor SO₂ pollutant. Besides that, in toilet 1 there occupants who are smoking during the measurement period were conducted.

4. Conclusion

The IAQ study in 4 selected toilets was carried out by measuring humidity, temperature, air velocity and SO₂ for assessing the conditions of the toilets that could trigger asthmatic symptom for those asthmatic people who are frequently used the toilets. The findings that can be drawn from this study are as follows:

- 1) It was found that SO₂ concentration in toilet 1 exceeded the TLV of SO₂ exposure for 10 minutes. This means that the toilet has the potential to trigger asthmatic symptom for those asthmatic people who frequently used this toilet.
- 2) Inadequate of air velocity/ventilation in toilet 1 for diluting indoor pollutant such as SO₂ contributes to the increment of SO₂ concentration. Additionally, indoor air humidity which was recorded indicates that this toilet is humid and it can affect to the growth of mold which also can trigger to asthmatic symptoms (Al-anzi & Salman, 2011).
- 3) The behavior of toilet user such as smoking at toilet can affect to the increases of indoor air pollutants which again increase the potential of triggering asthmatic symptoms

Acknowledgements

Thanks and gratitude to Ministry of Higher Education Malaysia and Research and Innovation Centre of Universiti Tun Hussein Onn Malaysia for funding this paper under Short Term Grant (STG) vote 1073.

References

- Al-anzi, F. S., & Salman, A. A. (2011). Correlation of Asthma Symptoms with Prevalence of Indoor NO₂ Concentration in Kuwait. *Journal of Environmental Protection*, 2011(2), 186–193. <http://dx.doi.org/10.4236/jep.2011.22021>
- Awbi, H. B. (1991). *Ventilation of Buildings* (Vol. 6, p. 309). London, UK: E & FN SPON. <http://dx.doi.org/10.1371/journal.pone.0022470>
- Balmes, J. R., Fine, J. M., & Sheppard, D. (1987). Symptomatic bronchoconstriction after short-term inhalation of sulfur dioxide. *Am. Rev. Respir. Dis.*, (136), 1117–1121.
- Bousquet, J., Jeffery, P. K., Busse, W. W., Johnson, M., & Vignola, A. M. (2000). Asthma, From Bronchoconstriction to Airways Inflammation and Remodeling. *American Journal of Respiratory and Critical Care Medicine*, 161, 1730–1745.
- Brunekreef, B., & Forsberg, B. (2005). Epidemiological evidence of effects of coarse airborne particles on health. *The European Respiratory Journal: Official Journal of the European Society for Clinical Respiratory Physiology*, 26(2), 309–18. <http://dx.doi.org/10.1183/09031936.05.00001805>
- Chartered Institution of Building Services Engineers (CIBSE) Guide B2, “Heating, Ventilating, Air Conditioning, and Refrigeration,” 2001.
- Donaldson, K., Gilmour, M. I., & Macnee, W. (2000). Commentary Asthma and PM 10. *Respiratory Research*, 1, 12–15.
- Frontczak, M., & Wargocki, P. (2011). Literature survey on how different factors influence human comfort in indoor environments. *Building and Environment*, 46(4), 922–937. <http://dx.doi.org/10.1016/j.buildenv.2010.10.021>
- Guo, P. (2012). Survival Analysis of Victims of Sulfur Oxide Air Pollution Suffering from COPD or Asthma in Yokkaichi, Japan, in Relation to Predisposing Exposure. *Journal of Environmental Protection*, 3(29), 1251–1259. doi:10.4236/jep.2012.329142
- Karakatsani, A., Analitis, A., Perifanou, D., Ayres, J. G., Harrison, R. M., Kotronarou, A., & Katsouyanni, K. (2012). Particulate matter air pollution and respiratory symptoms in individuals having either asthma or chronic obstructive pulmonary disease: A European multicentre panel study. *Environmental Health: A Global Access Science Source*, 11(1), 75. <http://dx.doi.org/10.1186/1476-069X-11-75>
- Koren, H. S. (1995). Associations between criteria air pollutants and asthma. *Environmental Health Perspectives*, 103Suppl(9), 235–42. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1518942&tool=pmcentrez&rendertype=abstract>
- Lee, S. (1997). Comparison of Indoor and Outdoor Air Quality at Two Staff Quarters in Hongkong. *Environmental International*, 23(97), 791-797.
- Luttinger, D., & Wilson, L. (2003). A study of air pollutants and acute asthma exacerbations in urban areas: status report. *Environmental Pollution*, 123(3), 399–402. [http://dx.doi.org/10.1016/S0269-7491\(03\)00025-3](http://dx.doi.org/10.1016/S0269-7491(03)00025-3)
- Massey, D., Masih, J., Kulshrestha, A., Habil, M., & Taneja, A. (2009). residential homes locations in central Indian region. *Building and Environment*, 44(10), 2037–2045. <http://dx.doi.org/10.1016/j.buildenv.2009.02.010>
- Muhamad-darus, F., Zain-ahmed, A., & Talib, M. (2011). Preliminary Assessment of Indoor Air Quality in Terrace Houses. *Health and the Environmental Journal*, 2(2), 8–14.
- Murdoch, J. R., & Lloyd, C. M. (2010). Chronic inflammation and asthma. *Mutation Research*, 690(1-2), 24–39. <http://dx.doi.org/10.1016/j.mrfmmm.2009.09.005>
- Protection, N. E. (2004). Review of the Practicability of a 10 Minute Sulfur Dioxide Standard Issues Paper.
- Restrepo, C. E. (2012). Asthma Hospital Admissions and Ambient Air Pollutant Concentrations in New York City. *Journal of Environmental Protection*, 3(29), 1102–1116. <http://dx.doi.org/10.4236/jep.2012.329129>
- Sheppard, D., Wong, W. S., Uehara, C. F., Nadel, J. J., & Boushry, H. A. (1980). Lower threshold and greater bronchomotor responsiveness of asthmatic subjects to sulfur dioxide. *Am. Rev. Resp. Dis.*, 122, 873-878. <http://dx.doi.org/10.1016/j.artmed.2004.01.013>
- The National Advisory Committee for Acute Exposure Guideline Levels for Hazardous. (2008). Sulfur Dioxide (CAS Reg. No. 7446-09-5) (AEGLS).

- Tseng, C. Y., Huang, Y. C., Su, S. Y., Huang, J. Y., Lai, C. H., Lung, C. C., & Liaw, Y. P. (2012). Cell type specificity of female lung cancer associated with sulfur dioxide from air pollutants in Taiwan: an ecological study. *BMC Public Health*, 12(1), 4. <http://dx.doi.org/10.1186/1471-2458-12-4>
- World Health Organization. (1987). Air Quality Guidelines for Europe. Copenhagen, Denmark.
- Zhang, Y. (2005). *Indoor Air Quality Engineering. Journal of Membrane Science* (Vol. 240). USA: CRC Press LLC.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

Reviewer Acknowledgements

Modern Applied Science wishes to acknowledge the following individuals for their assistance with peer review of manuscripts for this issue. Their help and contributions in maintaining the quality of the journal is greatly appreciated.

Modern Applied Science is recruiting reviewers for the journal. If you are interested in becoming a reviewer, we welcome you to join us. Please find the application form and details at <http://www.ccsenet.org/reviewer> and e-mail the completed application form to mas@ccsenet.org.

Reviewers for Volume 8, Number 5

Abdallah Al-Zyoud, Najran University, Kingdom of Saudi Arabia
Abdelhakim EL OUADIH, University of cadi AYYAD, Morocco
Alessandro Filisetti, European Centre for Living Technology, Italy
Andi Gusti Tantu, University of Makassar, Indonesia
Arvin Emadi, MAXIM INTEGRATED INC., USA
Christoph Maria Ravesloot, Rotterdam University of Applied Science, The Netherlands
Dinesh Sathyamoorthy, Ministry of Defence, Malaysia
Elham Sadat Yahyavi, Universiti Kebangsaan, Malaysia
Godone Danilo, Turin University, Italy
Hossein Chaji, University of Tabriz, Iran
James Manwell, University of Massachusetts in Amherst, USA
Jehn-Ruey Jiang, National Central University, Taiwan
Julien Wist, Universidad del Valle, Colombia
Julio Javier Castillo, National Technological University, Argentina
Liang Yu, Washington State University, China
Lim Hwee San, Universiti Sains Malaysia, Malaysia
Marek Brabec, Academy of Sciences of the Czech republic, Czech Republic
Mingxin Li, University of Delaware, USA
Nutsurang Pukkalanun, Surindra Rajabhat University, Thailand
Samendra Sherchan, University of Arizona, USA
Shang Yilun, Singapore University of Technology and Design, China
Tathagata Mukherjee, Seagate Technology, USA
Tomas Potuzak, University of West Bohemia, Czech Republic
V. Ananthaswamy, Madurai Kamaraj University, India
Valentina Valentina, Polytechnic University of Torino, Italy

Call for Manuscripts

Modern Applied Science (MAS) is an international, double-blind peer-reviewed, open-access journal, published by the Canadian Center of Science and Education. It publishes original research, applied, and educational articles in all areas of applied science. It provides an academic platform for professionals and researchers to contribute innovative work in the field. The scopes of the journal include, but are not limited to, the following fields: agricultural and biological engineering, applied mathematics and statistics, applied physics and engineering, chemistry and materials sciences, civil engineering and architecture, computer and information sciences, energy, environmental science and engineering, mechanics. The journal is published in both print and online versions. The online version is free access and download.

We are seeking submissions for forthcoming issues. All manuscripts should be written in English. Manuscripts from 3000–8000 words in length are preferred. All manuscripts should be prepared in MS-Word format, and submitted online, or sent to: mas@ccsenet.org

Paper Selection and Publishing Process

- a) Upon receipt of a submission, the editor sends an e-mail of confirmation to the submission's author within one to three working days. If you fail to receive this confirmation, your submission e-mail may have been missed.
- b) Peer review. We use a double-blind system for peer review; both reviewers' and authors' identities remain anonymous. The paper will be reviewed by at least two experts: one editorial staff member and at least one external reviewer. The review process may take two to four weeks.
- c) Notification of the result of review by e-mail.
- d) If the submission is accepted, the authors revise paper and pay the publication fee.
- e) After publication, the corresponding author will receive two hard copies of the journal, free of charge. If you want to keep more copies, please contact the editor before making an order.
- f) A PDF version of the journal is available for download on the journal's website, free of charge.

Requirements and Copyrights

Submission of an article implies that the work described has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the authorities responsible where the work was carried out, and that, if accepted, the article will not be published elsewhere in the same form, in English or in any other language, without the written consent of the publisher. The editors reserve the right to edit or otherwise alter all contributions, but authors will receive proofs for approval before publication.

Copyrights for articles are retained by the authors, with first publication rights granted to the journal. The journal/publisher is not responsible for subsequent uses of the work. It is the author's responsibility to bring an infringement action if so desired by the author.

More Information

E-mail: mas@ccsenet.org

Website: www.ccsenet.org/mas

Paper Submission Guide: www.ccsenet.org/submission

Recruitment for Reviewers: www.ccsenet.org/reviewer

The journal is peer-reviewed
The journal is open-access to the full text
The journal is included in:

CABI
Chemical Abstracts database
DOAJ
EBSCOhost
ERA
Google Scholar
LOCKSS
Open J-Gate

Polish Scholarly Bibliography (PBN)
ProQuest
Scopus
SHERPA/RoMEO
Standard Periodical Directory
Ulrich's
Universe Digital Library

Modern Applied Science Bimonthly

Publisher Canadian Center of Science and Education
Address 1120 Finch Avenue West, Suite 701-309, Toronto, ON., M3J 3H7, Canada
Telephone 1-416-642-2606
Fax 1-416-642-2608
E-mail mas@ccsenet.org
Website www.ccsenet.org/mas

