MODELLING FOR ENGINEERING AND HUMAN BEHAVIOUR 2018

2018

MATHEMATICAL
MODELLING IN
ENGINEERING
& HUMAN
BEHAVIOUR 2018

Valencia, Spain
July 16th-18th,
2018

L. Jódar, J. C. Cortés and L. Acedo(Editors)

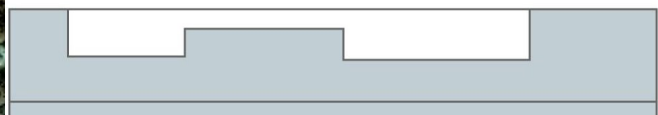## Instituto Universitario de Matemática Multidisciplinar

im²
Instituto de Matemática Multidisciplinar

UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

Ciudad Politécnica de la Innovación

MODELLING FOR ENGINEERING, & HUMAN BEHAVIOUR 2018

Instituto Universitario de Matemática Multidisciplinar

Universitat Politècnica de València

Valencia 46022, SPAIN

CONTENTS

A model for making choices with fuzzy soft sets in an intertemporal framework

José Carlos R. Alcantud[♭] [*], and María José Muñoz Torrecillas[†]

(♭) BORDA Research Unit and Multidisciplinary Institute of Enterprise (IME), University of Salamanca,

Campus Miguel de Unamuno, 37007 Salamanca, Spain,

(†) Universidad de Almería,

Facultad de Ciencias Económicas y Empresariales, La Cañada de San Urbano, s/n, 04120 Almería, Spain

November 30, 2018

1 Introduction

We present the main features of a model of choice where the alternatives are characterized by one fuzzy soft set in each of an indefinite number of periods. This model extends the standard model for choosing among fuzzy soft sets so that it can operate when the consequences of a decision extend over an infinite number of periods, or their termination date is unknown.

We explain how we can associate a characteristic fuzzy soft set with each modelisation in our framework. With this tool we are enabled to produce a decision making procedure for the selection of alternatives.

The target applications include portfolio selection in finance and evaluation of environmental issues among others [3].

2 Soft sets and fuzzy soft sets: notation and definitions

Let X denote a set. Then $\mathcal{P}(X)$ is the set of all non-empty subsets of X. A fuzzy subset (also, FS) A of X is a function $\mu_A : X \to [0, 1]$. For each $x \in X$, $\mu_A(x) \in [0, 1]$ is the degree of membership of x in that subset. The set of all fuzzy subsets of X will be denoted by $\mathbf{FS}(X)$.

In soft set theory we refer to a universe of objects U, and to a universal set of parameters E.

Definition 1 (Molodtsov [12]). Let A be a subset of E. The pair (F, A) is a *soft set* over U if $F : A \longrightarrow \mathcal{P}(U)$.

The pair (F, A) in Definition 1 is a parameterized family of subsets of U, and A represents the parameters. Then for every parameter $e \in A$, we interpret that $F(e)$ is the subset of U approximated by e, also called the set of e-approximate elements of the soft set.

---

[*]e-mail: jcr@usal.es

Other interesting investigations expanded the knowledge about soft sets. Soft set based decision making was pioneered by Maji, Biswas and Roy [11], and further applications of soft sets in decision making were given [4, **?**, 7, 11, 13].

Definition 2 (Maji, Biswas and Roy [10]). The pair (F, A) is a *fuzzy soft set* (henceforth, FSS) over U when $A \subseteq E$ and $F : A \longrightarrow \mathbf{FS}(U)$.

The set of all fuzzy soft sets over U will be denoted as $\mathcal{FS}(U)$. Any soft set can be considered as a fuzzy soft set with the natural identification of subsets of U with FSs of U. For example, if our set of options are films and they are parameterized by attributes, then fuzzy soft sets permit to deal with properties like "scary" or "funny" for which partial memberships are quite natural. However soft sets are suitable only when properties are categorical, e.g., "Oscar awarded" or "3D version available".

In real practice both U and A use to be finite. Then let k and n denote the respective number of elements of U and A. These soft sets can be represented either by $k \times n$ matrices or by a tabular form (cf. [1]). The k rows are associated with the objects, and the n columns are associated with the parameters. Both practical representations are binary, that is to say, all cells are either 0 or 1. One can proceed in a similar way in fuzzy soft sets, but now the possible values in the cells lie in $[0, 1]$.

Table 1 below compares the most important criteria for decision making when the alternatives are characterized by FSSs.

Ref.	Aggregation	Methodology	Solution	Other issues
[13]	Min operator	Scores from a comparison matrix	Unique	Many ties Information is lost by aggregation
[8]	Not provided	Choice value of level soft set	Not unique	Ties proliferate Richness introduces indeterminacy Additional inputs needed (e.g., threshold fuzzy set)
[1]	Product operator	Scores from new relative comparison matrix	Unique	Good power of discrimination
[9]	Not provided	Similarity measure	Unique	Use of subjective weights

Table 1: A list of the main fuzzy soft set based decision making procedures with their main characteristics.

3   An intertemporal model for FSSs

The spirit of a parameterized description of the universe can be merged with other characteristics that are not present in the original formulation of the (fuzzy) soft sets. Here we investigate the case where each attribute produces a possibly different fuzzy parameterization of the universe, for each of an indefinite number of periods.

3.1   The model

Let $\mathcal{S}$ denote the set of infinite sequences of the interval $[0, 1]$ (also called infinite utility streams in specialized literature, e.g., [2, 5]). Our intertemporal model of fuzzy soft sets over

U is defined by $\bar{F} : A \longrightarrow \mathbf{S}(U)$ where $\mathbf{S}(U)$ represents the mappings $U \longrightarrow \mathcal{S}$. In this fashion, for each attribute and each alternative, we express the degree of belongingness of such an alternative in each period of time.

In practical terms, where both $U = \{o_1, \ldots, o_m\}$ and $A = \{e_1, \ldots, e_n\}$ are finite, we can represent this information in a table where the cells are either finite or infinite sequences of membership degrees:

	e_1	e_2	$\ldots\ldots$	e_n
o_1	$(u_{11}^1, u_{11}^2, \ldots, u_{11}^t, \ldots)$	$(u_{12}^1, u_{12}^2, \ldots, u_{12}^t, \ldots)$	$\ldots\ldots$	$(u_{1n}^1, u_{1n}^2, \ldots, u_{1n}^t, \ldots)$
$\vdots$				
o_m	$(u_{m1}^1, u_{m1}^2, \ldots, u_{m1}^t, \ldots)$	$(u_{m2}^1, u_{m2}^2, \ldots, u_{m2}^t, \ldots)$	$\ldots\ldots$	$(u_{mn}^1, u_{mn}^2, \ldots, u_{mn}^t, \ldots)$

Table 2: Tabular representation of our intertemporal model for fuzzy soft sets.

With this notation we represent $\bar{F}(e_j)(o_i) = (u_{ij}^1, u_{ij}^2, \ldots, u_{ij}^t, \ldots) \in \mathcal{S}$, hence u_{ij}^t means the degree of membership of o_i to the fuzzy set of elements that verify attribute e_j in period t.

3.2 Some relationships

It is natural to embed the FSS model in this context. With each $(F, A) \in \mathcal{FS}(\mathrm{U})$ we associate $\bar{F} : A \longrightarrow \mathbf{S}(U)$ such that for each e_j and o_i, $\bar{F}(e_j)(o_i) = (F(e_j)(o_i), \ldots, F(e_j)(o_i), \ldots)$. This is the intertemporal assignment where at each moment, the degree of membership of alternative o_i to the fuzzy set of elements that verify attribute e_j is constant.

Conversely, we can use *reduction mechanisms* in order to associate a FSS with each intertemporal modelization. The easiest mechanisms are applied cell-by-cell. For example, one can select the evaluation at a fixed period (e.g., the first one); or in finite instances, their highest/lowest evaluation, their (either arithmetic or geometric) average, ... Under infinity of the periods, natural modifications by supremum/infimum or discounted sums serve this purpose.

4 Decision making in intertemporal FSSs

We proceed to define a procedure for prioritizing the alternatives in Table 2. It consists of three basic steps.

Algorithm for decision making

Inputs: An intertemporal table of fuzzy soft sets (in the notation of Table 2). A reduction mechanism. A fuzzy soft set decision making procedure (e.g., from Table 1).

1: Associate a FSS with the original intertemporal information by the recourse to the selected reduction mechanism.

2: Prioritize the alternatives in the reduced FSS by the selected decision making procedure.

3: The result of the decision is any object o_k that is at the top of the ranking in the previous step.

---

5   An example with a finite horizon

We need to decide among three development plans for the next 4 years. Their adequacies for 4 affected provinces are given by the following intertemporal table of fuzzy soft sets:

	e_1	e_2	e_3	e_4
o_1	$(0.2, 0.3, 0.3, 0.4)$	$(0.5, 0.5, 0.6, 0.6)$	$(0.7, 0.7, 0.6, 0.6)$	$(0.6, 0.5, 0.5, 0.6)$
o_2	$(0.6, 0.5, 0.5, 0.4)$	$(0.3, 0.4, 0.6, 0.6)$	$(0.4, 0.5, 0.5, 0.5)$	$(0.6, 0.6, 0.4, 0.4)$
o_3	$(0.3, 0.3, 0.4, 0.5)$	$(0.5, 0.5, 0.7, 0.6)$	$(0.5, 0.6, 0.6, 0.5)$	$(0.5, 0.4, 0.6, 0.6)$

We use the arithmetic average reduction mechanism in order to obtain the FSS whose tabular representation is

	e_1	e_2	e_3	e_4
o_1	0.3	0.55	0.65	0.55
o_2	0.5	0.475	0.475	0.5
o_3	0.375	0.575	0.55	0.525

Now we apply the algorithm in [1], which produces the following comparison table and score table:

	o_1	o_2	o_3		Row-sum (R_i)	Column-sum (T_i)	Score (S_i)
o_1	0	0.491	0.199	o_1	0.69	0.593	0.097
o_2	0.4	0	0.25	o_2	0.65	0.825	-0.175
o_3	0.193	0.335	0	o_3	0.528	0.449	0.079

By looking at the S_i scores, we conclude that the prioritization of the plans should be $o_1 \succ o_3 \succ o_2$.

6   Conclusions

We have paved the way to analyzing choices in soft computing models, when their consequences extend over time and do not terminate at a fixed date. Long-term projects (like public investments or environmental actions) are within the range of the potential applications.

Although we have worked under the assumption that uncertain knowledge is modelled by fuzzy soft sets, the intertemporal analysis makes sense in other popular and applicable frameworks: separable fuzzy soft sets [6], hesitant fuzzy sets, et cetera. This will be the subject of separate analyses in the future.

References

[1] J. C. R. Alcantud, A novel algorithm for fuzzy soft set based decision making from multi-observer input parameter data set, Information Fusion 29 (2016) 142-148.

[2] J. C. R. Alcantud, Inequality averse criteria for evaluating infinite utility streams: The impossibility of Weak Pareto, Journal of Economic Theory 147 (2012) 353-363

[3] J. C. R. Alcantud, M. J. Muñoz Torrecillas, Intertemporal choice of fuzzy soft sets. Symmetry 10(9) (2018), 371.

[4] J. C. R. Alcantud, S. Cruz, M. J. Muñoz Torrecillas, Valuation fuzzy soft sets: A flexible fuzzy soft set based decision making procedure for the valuation of assets, Symmetry 9 (2017), 253

[5] J. C. R. Alcantud, M. D. García-Sanz, Paretian evaluation of infinite utility streams: an egalitarian criterion, Economics Letters 106 (2010), 209-211

[6] J. C. R. Alcantud, T. J. Mathew, Separable fuzzy soft sets and decision making with positive and negative attributes. Applied Soft Computing 59 (2017), 586-595.

[7] J. C. R. Alcantud, G. Santos-García, E. H. Galilea, Glaucoma diagnosis: A soft set based decision making procedure, in: J. M. Puerta et al. (eds.), Advances in Artificial Intelligence, vol. 9422 of LNCS, Springer, 2015, pp. 49–60.

[8] F. Feng, Y. Jun, X. Liu, L. Li, An adjustable approach to fuzzy soft set based decision making, Journal of Computational and Applied Mathematics 234 (2010) 10–20.

[9] Z. Liu, K. Qin, Z. Pei, A method for fuzzy soft sets in decision-making based on an ideal solution, Symmetry 9(10) (2017) 246.

[10] P. Maji, R. Biswas, A. Roy, Fuzzy soft sets, Journal of Fuzzy Mathematics 9 (2001) 589–602.

[11] P. Maji, R. Biswas, A. Roy, An application of soft sets in a decision making problem, Computers & Mathematics with Applications 44 (2002) 1077–1083.

[12] D. Molodtsov, Soft set theory-first results, Computers & Mathematics with Applications 37 (1999) 19–31.

[13] A. Roy, P. Maji, A fuzzy soft set theoretic approach to decision making problems, Journal of Computational and Applied Mathematics 203 (2007) 412–418.

Methylphenidate and the Self-Regulation Therapy increase happiness and reduce depression: a dynamical mathematical model

Salvador Amigó[1]*, *Joan C. Micó***, *Antonio Caselles****

(*) Departament de Personalitat, Avaluació i Tractaments Psicològics.
Universitat de València,
Av. Blasco Ibáñez 21, 46010, ciutat de Valencia, Spain.
(**) Institut Universitari de Matemàtica Multidisciplinar.
Universitat Politècnica de València.
Camí de Vera s/n., 46022, ciutat de Valencia, Spain.
(***) IASCYS member, Departament de Matemàtica Aplicada.
Universitat de València.
Dr. Moliner 50, 46100 Burjassot, Spain.

1. **Introduction**

Methylphenidate (MPH) is a stimulant drug that produces effects such as euphoria, vigor and kindness, as well as anxiety and confusion [1]. The work [2] presents a study with 23 healthy subjects: 12 subjects reported a general pleasant effect and 9 subjects reported a general non-pleasant effect.

Besides, the Self-Regulation Therapy (SRT) is a psychological procedure based on learning and suggestion. It has been specially designed to facilitate the reproduction of drug effects, imitation and re-experimenting effects of drugs [3]. In addition, the SRT is capable to reproduce the MPH effects on personality and mood, as well as its influence on different biological indicators such as glutamate, and the DRD3 and c-fos regulator genes [4-6].

This paper presents an experiment to study the dynamical effect of a MPH single dose (10 mg) in two mood scales: happiness and depression [7]. A previous study demonstrated that the SRT applied on drug consumers reproduced mood, increasing happiness and decreasing depression [8]. In addition, a dynamical mathematical model, the so-called response model, was provided in the work [9] to describe the personality change due to the MPH and reproduced by the SRT. Also a dynamical mathematical model, a modified version of the response model, is here used to describe the mood change due to the MPH [10]. In addition, the dynamical response to the SRT is demonstrated that can be described by the new response model in a healthy subject.

2. Participant, design and procedure

The participant was a 56-year-old man who is a University of Valencia staff member. A single-case experimental ABC design was used. In phase A, the participant received no treatment. At the start of phase B, the participant consumed 10 mg of MPH. In phase C, the participant underwent the SRT to reproduce the effects of MPH, but did not consume this drug. The participant filled in a sheet of adjectives every 7.5 minutes over a 3-hour period. These adjectives measure happiness and depression mood [7]. For the mathematical analysis, the modified response model was applied, whose usefulness has been shown to model the dynamic effect of a stimulant drug.

3. The response model

The response model is given by the integro-differential equation:

$$\left.\begin{array}{c} \frac{dy(t)}{dt} = a(b - y(t)) + p \cdot s(t) \cdot y(t) - q \cdot \int_0^t e^{\frac{x-t}{\tau}} \cdot s(x) \cdot y(x)\, dx \\ y(0) = y_0 \end{array}\right\} \tag{1}$$

In Eq. 1, *y(t)* represents the mood dynamics, i.e., happiness or depression; and *b* and *y0* are respectively its tonic level and its initial value. Its dynamics is a balance of three terms, which

---

[1] E-mail: salvador.amigo@uv.es

provide the time derivative of the mood dynamics studied: the homeostatic control $(a(b - y(t)))$, i.e., the cause of the fast recovering of the tonic level b, the excitation effect $(p \cdot s(t) \cdot y(t))$, which tends to increase the dynamics, and the inhibitor effect $(q \cdot \int_0^t e^{\frac{x-t}{\tau}} \cdot s(x) \cdot y(x)\, dx)$, which tends to decrease the dynamics and is the cause of a continuously delayed recovering. Parameters a, p, q and τ are named respectively the homeostatic control power, the excitation effect power, the inhibitor effect power and the inhibitor effect delay. In addition, the $s(t)$ time function represents the dynamics of the stimulus. It can be described by the equation:

$$s(t) = s_0 e^{-\beta \cdot t} + \begin{cases} \frac{\alpha \cdot M}{\beta - \alpha}(e^{-\alpha \cdot t} - e^{-\beta \cdot t}) : & \alpha \neq \beta \\ \alpha \cdot M \cdot t \cdot e^{-\alpha \cdot t} : & \alpha = \beta \end{cases} \tag{2}$$

In Eq. 2 *m(t)* is the non-assimilated methylphenidate amount, M is the initial amount of methylphenidate of a single dose and α is the methylphenidate assimilation rate. In addition *s(t)* represents the stimulus, i.e., the amount in organism of the methylphenidate non-consumed by cells, s_0 is the amount of methylphenidate present in organism before the dose intake, and β is the methylphenidate metabolizing rate.

In the calibration process of Eqs. 1 and 2, M=10 mg for Phase B, and M is calibrated in Phase C, while $s_0 = 0$, due to the individual has not consumed methylphenidate for very long. The calibration consists in finding the optimal parameter values that minimize the square sum of the difference between the experimental values and the theoretical ones in both Phases B and C for both scales: happiness and depression. However, both scales share the same M, α and β values in the same phases. The strength of the calibration is measured by the determination coefficient (R^2). In addition, the residuals' randomness is provided by the p-value of the Anderson-Darling test, which reports if the residuals distribute as a N(0,std), i.e., as a Normal distribution of zero mean and constant standard deviation (std), being std the standard deviation of the residuals.

4. Results
In Figs. 1, 2 and 3, the evolution of the happiness can be observed in the three phases of the experiment (A: Base Line; B: 10 mg of MPH; C: SRT). Note that in Phase B the happiness holds an inverted-U shape, reaching a maximum two hours after starting the phase. In Phase C, the SRT reproduces the same dynamical shape but reaching its maximum approximately half an hour after starting the phase, decreasing much earlier than Phase B. The same behavior was found in the above cited work [9].

On the other hand, Figs. 4, 5 and 6 show the dynamical evolution of the depression for the three phases of the experiment (A: Base Line; B: 10 mg of MPH; C: SRT). Note in Phase B that, as a consequence of the MPH dose (10 mg), the depression decreases quickly, reaching its lowest level approximately an hour and a half after starting the phase. Similarly than the previous case, the SRT reproduces the MPH dynamics, but much earlier than in the previous case, reaching its lowest level approximately half an hour after starting the phase.
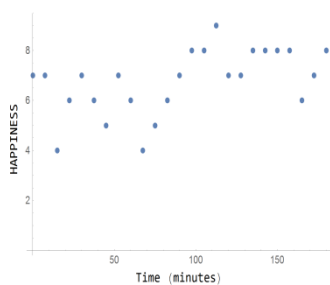
Fig. 1. Phase A: Base line.　**Fig. 2**. Phase B: R^2=0.87. P-value=0.79.　**Fig. 3**. Phase C: R^2=0.80. P-value=0.88.
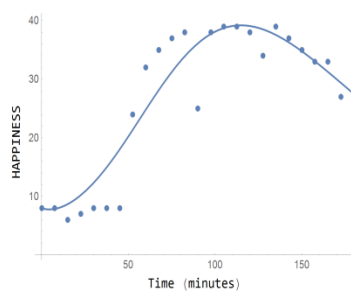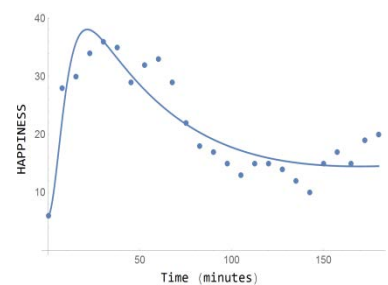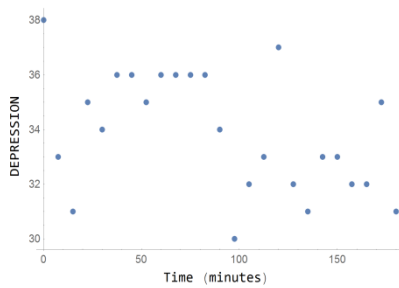
Fig. 4. Phase A: Base line. **Fig. 5**. Phase B: R^2=0.94. P-value=0.58. **Fig. 6**. Phase C: R^2=0.86. P-value=0.94.
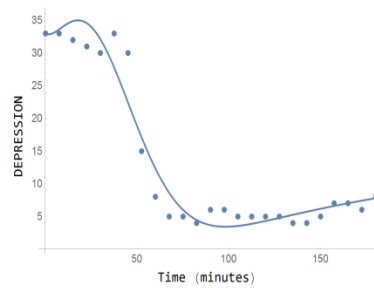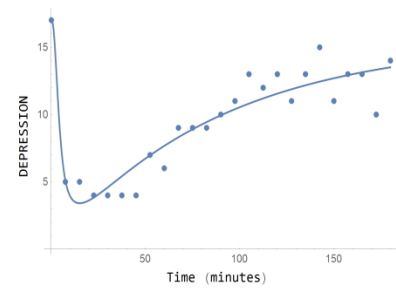
5. Discussion

The SRT has demonstrated its efficacy to reproduce the effects on the mood similarly to the mood observed as a consequence of a single dose of MPH. Concretely: increasing the happiness and reducing the depression. However, the SRT effect is shorter but of similar power. In addition the response model has been able to reproduce the mood dynamics, for both the MPH and the SRT.

These results open a way to the therapeutic application of the method presented in different behavioral disciplines, such as psychology, psychiatry or neurology. To do this, the patients need to be able to reproduce the MPH effects in several sessions. This assumption has been found in previous works [11], although a habituation effect can be observed.

Besides, there is clinical evidence that the results of the experiment can be used to treat the depression of a patient by using the MPH and the SRT to reproduce euphoria effects [12].

However, more scientific evidence is necessary to support the therapeutic use of this method, although this paper confirms its potential therapeutic benefit. In addition, having a mathematical tool such as the response model to reproduce the different dynamics is important to predict the therapeutic efficacy, as well as to reproduce the dynamical evolution of a treatment.

References

[1] Kollins, S.H., MacDonald, E.K. y Rush, C.R. *(*2001*)*. Assessing the abuse potential of methylphenidate in nonhuman and human subjects. *Pharmacology, Biochemistry and Behavior, 68,* 611-627.

[2] Volkow, N.D., Wang, G.J., Fowler, J.S., Logan, J., Gatley, S.J., Gifford, A., Hitzemann, R., Ding, Y.S. y Pappas, N. (1999). Prediction of reinforcing responses to psychostimulants in humans by brain dopamine D2 receptor levels. *American Journal of Psychiatry, 156*, 1440-1443.

[3] Amigó, S. (2005). *La teoría del rasgo único de personalidad. Hacia una teoría unificada del cerebro y la conducta* (The Unique Personality Trait Theory. Towards a unified theory of brain and behavior). Editorial de la Universitat Politècnica de València. Valencia, Spain.

[4] Amigó, S., Caselles, A., Micó, J.C. and García, J.M. (2009). Dynamics of the unique trait of personality: blood's glutamate in response to methylphenidate and conditioning. *Revista Internacional de Sistemas*, *16*, 35-40.

[5] Amigó, S., Caselles, A. and Micó, J. C. (2013). Self-regulation therapy to reproduce drug effects: A suggestion technique to change personality and DRD3 gene expression. *The International Journal of Clinical and Experimental Hypnosis, 61,* 282–304.

[6] Micó, J.C., Amigó, S. and Caselles, A. (2012). Changing the General Factor of Personality and the c-fos expression with methylphenidate and Self-Regulation Therapy. *The Spanish Journal of Psychology*, *15*, 850-867.

[7] Del Pino-Sedeño, T., Peñate, W. and Bethencourt, J. M. (2010). La escala de valoración del estado de ánimo (EVEA): análisis de la estructura factorial y de la capacidad para detectar cambios en estados de ánimo. *Análisis y Modificación de Conducta, 36*, 19-32.

[8] Amigó, S. (2014). Drugs, self-control and happiness. Comunicación presentada en el 9th Congress of the EUS-UES Globalization and Crisis. Complexity and governance. Valencia (Spain).

[9] Amigó, S., Micó, J.C. and Caselles, A. (2017). Methylphenidate and Self-Regulation Therapy: A systemic mathematical model. Comunicación presentada en Mathematical Modelling in Engineering & Human Behaviour 2017 Conference. Valencia (España).

[10] Micó, J.C., Amigó, S. and Caselles, A. (2018). Advances in the General Factor of Personality Dynamics, *Revista Internacional de Sistemas, 22*, 34-38.

[11] Amigó, S., Micó, J.C. and Caselles, A. (2018). Learning to be a psychostimulants addict with Self-Regulation Therapy. *Revista Internacional de Sistemas*, *22*, 13-21.

[12] Amigó, S. (1997). Uso potencial del metilfenidato y la sugestión en el tratamiento psicológico y en el aumento de las potencialidades humanas: un estudio de caso". *Análisis y Modificación de Conducta, 23*, 863-890.

A procedure to predict the short-term glucose level in a diabetic patient which captures the uncertainty of the data.

C. Burgos[(a)] *, J.C. Cortés[(a)] †, D. Martínez-Rodríguez[(a)] ‡
J. I. Hidálgo [(b)] §, R.-J. Villanueva[(a)] ¶

(a) Instituto Universitario de Matemática Multidisciplinar, Universitat Politècnica de València

(b) Departamento de Arquitectura de Computadores y Automática,Universidad Complutense de Madrid

November 26, 2018

1   Introduction

Diabetes Mellitus is an autoimmune disease that has a high prevalence in the world population. It is caused by a defect in the production (or effectiveness) of insulin, which means that the glucose we take from food is not processed correctly and therefore increases blood glucose levels. To avoid complications it is necessary to maintain glucose levels in a healthy range so is important to predict the glucose level value in a period of time.

We propose an adaptation of the model introduced in [1] to describe the evolution of the levels of glucose of a patient during an hour, taking into account the uncertainty due to measurements. The obtained parameters that describe the glucose levels during this hour will allow us to predict the glucose levels over the next four hours. This prediction will be accurate at

---

*e-mail: clabursi@posgrado.upv.es

†e-mail: jccortes@imm.upv.es

‡e-mail: damarro3@etsii.upv.es

§e-mail: damarro3@etsii.upv.es

¶e-mail: rjvillan@imm.upv.es

the beginning of the 4 hours, but it will deteriorate as the time goes on. In order to avoid this inconvenience and improve the accuracy, every half an hour we will remove the first 30 minutes of data used for calibration and add the new 30 minutes of glucose data.

This abstract is organized as follows. In Section 2 we introduce the model and the data uncertainty treatment. In Section 3 we describe the aforementioned approach to predict the glucose level of a patient. Then, we apply this technique to one particular patient, and the obtained results are shown in Section 4. Finally, Section 5 is devoted to conclusions.

2　Model description and data uncertainty treatment

In this section we introduce the model to predict the glucose level of a diabetic patient. The model we are going to consider, is the one adapted from [1]. We use this model because the model parameters correspond with therapeutic parameters used in the daily treatments to the patients. The model is given by the following system of difference equations.

$$U_{t+1} = U_t + V_t, \tag{1}$$

$$V_{t+1} = V_t - 2a_g V_t - a_g^2 U_t + K_g a_g^2 Ch_t, \tag{2}$$

$$G_{t+1} = G_t - X_t G_t - S_{g_0} G_t + U_{endo} + C\frac{U_t}{M}, \tag{3}$$

$$X_{t+1} = X_t - a_x X_t + a_x X_t^1, \tag{4}$$

$$X_{t+1}^1 = X_t^1 - a_x X_t^1 + K_x a_x \frac{I_t}{M}, \tag{5}$$

where U_t represents the gut glucose absorption at time t, V_t is the variation rate of the gut glucose absorption at time t, G_t stands for the level of Glucose at time t, X_t is the insulin action and X_t^1 represents the intermediate insulin action at time t.

The parameters of the model are related with the daily clinics of the patient, is to say, they are related with the biology of the patient. Ch_t and I_t are the level of ingested carbohydrates and insulin, respectively, C is the constant 50/9, M is the weight of the patient. Also a_g is the inverse of the meal time constant, K_g is the unitless bioavailability of the meal of

interest, S_{g_0} is the glucose effectiveness at zero insulin, U_{endo} is the insulin independent endogenous glucose production, a_x is the inverse of the insulin absorption/action time constant and K_x is the insulin sensitivity.

As we said before, we measure the glucose with an electronic device which measurement error is about 5% of the measured value. Furthermore, we want the model also captures the measurement uncertainty. To deal with this problem, we are going to assume that instead of having a single number for the glucose level, we have a Gaussian random variable with mean the glucose level value and standard deviation of 5% of the glucose level value. Then, we compute the percentiles 2.5 and 97.5 of each random variable (in each datum) and let us denote them by LP_t and UP_t respectively. These 95% confidence intervals will allow us to seek the sets of parameters that best capture the data uncertainty expressed via the 95% confidence intervals.

3   Procedure design

In this section, we describe a method to determine the sets of parameters that capture the uncertainty in the first hour and then predict 4 hours more with the same set of parameters. It is expected that the predictions obtained on the first minutes are more reliable than those obtained on the last minutes of these 4 hours. To avoid this deterioration every 30 minutes, we feedback the model rejecting the first half an hour and adding the new half an hour data. Thus, we will obtain new sets of parameters able to predict the glucose level over the next 4 hours.

To seek the set of parameters that capture the best the data uncertainty, we define for each set of parameters $Par = (a_g, K_g, S_{g_0}, U_{\text{endo}}, a_x, K_x)$ the following fitness function

$$FF(Par) = \left| \sum_{t=1}^{60} g(G_t^{Par}) \right|, \tag{6}$$

where

$$g(G_t^{Par}) = \begin{cases} 0, & \text{if } LP_t \leq G_t^{Par} \leq UP_t, \\ \min\left\{ G_t^{Par} - LP_t, UP_t - G_t^{Par} \right\} & \text{if } G_t^{Par} \geq LP_t \text{ or } UP_t \geq G_t^{Par}, \end{cases} \tag{7}$$

that is, the function is zero if the glucose level returned by the model lies inside the 95% confidence interval, and the minimum of the distance between

the glucose level and each percentile, otherwise.

Then, applying several times an rPSO algorithm, [5], we save all the set of parameters of all the rPSO iterations with their error. Among them, we want to select those sets of model parameters such that, when we substitute them into the model, retrieve the glucose level output given by the model in the same instants we have the glucose data and calculate their 95% confidence interval, these are as close as possible to the 95% confidence intervals of data. In order to select those outputs that capture the best the data uncertainty, we define M as the total amount of stored rPSO iterations, let E be the iterations and let us denote by E_i, $0 \leq i \leq M$ the i-th iteration. Then, we apply the following selection algorithm to select the best 100 iterations:

1. Initialization.

   - Initialize N index subsets $S_1, \ldots, S_N$ with 100 elements of the set $\{1, \ldots M\}$ (particles) chosen randomly without repetition. Evaluate the fitness of all the particles $F(S_1), \ldots, F(S_N)$.

   - Define the individual best fitness as $S_i^{best} = S_i$, $i = 1, \ldots, N$ and the global best fitness S_{global}^{best} as the $S_i^b est$ which fitness is the minimum.

2. For $i = 1, \ldots, N$, we extract without repetition 100 iterations from the union of the current particle, its individual best and the global best, and we denoted it as S_i. Evaluate the fitness of all the new particles $F(S_1), \ldots, F(S_N)$.

3. Update the individual best fitness $S_i^{best}, i = 1, \ldots, N$ and the global best fitness S_{global}^{best}. Go to Step 2.

4   Results

In Figures 1, we can see how evolves the calibration-selection procedure described in the previous section as the times goes on. Only 3 calibrations and predictions are shown for illustration purposes. We must say that we have chosen to test the described procedure the period starting where the patient is still sleeping, then wakes up and have breakfast until just before the lunch time. In this period there is a high and sudden change (wake up and breakfast) in the trajectory of the glucose level.
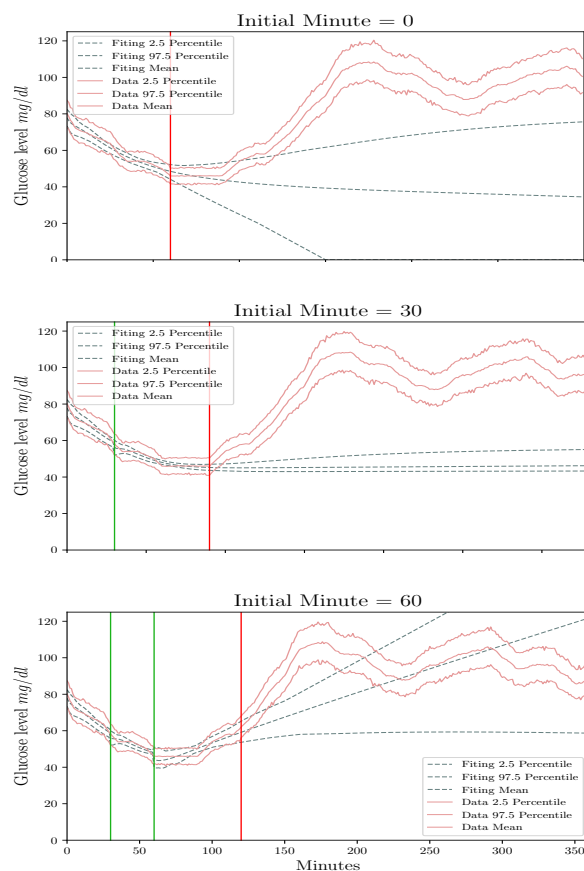
Figure 1: Feedback and fitting and prediction at the minutes $0, 30, 60$. On left line of the greed lines we can see the fitting obtained in the first and second step respectively in the firsts half hours. Then, between the red and green lines we can see the fitting in this third step. On the right of the red line the 4 hour predictions are shown.

5   Conclusions

The problem of predicting the glucose levels for diabetic patients is very complex. To face it, we propose an adapted model introduced in [1]. We design an appropriate computational technique handling the data uncertainty to predict with confidence intervals in the short-term the glucose levels of a patient.

   Even though the obtained results are partial and a thorough examination

of the way we perform the experiment is required, we think that we will be able to tune the procedure to provide to the doctors, in advance, tools to make appropriate and accurate decisions in the patient's treatment.

Acknowledgements

References

[1] A. Bock, G. Francois, D. Gillet. A therapy parameter-based model for predicting blood glucose concentrations in patiens with type 1 diabetes. Computed Methods and Programs in Biomedicine. 118 (2015) 107-123, `http://dx.doi.org/10.1016/j.cmpb.2014.12.002`.

[2] C. Anaya, C. Burgos, J. C. Cortés, R. J. Villanueva. Capturing the Data Uncertainty Change in the Cocaine Consumption in Spain Using an Epidemiologically Based Model `http://dx.doi.org/10.1155/2016/1758459`. Abstract and Applied Analysis. (ISSN: 1085-3375). Hindawi Publishing Corporation.

[3] C. Burgos, J.C. Cortés, C. Lombana, D. Martínez, R. J. Villanueva. Uncertainty quantification of the users of electronic commerce over the next few years. Proceedings of the conference in *Mathematical Modelling in Engineering and Human Behaviour 2017*. ISBN: 978-84-697-8505-8, pgs./pp.: 202-207.

[4] C. Cobelli, C. Dalla-Man, M. Gram-Pedersen, A. Bertoldo, G. Toffolo. Advancing our understanding of the glucosa system via Modeling: A perspective. *IEEE Transactions On Biomedical Engineering*, vol 6, num 5. (2014).

[5] C. Jacob, N. Khemka, Particle Swarm Optimization in *Mathematica*. Exploratory Toolkit for Evolutionary and Swarm-Based Optimization, The Mathematica Journal, Wolfram Media, 11:3 (2009).

Assessing organizational risk in industry by evaluating interdependencies among human factors through the DEMATEL methodology

Silvia Carpitella[1,2], Fortunato Carpitella[3], Antonella Certa[2], Julio Benítez[1], Joaquín Izquierdo[1]

[1] Instituto Matemático Multidisciplinar/Universitat Politècnica de València, València, Spain.

[2] Dipartimento dell'Innovazione Industriale e Digitale (DIID)/Università degli Studi di Palermo, Palermo, Italy.

[3] Studio di Ingegneria Carpitella, Trapani, Italy.

1. Introduction

This contribution proposes a Multi-Criteria Decision-Making (MCDM)-based approach to support organizational risk assessment in industrial environments.

Clerici *et al.* (2016) affirm that an organization is a plurality of "human elements", and risks often depend on organizational criticalities, whose reduction can be undertaken by implementing effective human resource management (HRM). In particular, HRM is defined as a system of structured procedures aimed at optimizing the manpower management in a company (Azadeh and Zarrin, 2016), its workers being the most valuable assets of the organization (Boatca and Cirjaliu, 2015; Carpitella *et al.*, 2017). As asserted by Cirjaliu and Draghici (2016), nowadays companies seek to continuously improve the well-being and satisfaction of human resources within their own operative environments. An important aspect to take into account within this context is integrated by human factors and ergonomics (HF/E), whose optimal management is crucial to achieve important objectives.

The importance of this concept is broadly shared in literature. Wilson (2014) asserts that any understanding of systems ergonomics must be related to the idea of systems engineering. Hassall *et al.* (2015) stress that analyses based on human factors and ergonomics are commonly used to improve safety and productivity, particularly in complex systems. Sobhani *et al.* (2017) underline as the improvement of workplace ergonomic conditions gives opportunities to better deal with production variations and optimize the performance of operation systems.

Given the importance of aspects related to HF/E within industrial workplaces, the Decision-Making Trial and Evaluation Laboratory (DEMATEL) method, firstly

developed by Fontela and Gabus (1974; 1976), is herein suggested as mathematical model to evaluate mutual relationships of some of the most important human factors involved in industrial processes. Among various MCDM methods proposed in literature, the DEMATEL is particularly helpful to take into account existing interdependencies among the main elements involved in any complex decision-making problem, on the basis of judgments attributed by a team of experts. It is clear, indeed, the high degree of correlation bonding human factors in any process led by humans. These interdependencies are eventually represented by means of a graphical chart.

2. Human factors and ergonomics in industry

Amount and intensity of human interactions with processes generally depend on the field in which the organisation operates. It is neither possible nor convenient totally eliminating the human contribution to processes, even when a high degree of automation is pursued, such as in manufacturing industries (Choe *et al.*, 2015). On one hand, industries with high production volumes may consider machines and computers as faster and more reliable than humans in leading automatic operations. On the other hand, the more customised the manufacturing process, the more crucial the role of human factors.

On the basis of what expressed so far, organisational risk assessment in industrial environments is conducted with the aim of evaluating, eliminating or at least minimise risks related to ineffective manners of work, in terms of methods and operations management from humans. Such kind of risks derives from psychological and physical conditions that negatively impact on the broad quality of work and life.

In particular, when leading organisational risk assessment, the main areas reported in Table 1 are analysed with a deep level of detail. The purpose consists in highlighting the presence of possible criticalities related to human factors and ergonomics within each area, which could potentially damage the global wellness and health of workers, and then the performance of the whole organisation.

Table 1. Description of investigated area related to human factors and ergonomics

ID	Investigated area focused on HF/E
A_1	Organisational culture and role
A_2	Career development and job stability
A_3	Communication, information, consultation and participation of workers
A_4	Training, awareness and competence
A_5	Operational control: indication of measures and instruments
A_6	Extraordinary situations and changes management
A_7	Outsourcing and interference management
A_8	Workload and working hours

We propose a MCDM-approach to rank the reported areas and then focus on major criticalities related to human factors. In particular, the DEMATEL methodology is suggested to select those areas more influencing each other (Carpitella *et al.*, 2018). This approach is useful to suggest an order in planning and implementing reduction measures of organisational risk.

3. The DEMATEL to increase the level of safety in industrial processes

The implementation of the DEMATEL methodology can be summarised through the following steps.

- Clear definition of the problem under analysis, in terms of goal and main elements/factors involved.

- Building the non-negative matrices $X^{(k)}$, where $1 \leq k \leq H$, H being the number of experts, expressing judgments on the mutual influence between pairs of elements. Elements $x_{ij}^{(k)}$ ($i, j = 1, \ldots, n$, n being the number of compared elements) represent the numerical values encoding the judgments. The meanings of those numerical values are defined as follows: 0 (no influence), 1 (very low influence), 2 (low

influence), 3 (high influence), 4 (very high influence). The main diagonal values of any of these matrices are zero.

- Building the direct-relation matrix A, incorporating the matrices filled in by the experts. A is a $n \times n$ squared matrix whose entries a_{ij} are obtained by:

$$a_{ij} = \frac{1}{H} \sum_{k=1}^{H} x_{ij}^{(k)}. \tag{1}$$

- Building the normalized direct-relation matrix $D = s\,A$, where s is given by:

$$s = \min \left[\frac{1}{\max_{1 \le i \le n} \sum_{j=1}^{n} a_{ij}}, \frac{1}{\max_{1 \le j \le n} \sum_{i=1}^{n} a_{ij}} \right]. \tag{2}$$

- Calculating the total relation matrix T, incorporating direct and indirect effects, calculated as the sum of the series of powers of D, given by

$$T = D(I - D)^{-1};$$
$$(3)$$

where I is the identity matrix.

- Obtaining a causal diagram by previously defining r_i and c_i as $n \times 1$ and $1 \times n$ vectors respectively representing the sum of rows and sum of columns of the total relation matrix T. The sum $r_i + c_i$ gives the overall effect of element i and the subtraction $r_i - c_i$ helps in dividing the elements into cause (if the subtraction is positive) and effect (if the subtraction is negative) groups.

- Drawing the chart by mapping the dataset of $(r_i + c_i, r_i - c_i)$, after having established a proper threshold to avoid taking into account also negligible effects. A threshold value is finally determined as the average value of the elements belonging to T.

To exemplify the applicability of our approach, a real-world case study is developed to evaluate interdependencies among the areas focused on human factors with relation to a manufacturing process of a Sicilian firm.

4. Real-world case study of a Sicilian winery

The case study refers to a manufacturing firm, precisely a winery located in the city of Trapani, in the isle of Sicily (Italy). The DEMATEL is herein applied to evaluate

interdependencies among the human factors of Table 1 related to the wine bottling process carried out by the company.

This process is composed of 13 different phases, reported in Figure 1, and takes place in the area dedicated to delivery and production. In the mentioned area there are three fixed stations and a mobile position, respectively occupied by the following operators:

1. W_1, worker dedicated to control that bottles are filled in and plugged;
2. W_2, worker dedicated to control the global quality of bottles;
3. W_3, worker dedicated to wrap final products;
4. W_4, worker dedicated to carry out the following two activities: raw materials (empty bottles, labels and corks) and packaging supply; handling of wrapped final products.
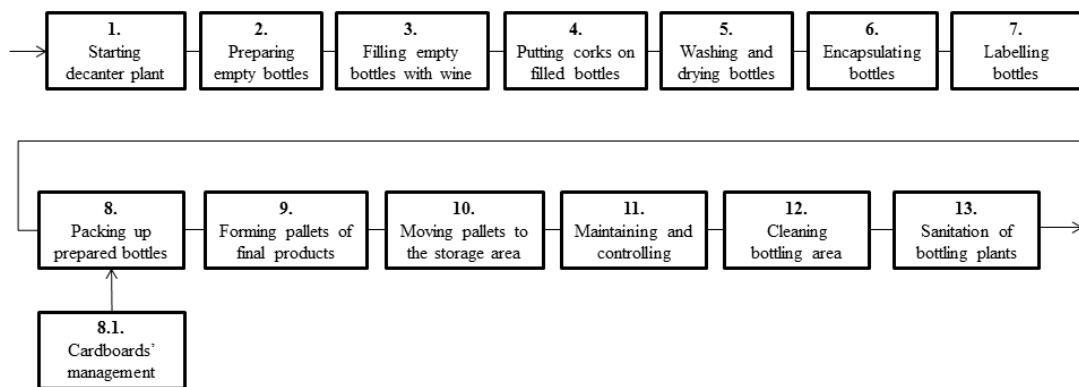
Figure 1. Phases of the bottling process

Three experts in the field ($H = 3$) were involved to apply the DEMATEL, namely the enologist, the department chief and the technical consultant. Each decision-maker was asked to evaluate the direct influence between any two human factors by means of integer scores from 0 to 4. Three non-negative square matrices X^1, X^2, X^3 were collected and then aggregated to obtain the direct-relation matrix $A = [a_{ij}]$ of size 8×8 (Table 2). Table 3 and Table 4 respectively report the total-relation matrix and the final ranking of areas, whereas Figure 2 represents the final chart (by considering a threshold of 0.872).

Table 2. Direct-relation matrix A

A	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8
A_1	0.000	4.000	2.333	2.333	2.333	2.333	2.333	4.000
A_2	4.000	0.000	3.000	3.000	2.333	2.333	2.333	3.000
A_3	2.000	3.333	0.000	2.333	2.333	2.333	2.333	2.333
A_4	2.333	3.333	2.333	0.000	2.000	3.000	3.333	2.000
A_5	2.333	2.333	2.333	2.333	0.000	2.000	2.333	4.000
A_6	2.333	2.333	1.333	3.000	1.667	0.000	4.000	4.000
A_7	2.333	2.333	2.667	4.000	1.667	3.667	0.000	2.333
A_8	4.000	2.667	2.667	1.333	3.667	4.000	2.333	0.000

Table 3. Total direct-relation matrix T

T	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	$r_i + c_i$	$r_i - c_i$
A_1	0.842	1.019	0.822	0.878	0.801	0.943	0.907	1.070	14.492	0.072
A_2	1.001	0.871	0.851	0.909	0.805	0.948	0.915	1.042	14.769	-0.085
A_3	0.814	0.883	0.630	0.780	0.707	0.832	0.804	0.891	12.574	0.109
A_4	0.876	0.935	0.770	0.734	0.735	0.909	0.892	0.934	13.546	6.786
A_5	0.857	0.876	0.752	0.803	0.638	0.852	0.831	0.984	12.634	0.551
A_6	0.898	0.918	0.751	0.872	0.742	0.814	0.936	1.027	14.244	-0.331
A_7	0.896	0.923	0.800	0.912	0.741	0.958	0.783	0.969	13.991	-0.025
A_8	1.025	1.002	0.856	0.872	0.873	1.031	0.939	0.953	15.422	-0.318

Table 4. Final ranking of areas

Ranking of areas	$r_i + c_i$ ↓
A_8	15.422
A_2	14.769
A_1	14.492
A_6	14.244
A_7	13.991
A_4	13.546
A_5	12.634
A_3	12.574
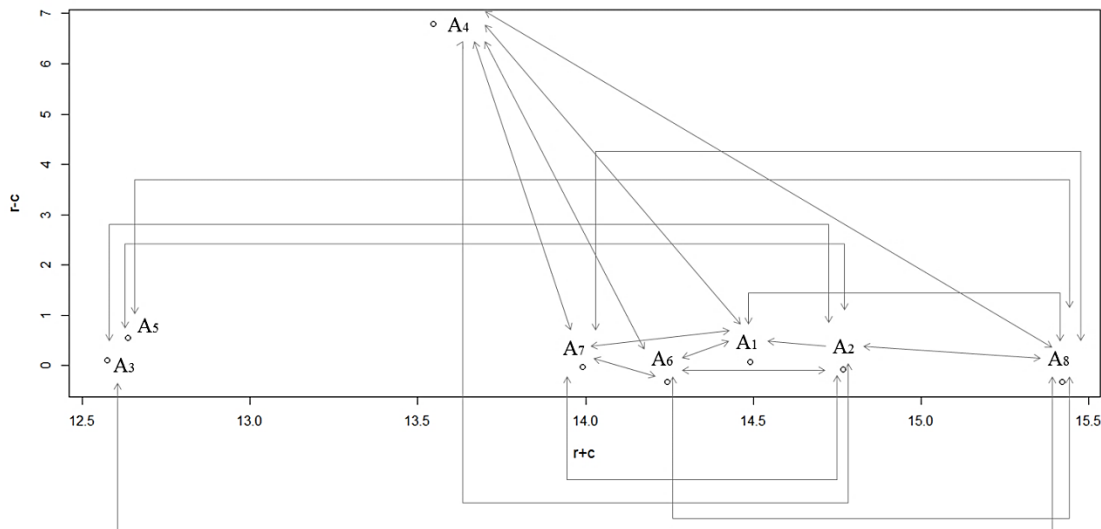
Figure 2. DEMATEL chart

The arrows in the chart represent relations of influence and they bond two areas A_i and A_j if the related value of the total-relation matrix is $T(A_i, A_j) \geq 0.872$.

As we can observe, the area A_8 occupying the first position in the ranking should be more carefully monitored during the process of organisational risk management related to the wine bottling process object of the case study. In other terms, aspects related to workload and working hours should be object of analysis and readjustment in order to globally enhance organisation quality. Indeed, variations implemented within this area can correspond to variations of all the other aspects. Lastly, being the value of $r_i - c_i > 0$ for the mentioned area, we can consider it as a cause-element.

Acknowledgements

References

Azadeh, A., Zarrin, M. (2016). An intelligent framework for productivity assessment and analysis of human resource from resilience engineering, motivational factors, HSE and ergonomics perspectives. Safety Science, 89, 55–71.

Boatca, M.E., Cirjaliu, B. (2015). A Proposed Approach for an Efficient Ergonomics Intervention in Organizations. Procedia Economics and Finance, 23, 54–62.

Carpitella, S., Certa, A., Enea, M., Galante, G.M., Izquierdo, J., La Fata, C.M. (2017). Human Reliability Analysis to support the development of a software project. Proceedings of the 23th ISSAT International Conference on Reliability and Quality in Design, Chicago, Illinois, USA, August 4-6, 214-218.

Carpitella, S., Certa, A., Izquierdo, J. (2018). DEMATEL-based consensual selection of suitable maintenance KPIs. Proceedings of the 24th ISSAT International Conference on Reliability and Quality in Design, Toronto, Canada, August 2-6, 270-274.

Choe, P., Tew, J.D., Tong, S. (2015). Effect of cognitive automation in a material handling system on manufacturing flexibility. International Journal of Production Economics, 170, Part C, 891–899.

Cirjaliu, B., Draghici, A. (2016). Ergonomic Issues in Lean Manufacturing. Procedia - Social and Behavioral Sciences, 221, 105–110.

Clerici, P., Guercio, A., Quaranta, L. (2016). Human Management System for Occupational Health and Safety. Tipolitografia INAIL, Milano.

Fontela, E., Gabus, A. (1974). DEMATEL, innovative methods, Technical report no. 2, Structural analysis of the world problematique. Battelle Geneva Research Institute.

Fontela, E., Gabus, A. (1976). The DEMATEL Observe. Battelle Institute, Geneva Research Center.

Hassall, M., Xiao, T., Sanderson, P., Neal, A. (2015). Human Factors and Ergonomics. International Encyclopedia of the Social & Behavioral Sciences (Second Edition), 297–305.

Sobhani, A., Wahab, M.I.M., Neumann, W.P. (2017). Incorporating human factors-related performance variation in optimizing a serial system. European Journal of Operational Research, 257 (1), 69–83.

Wilson, J.R. (2014). Fundamentals of systems ergonomics/human factors. Applied Ergonomics, 45 (1), 5–13.

Selection of an anti-torpedo decoy for the new frigate F-110 by using the GMUBO method

Rafael M. Carreño*[1], *Javier Martínez***, *J. Benito Bouza****

(*) Centro Universitario de la Defensa-Marín.
Universidad de Vigo.
Plaza de España s/n, 36920 Marín, Spain.
(**) Escuela Superior de Ingeniería y Tecnología.
Universidad Internacional de La Rioja.
(***) Departamento de Diseño en la Ingeniería.
Universidad de Vigo.
Escuela de Ingeniería Industrial, C/ Torrecedeira 86 - 36208 Vigo, Spain.

Introduction

Currently, the Spanish Navy has two types of frigates in service, the F-80 class and the F-100 one. F-80 frigates will be replaced in 2022, after 35 years of service. It is planned that F-80 units will be replaced by brand-new F-110 frigates. The F-110 project is in the conceptual design phase and one of the objectives is to provide the new frigate F-110 with a remarkable anti-submarine capability. Frigates are surface warships and anti-submarine warfare (ASW) is one of the most complex concerns in a surface warship (González-Cela, Bellas, Martínez, Touza, & Carreño, 2018).

The presence of an enemy submarine in the area of operations of a surface warship is a dangerous threat against the ship. Therefore, one of the important decisions that must be taken is the choice of the best anti-torpedo decoy that will be implemented in the F-110.

When a torpedo is launched, it goes straight to the target because it is fitted with an electronic device that enables it to find and hit the target. An anti-torpedo decoy is an acoustic device used as countermeasure to avoid the attack of torpedoes. A decoy works by transmitting the emulated ship's signature to confuse the torpedo (Liang & Wang, 2006).

This paper addresses the problem of the selection of the best anti-torpedo decoy to be implemented in the new F-110 frigates. The methodology described below was carried out in order to solve this problem. According to the Navy guidelines, two possible decoy alternatives were chosen: towed device (N) and expandable device (L), (Ercís, 2013). A group of experts from the Navy established the proper criteria and the Analytic Hierarchy Process (AHP) method was applied to determine the best anti-torpedo decoy (Saaty, 1990). The result of applying the AHP did not produce a solution to the problem since none of the decoys obtained a better score to the other one to make a decision. This allowed implementing a new method for decision-making, the Graphic Method of Measurement of Uncertainty Beyond Objectivity (GMUBO). The new method integrates the uncertainty in the AHP method and can help the Spanish Navy Staff (EMA) to make a decision.

Methodology

First of all, an evaluation by experts from the Navy was done in order to establish relevant criteria and sub-criteria. The chosen criteria had been previously monitored by the EMA so a survey was administered to experts of the Navy. The survey results allowed to determine the weights of the criteria which were used in the AHP method. The sub-

---

[1] E-mail: rafaelcarreno@uvigo.es

criteria and criteria are presented in the chart 1, where all values are fictitious due to real ones are classified information.

Secondly, the AHP method was applied to determine the best anti-torpedo decoy. The decision process based on the AHP considers a finite number of alternatives x_i, for i from 1 to n. A score is assigned to each alternative (w_i is the score of alternative x_i), providing a weight vector. A square matrix of pairwise comparison is used to solve a Multi-Criteria Decision Making (MCDM).

$$A = (a_{ij})$$

Where $a_{ij} = 1/a_{ji}$ and $a_{ii} = 1$ for all i from 1 to n (Saaty, 1990). Saaty proposed a consistency index (CI) to evaluate the consistency of the pairwise comparison matrix, CI $= (\lambda_{max} - \text{n})/(\text{n} - 1)$.

To deduce the weight vector, is used the eigenvector theory. The method consists of finding the Perron-Frobenius eigenvector (Brunelli, 2015), which corresponds to the maximum eigenvalue of the pairwise comparison matrix, that is $A \cdot w = \lambda_{max} \cdot w$.

The AHP decomposes a problem into a hierarchy of smaller sub-problems which can more easily be evaluated. Thus, the AHP provides a hierarchy of goal, criteria (c_i), sub-criteria (sc_{ij}) and alternatives (a_i).

Thirdly, for the logistics sub-criteria "storage volume" (SV), a utility function were used. This allows to add objective data of the real volume occupied by each decoy in a warship. Similarly, a utility function is used to evaluate the sub-criteria "reaction time" (RT). This adds objectivity since the real values of the time it takes to make effective use of the decoy of each of the alternatives are known (Marzouk & Moselhi, 2003). Thus, the following linear utility functions were considered:

$$y_{SV} = A \cdot x_{SV} + B \Leftrightarrow \begin{cases} x_{SV} = 5 \ m^3 \Rightarrow y_{SV} = 0 \\ x_{SV} = 1 \ m^3 \Rightarrow y_{SV} = 1 \end{cases}$$

$$y_{RT} = C \cdot x_{RT} + D \Leftrightarrow \begin{cases} x_{RT} = 30 \ s \Rightarrow y_{RT} = 0 \\ x_{RT} = \ \ 0 \ s \Rightarrow y_{RT} = 1 \end{cases}$$

Once the AHP was applied, the results are shown in the table 1. For a better analysis of this result, a sensitivity analysis was carried out as is shown in figures 1, 2 and 3.

Finally, the GMUBO method considers uncertainty of the process. In order to achieve a robustness in the results, the uncertainty in the process must be integrated. To do this, different scenarios were considered. These scenarios are changes to the weightings of the objective. Then, uncertainty of the alternatives, considered as grey numbers, was calculated. Subsequently, the best alternative was determined, taking into account that the scenarios are not controllable by the decision maker.

In the AHP method, a criteria comparison matrix is multiplied by a priority vector and an overall priority vector (OPV) is obtained. The OPV determines a hierarchy on the selection of alternatives and provides a first selection that does not consider uncertainty. A decision maker do not know which scenario is going to arise. Then, it is necessary to repeat the process considering a number of scenarios. Each scenario allows to obtain a corresponding OPV. Keeping this in mind a Penalties Matrix (*m* x *n*) is built:

$$P = (c_{ij})$$

Consider $A = \{a_1, a_2 \ldots a_m\}$ the set of alternatives and $S = \{s_1, s_2 \ldots s_n\}$ the set of scenarios. In P matrix, each c_{ij} is the penalty obtained after choosing the alternative a_i when the given scenario is s_j.

To choose the best alternative two calculations are performed. On the one hand, a measure of the uncertainty is needed. On the other hand, weighted sums calculation is carried out. The best alternative should have the highest value of weighted sum and the lowest value for uncertainty.

Since the penalties are considered as grey numbers, the following expression is a measure of the uncertainty for each alternative:

$$g_i^0 = \frac{1 \cdot c_{im} + \sum \mu_{ij} \cdot c_{ij} + 0 \cdot c_{ik}}{c_{ik} - c_{im}}$$

Where $c_{ik} = \max_{j \neq k, m} c_{ij}$ and $c_{im} = \min_{j \neq k, m} c_{ij}$, with $\mu_{ij} \in [0,1]$.

Now the weighted sum is calculated for each alternative i:

$$W_i = \lambda_{i1} \cdot p_{i1} + \cdots + \lambda_{nn} \cdot p_{nn}$$

The best alternative should have the highest value of weighted sum and the lowest value for uncertainty. If both values lead to more than one alternative, then the alternative with the greatest final sum FS must be selected.

$$FS = W_i + (max g_i^0 - g_i^0) + 1/\sum c_{ij}$$

Results and discussion

The AHP method has wide applicability and allows dealing with complex problems by synthesizing them. Moreover, it establishes a ratio scale that makes easy the measurement (Forman & Gass, 2001). The AHP has an axiomatic foundation and uses a clearly defined mathematical structure (Saaty, 1986).

Although the criteria, sub-criteria and their weights were provided by a survey conducted with EMA experts, a utility functions had to be used for two of the sub-criteria. Specifically, the sub-criteria of SV and RT, since the real values of the storage volume and the reaction times were available. This was possible because the technical features of the different alternatives were available, all of which allowed the process to be more objective.

Table 1 provide the final decision matrix. Then, as is shown in column 1, there is no clear preference for one of the alternatives. As shown in Table 1, the results are clearly different if only one of the criteria is taken into account. Specifically, alternative N acquires a clear advantage over alternative L if only the "Logistics" criterion is considered in the evaluation. Similarly, if only the criterion "Operational Capabilities" is taken into account, it is observed that alternative L is better than alternative N.

It follows that, once the AHP method is applied, practically equal results are obtained for both N and L alternatives. To be exact, it is obtained that the alternative L is slightly preferable to the alternative N, but with only a 1 percent difference between L and N. Because this difference is very small, it can be concluded that given the available information, it is not advisable to establish which alternative is better.

Figures 1, 2 and 3 show a sensitivity analysis obtained by varying the weights assigned to the criteria. The sensitivity analysis allows to observe that a small variation

in the weights assigned to the criteria produces that the chosen alternative changes. Therefore, the result obtained in the problem is very sensitive to variations in the influence or weight of the two criteria "Logistics" and "Operational Capabilities".

This means that, according to this result, it is not advisable to choose any of the two alternative solutions if you want to be certain of success. This is because the method does not clearly decide which alternative is preferable to the other. This is one source of uncertainty in the decision-making process. The AHP is a method that does not handle uncertainty although it is a good starting point to implement a better solution.

Different options were evaluated: reevaluation and addition of more criteria, the use of complementary calculations, simulations or sophisticated software, and the integration of uncertainty in the AHP method. These options are discussed below.

Initially, by adding new criteria to the hierarchy to proceed to a re-evaluation, applying again the AHP. This option was discarded because it required more time and did not take into account the uncertainty inherent in the process. It should also be noted that this solution would imply that the experts would not have done their job correctly. However, the EMA took special care to define the most important criteria for decision-making. Moreover, the experts assigned the scores according to their experience and current environment.

Then, some mathematical techniques were considered as the theory of probability and fuzzy logic, among others, which do consider uncertainty. The implementation of these techniques involved complex calculations, simulations and sophisticated software. Actually, decision-makers do not have time to develop techniques that require excessive amounts of knowledge. Hence, ease of use and an exactness are basic features to encourage decision-makers in using of useful techniques.

Finally, uncertainty was implemented using the GMUBO method, which is easy to use and useful to help decision-making under uncertainty, as well as providing a very useful graphical tool (Figure 4). GMUBO provides two vectors that are combined subsequently. On the one hand, a measure of the uncertainty given by the degree of greyness of each alternative. On the other hand, for a given alternative, the inverse of the sum of its penalties is measured. Both vectors can provide a clear and only alternative. However, there are situations where each vector leads to a different alternative. In these cases, the final sum (FS) would resolve the discrepancy.

FS measures the suitability of an alternative considering all the scenarios that can arise. FS controls and minimize the effect of a very small uncertainty that could modify the choice of the best alternative. It is considered that for a given alternative, the inverse of the sum of its penalties should be as large as possible to avoid alternatives with both high penalty values and a very small uncertainty.

GMUBO considers uncertainty and allows to measure the robustness of the selected alternative. Furthermore, it is a helpful method for decision-makers since it has a great ease of use.
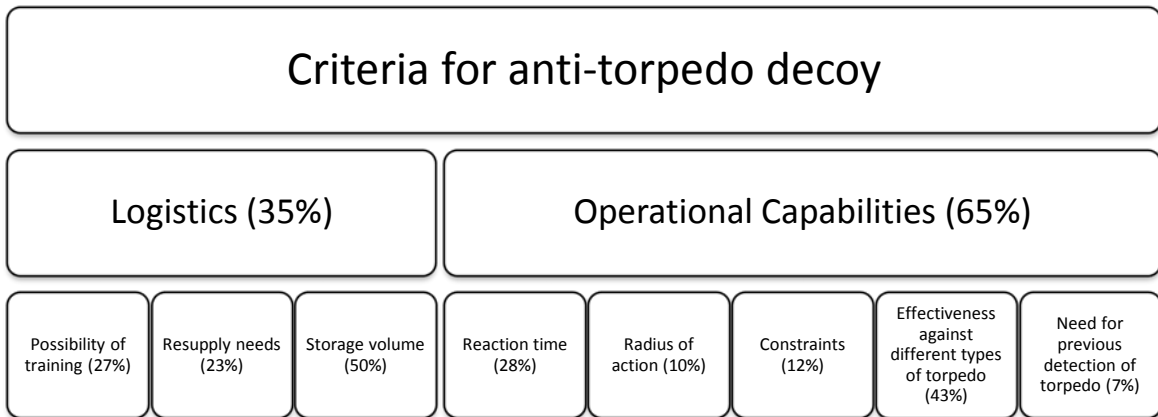
Chart 1: Sub-criteria and criteria used in AHP method.
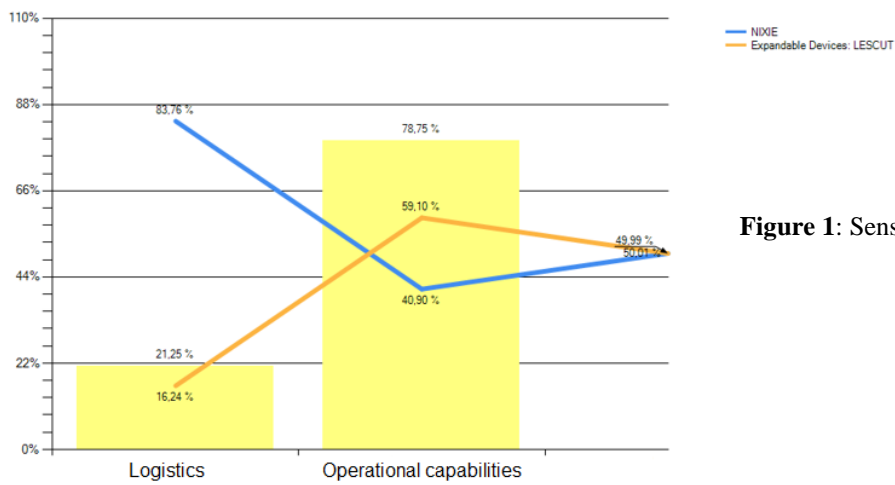
Figure 1: Sensitivity analysis
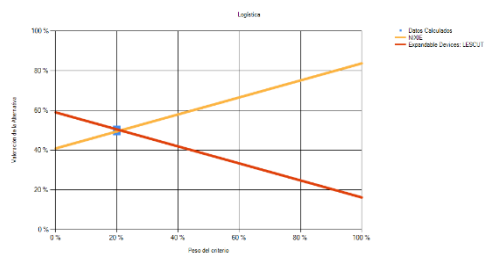
Figure 2: Sensitivity of the weight assigned to the Logistics criterion
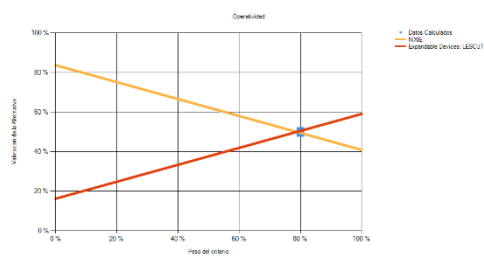
Figure 3: Sensitivity of the weight assigned to the Operational Capabilities criterion
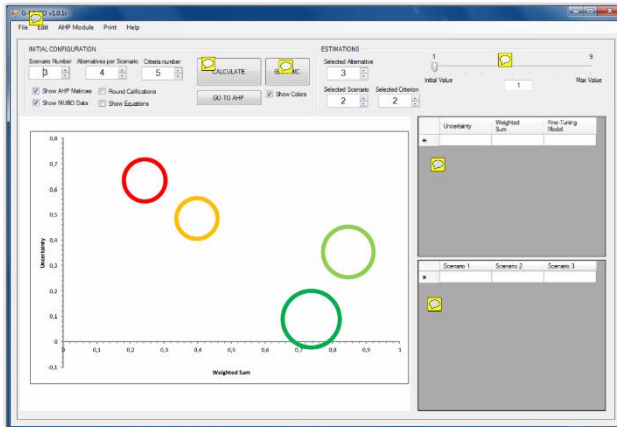
Figure 4: Graphical interface of the new method

	Decision matrix	
	Logistics (35%)	Operational capabilities (65%)
Alternatives	sub-criteria	sub-criteria
N (49.5%)	84%	41%
L (50.5%)	16%	59%

Table 1: Final decision matrix.

References

- Brunelli, M. (2015). Priority vector and consistency. In *Introduction to the Analytic Hierarchy Process* (pp. 17-31). Springer, Cham.
- Ercís, F. (2013). *Design and construction of unmanned underwater vehicle* (Doctoral dissertation). Middle East Technical University, Ankara, Turkey.
- Forman, E. H., & Gass, S. I. (2001). The analytic hierarchy process—an exposition. *Operations research*, 49(4), 469-486.
- González-Cela, G., Bellas, R., Martínez, J., Touza, R., & Carreño, R. (2018). Optimal Design of Spanish Navy F-110 Frigates Combat Information Center. *Naval Engineers Journal*, 130(1), 79-90.
- Liang, K. H., & Wang, K. M. (2006, December). Using simulation and evolutionary algorithms to evaluate the design of mix strategies of decoy and jammers in anti-torpedo tactics. In *Simulation Conference, 2006. WSC 06. Proceedings of the Winter* (pp. 1299-1306). IEEE.
- Marzouk, M., & Moselhi, O. (2003). A decision support tool for construction bidding. *Construction Innovation*, 3(2), 111-124.
- Saaty, T. L. (1986). Axiomatic foundation of the analytic hierarchy process. *Management science*, 32(7), 841-855.
- Saaty, T. L. (1990). How to make a decision: the analytic hierarchy process. *European journal of operational research*, 48(1), 9-26.

A modelling methodology based on General Systems Theory

Antonio Caselles[1]

IASCYS member
Departament de Matemàtica Aplicada. Universitat de València (retired).
Dr. Moliner 50, 46100 Burjassot, Spain.

1. **Introduction**

The General Systems Theory (GST) appeared in the middle of the last century at the hands of Von Bertalanffy [1] with the intention to introduce a global vision of things (which implies interdisciplinary collaboration among specialists) and maybe to become a universal language common to all academic disciplines. Since then, several attempts to formalize GST have took place (for instance: [2-6]), one of them [7-13] has given rise to a modelling methodology that has yield many studies and mathematical models applied mainly to the social sciences (for instance: [14-20]). The following tries to be an outline of this methodology. For a comparative analysis of some possible approaches to the formalization of the GST you can see the one of Caselles [21] in the EOLSS Encyclopaedia.

2. **Basic concepts**

The basic idea of the formalization of Caselles (sensitized by the criticisms from many authors to the previous formalizations) is that this is really useful. In other words, it can serve as a basis for a methodology and its associated software tools that facilitate the process of building models of real systems. This implies a definition of basic concepts (ontology) and a form of acquisition and representation of knowledge (epistemology) that takes into consideration the two original ideas of GST: global vision and universal language, besides the practical utility.

The definition of system as "set of interrelated elements" firstly leads to the concept of *structure* [7] that would be formalized, given a set *A,* as a subset of the Cartesian product of *AxA*, i.e., something that would be translated into a diagram of arrows or a binary relationship, showing connections, influences, dependencies, etc., among the considered elements. Some derived concepts would be: the "*structural application*" which would make correspond the set of its influencers to each element; *input, output, strict, isolated, level 1, 2,..., n elements*; and others such as: *direct influence, indirect influence, loop, etc*.
Secondly, the term "related" also gives place, depending on the type of relationship in study, to the concept of *behaviour* that would imply that the elements being considered are represented by mathematical variables (symbols with a domain of possible values, not necessarily numeric). Then, the behaviour of the system, which would be based on its structure (the variables that influence each of them are known), would allow us to know the value that would take each variable of the system from the values that would take the other variables, and what is the form of the function that allows to determine this value. This set of functions (determined under the form of equations, tables, logical rules or algorithms) would be that would define the behaviour of the system.

Thirdly, it is necessary to classify systems with behaviour into *static and dynamic, deterministic and stochastic, temporal, spatio-temporal, learning, etc*. And the variables into *state-variables or not, dimensioned or not, with uncertainty or not*, and the input variables into *constants or value-changing, etc*. Finally, having a list of variables, each one with its own features, and a list of functions, one for each output variable, it proceeds to check that we have a *hierarchical system*, i.e. a system without loops. Otherwise calculations are not possible. The elimination of loops implies the introduction in the system of state variables (those which depend on its previous value or earlier ones) and their corresponding *memory variables* (which should also be input variables). The hierarchysation of the system consists in to classify variables by levels according to the level that have the variables of lower level on which they depend. For instance, level-1 variables would be the input ones (data); level-2 variables would be those depending only on level-1 ones; level-3 variables would depend only on level-1 and level-2 variables, etc. This hierarchy allows setting the order of calculation. Obviously the memory variables must be updated at each time step.

3. **Tools**

The second part of the methodological problem, the first is the representation of knowledge, already explained (structure and behaviour: model), is the use of this knowledge for the resolution of problems. This requires auxiliary tools which, in our case, we have reduced to the following software:

[1] E-mail: Antonio.caselles@uv.es

SIGEM, an intelligent system generator of computerized models [10, 13, 22, 23], is an automatic programming tool. Effort required by the development and debugging of a computer program is well known by all programmers. SIGEM tries to minimize this effort. In short, the only thing the programmer should do is to prepare two text files: a file with the list of variables with their characteristics represented with symbols (with uncertainty or not, dimensioned or not, state-variable or not, etc.), and another file with the list functions (one for each output variable, and not necessarily ordered). SIGEM collects, in an interactive way, some additional data about the system and, as a result of its work, it develops three programs: a data manager, a simulator/optimizer and a results manager. These programs would not have more errors than those existing in the referred text files (lists of variables and functions). SIGEM is currently programmed in Visual Basic 6 and the programs that it produces also are (they can be executed inside an EXCEL spreadsheet).

REGINT is an interactive search engine and adjuster of functions [23, 24] which are linear combination of other more simple functions, linear or non-linear (identity, product, exponential, logarithm, cosine, etc.). It uses as data a table of up to 13 columns (for a dependent variable and up to 12 independent ones) and up to 60000 rows or points. It allows several search options (exhaustive, sampling and genetic algorithm) and determination options for the degree of adjustment (by R^2, s^2, etc.). It allows obtaining deterministic or stochastic type functions (estimating the mean and the standard deviation as functions of the independent variables) that may enter as such in SIGEM.

EXTRAPOL is an extrapolator with confidence intervals of the type developed by REGINT or other functions. It is useful to design strategies and scenarios for the future to be simulated.

While non-numeric or mixed type computerized models can be built with SIGEM, sometimes there are no more information available than that can be deduced from opinions of experts in relation to the elements or functions of a system. In these cases it may be more practical to try to computerize the system and the resolution of the problem with other tools. For this reason our methodology includes two new programs:

DIFU is a fuzzy cognitive maps analyzer. It is based on estimating the degree of direct influence on each ranked pair of variables by means of expert opinions, resulting in an array of data. The result is another array with the total influences (the direct one plus the indirect ones by all possible influence chains). The program friendly presents these results indicating those elements having the greatest influence over each one of them, from the most positive one to the most negative one.

CISTE is a cross-impact simulator. It needs as data, the previous and current values of each output variable and forecasts for the future of the input variables, in addition to the previously mentioned array of direct influences. Events (with their respective estimated probability) are also permitted. The model may be either deterministic or stochastic (when each value of a variable or impact is introduced with their respective mean and standard deviation). In this last case, the forecasts would come with its respective confidence intervals or standard deviations.

4. **The modelling process**

As regards the overall modelling process [9, 19, 23], consider the following steps or "life cycle" since, although it seems, it is not a strictly sequential process: normal is that in each stage it is necessary to update the previous ones.
1. To find the list of elements/factors/variables.
    A. To state the problem: objectives, constraints, assumptions, types of data, types of results.
    B. To seek the involved factors from experts (bibliography, *Brainstorming, Delphi*).
    C. To find the variable or variables representing each involved factor (with its respective measurement unit and possible values).
    D. To classify them into input and output variables (black box diagram).
2. To seek from experts (bibliography, *Brainstorming, Delphi*) which variables have a causal or previous influence on each of them (causal diagram, blocks diagram).
3. To shape the relations/functions that allow giving a value to each variable based on the values of those influencing it. It requires prior knowledge of the subject in study. In some cases, it will be necessary to look for and adjust equations from tables (REGINT). If the system is dynamic (with state variables) it may be interesting to build a hydrodynamic diagram or Forrester [25] diagram.
4. To build the corresponding computer programs (SIGEM).
5. To verify the model and computer programs: the desired data must produce the corresponding expected results.
6. To validate the model: to be sure about the utility of the model to solve the proposed problem. The "prediction of the past" method is usual in this phase but other methods are possible, for instance: "prediction of the future" (waiting for its arrival), acceptation by experts, or acceptation by those assuming the corresponding risks.

---

[1] E-mail: Antonio.caselles@uv.es

7. Once the model is validated it proceeds to use it to solve the stated problem: to design experiments to be performed with the simulator, or starting the optimization procedure (designing strategies and scenarios, genetic algorithm, etc.).

Conclusions

As it has been shown in application cases publications [14-20], the suggested formalization of the GST and the corresponding working methodology and software tools have been demonstrated to be very useful for modelling systems, using a universal language, and solving problems.

References

[1] L. von Bertalanffy. General Systems Theory. George Brazilier. New York. 1968.

[2] G.J. Klir. An approach to General Systems Theory. D. Van Nostrand Co. Litton Educational Publishing International.1969.

[3] M. D. Mesarovic, Y. Takahara. Abstract Systems Theory. Springer Verlag. Berlin. 1989.

[4] Y. Lin, Y. Ma. Remarks on the concept of dynamical System. Cybernetics and Systems: An International Journal, 20 (1989) 435-450.

[5] A. W. Wymore. Wattled Theory of Systems. In Trends in General Systems Theory, G. J. Klir Ed. Wiley Interscience. New York. 1972.

[6] B. P. Zeigler. Theory of modelling and simulation. R.E. Krieger P.C. Inc. Malabas (FL 329590). 1976.

[7] A. Caselles. Structure and Behavior in General Systems Theory. Cybernetics and Systems: An International Journal, 23 (1992) 549-560.

[8] A. Caselles. Simulation of Large Scale Stochastic Systems. In Cybernetics and Systems'92. R. Trappl (ed.). World Scientific. Singapore. 1992. 221-228.

[9] A. Caselles. Systems Decomposition and Coupling. Cybernetics and Systems: An International Journal, 24 (1993) 305-323.

[10] A. Caselles. Improvements in the Systems Based Program Generator SIGEM. Cybernetics and Systems: An International Journal. 25 (1994) 81-103.

[11] A. Caselles. Goal-Seeking Systems. In Cybernetics and Systems Research '94. R. Trappl (ed.). World Scientific Publishing Corp. Singapore. 1994. 87-94.

[12] A. Caselles. Systems Autonomy and Learning from Experience. Advances in Systems Science and Applications, Inauguration Issue (1995) 97-102.

[13] A. Caselles. Building Intelligent Systems from General systems Theory. In Cybernetics and Systems Research '96. R. Trappl (ed.). Austrian Society for Cybernetic Studies. Vienna . 1996. 49-54.

[14] A. Caselles, L. Ferrer, I. Martínez de Lajarza, R. Pla, R. Temre. (1999). Control del desempleo por simulación (Controlling unemployment by simulation). Universitat de València. Valencia (Spain).

[15] A. Caselles-Moncho, L. Ferrandiz-Serrano, E. Peris-Mora. Dynamic simulation model of a coal thermoelectric plant with a flue gas desulphurization system. Energy Policy, 34 (2006) 3812–3826.

[16] A. Caselles, J.C. Micó, S. Amigó. Cocaine addiction and personality: A mathematical model. British Journal of Mathematical and Statistical Psychology, 63 (2010) 449–480.

[17] J.C. Micó, A. Caselles, D. Soler, P.D. Romero. Formalism for discrete multidimensional dynamic systems. Kybernetes, 45(10) (2016) 1555 – 1575.

[18] M.T. Sanz, A. Caselles, J.C. Micó, D. Soler. Including an environmental quality index in a demographic model. Int. J. Global Warming, 9(3) (2016).

---

[1] E-mail: Antonio.caselles@uv.es

[19] D. Soler, M.T. Sanz, A. Caselles, J.C. Micó. A stochastic dynamic model to evaluate the influence of economy and well-being on unemployment control. Journal of Computational and Applied Mathematics 330 (2018) 1063-1080.

[20] M.T. Sanz, A. Caselles, J.C. Micó, D. Soler. A stochastic dynamical social model involving a human happiness index. Journal of Computational and Applied Mathematics 340 (2018) 231-246.

[21] A. Caselles. Formal Approaches to Systems. In Systems science and cybernetics (F. Parra Luna Ed.) Vol. 2. (ISBN: 978-1-8426-653-7), Enciclopedia for Life Support Systems (EOLSS) (UNESCO). Eolss Publishers Co: Ltd. Oxford. United Kingdom. 2008. 312-338.

[22] A. Caselles. SIGEM: a realistic models generator expert system. In Cybernetics and Systems '88. R. Trappl (ed.). Kluwer Academic Publishers. Dordrecht. 1988. 101-108.

[23] A. Caselles. Modelización y simulación de sistemas complejos (Modeling and simulation of complex systems). Universitat de València. Valencia (Spain). 2008.   (Available in http://www.uv.es/caselles as well as software).

[24] A. Caselles. A tool for discovery by complex function fitting. In Cybernetics and Systems Research'98. R. Trappl (ed.). Austrian Society for Cybernetic Studies. Vienna. 1998. 787-792.

[25] J. Forrester. Industrial Dynamics. M.I.T. Press. Cambridge, MA. 1961.

---

[1] E-mail: Antonio.caselles@uv.es

Dynamics of the general factor of personality as a consequence of alcohol consumption

Salvador Amigó *, Antonio Caselles**, Joan C. Micó ***, Maria T. Sanz ****, David Soler ***

(*) Departament de Personalitat, Avaluació i Tractaments Psicològics. Universitat de València, Av. Blasco Ibáñez 21, 46010. València, Spain.
(**) IASCYS member, Departament de Matemàtica Aplicada.
Universitat de València (retired).
Dr. Moliner 50, 46100 Burjassot, Spain.
(***) Institut Universitari de Matemàtica Multidisciplinar.
Universitat Politècnica de València.
Camí de Vera s/n., 46022, ciutat de Valencia, Spain.
(****) Departament de Didàctica de la Matemàtica. Universitat de València, Avda. Tarongers, 4, 46022 València, Spain.

1. Introduction

The social importance of alcohol consumption in our Western Society cannot be neglected. Its common consumption is culturally accepted in suitable doses, but its relationship with personality disorders should be studied [1, 2]. However, the prediction about what a suitable dose is or what misuse is depends on the individual personality [3, 4].

The General Factor of Personality (GFP) and the response model are, respectively, the suitable psychological and mathematical tools to study the alcohol misuse. On a hand, the GFP is a trait that occupies the apex of the hierarchy of personality, and extends from an impulsiveness-and-aggressiveness pole (approach tendency) to an anxiety-and-introversion pole (avoidance tendency) [5]. In addition, the General Factor of Personality Questionnaire (GFPQ) proposed in the work [5] presents a questionnaire constructed specifically to assess GFP. Another way to measure GFP is the Five-Adjective Scale of the General Factor of Personality (GFP-FAS). The 5 adjectives are: adventurous, daring, enthusiastic, merry and bored. However, it can integrate all basic traits of personality [22]. Its validity and its relationship with the GFPQ to measure the GFP is proved in the works [6, 7]. Note from these works that extraversion is another way to refer to GFP, and it has a broader meaning than that generally implied in current personality research. In addition, the GFP-FAS scale has a trait-format (GFP-T) (how extraverted is an individual in general), which represents the individual stable personality, and a state-format (GFP-S) (how extraverted is an individual in a concrete situation), which represents the individual situational personality. Thus, the suitable way to measure the GFP dynamical response to a stimulus such as a stimulant drug is to determine the time evolution of the GFP-S.

Besides, the response model is capable to predict the short-term effects of a dose of alcohol on GFP and to report the results of an alcohol intake experiment. In fact, the dynamical GFP pattern (identified by the time evolution of the GFP-S scores) has a typical inverted-U pattern [8-10]. The inverted-U pattern was already identified by Solomon & Corbit [11] Grossberg [12] and Amigó [8], as the typical personality response to a stimulant drug. Moreover, these works report that, in the presence of a stimulus, the inverted-U is a consequence of a balance between an excitation effect and a delayed inhibitor effect. Additionally, those individuals with higher GFP-T scores have higher excitation and inhibitor effects (measured by the time evolution of the GFP-S scores). Oppositely, the individuals with lower GFP-T scores have lower excitation and inhibitor effects (also measured by the time evolution of the GFP-S scores). Note that, although alcohol is considered a depressant drug, its acute effects reproduce generally an inverted-U, referred in the literature about alcohol as biphasic effect, similar to a stimulant drug, such as literature demonstrates [13-15].

2. The experiment

Fifty volunteers presented to participate in the experiment, all of them from Valencia (Spain). Some selection rules were applied on them to be accepted as participants in the experiment:
a)   Do not have incompatible medication with alcohol.
b)   Come accompanied to the experiment.
c)   Do not work the day of the experiment.
d)   Do not be abstemious.
e)   Do not be alcoholic.
f)   For the control group: have had bad experiences with alcohol.

These selection rules provided thirty seven participants, divided into two groups: the experimental group (28 alcohol consumers) and the control group (9 non consumers). From them there were 10 males (27%) and 27 females (73%). The mean age was 32.84 (SD=11) with ages ranging between 20 and 55 years. The mean weight was 64.18 kg with weights ranging between 50 and 94 kg.

All the participants completed The Five-Adjective Scale of the General Factor of Personality (GFP-FAS) in trait-format (GFP-T) and state-format (GFP-S) before alcohol consumption. The participants in the experimental group (28) received *M*=26.51 g. of alcohol and a slight food, while the participants in the control group (9) just received the food. Every participant filled the GFP-S each 7 minutes. The response model calibration to the GFP-S scores is demonstrated that reproduces the biphasic GFP dynamics as a consequence of an alcohol dose intake described by the literature, i.e., a stimulant-like or excitation effect balanced by a sedative-like or inhibitor effect. In fact, the response model predicts that the high scores of GFP-T provide a stronger stimulant-like effect and a stronger inhibitor effect. Thus, the response model is a useful mathematical tool to predict those individuals inclined to the alcohol misuse.

3. The response model

The response model is the mathematical tool used to compute the short term dynamics of the GFP as a result of a stimulus produced by a single dose intake of a drug, such as it has been used in [9, 16, 17-19]. The kinetic part of the response model provides the evolution of the alcohol amount in organism, after being consumed by an individual. It is given by:

$$s(t) = \begin{cases} \frac{\alpha \cdot M}{\beta - \alpha} \left(e^{-\alpha \cdot t} - e^{-\beta \cdot t} \right) : \ \alpha \neq \beta \\ \alpha \cdot M \cdot t \cdot e^{-\alpha \cdot t} : \ \alpha = \beta \end{cases} \tag{1}$$

The *s(t)* variable represents the stimulus, i.e., the amount in organism of the alcohol non-consumed by cells, assuming that the amount of alcohol present in organism before the dose intake is zero, due to the experimental conditions, which obligates the participants to the non-alcohol consumption since the afternoon prior to the experiment. *M* is the drug initial amount, α is the drug assimilation rate and β is the drug metabolizing rate. The dynamics of the GFP is given by the integro-differential equation:

$$\left. \begin{array}{c} \frac{dy(t)}{dt} = a(b - y(t)) + \frac{p}{b} s(t) - b \cdot q \cdot \int_0^t e^{\frac{x-t}{\tau}} \cdot s(x) \cdot y(x) \, dx \\ y(0) = y_0 \end{array} \right\} \tag{2}$$

In Eq. 2, *y(t)* represents the GFP dynamics; and *b* and *y0* are respectively its tonic level and its initial value. Its dynamics is a balance of three terms, which provide the time derivative of the GFP: the homeostatic control $(a(b - y(t)))$, i.e., the cause of the fast recovering of the tonic level *b*, the excitation effect $(\frac{p}{b} s(t))$, which tends to increase the GFP, and the inhibitor effect $(b \cdot q \cdot \int_0^t e^{\frac{x-t}{\tau}} \cdot s(x) \cdot y(x) \, dx)$, which tends to decrease the GFP and is the cause of a continuously delayed recovering. Parameters *a*, *p*, *q* and τ are named respectively the homeostatic control power, the excitation effect power, the inhibitor effect power and the inhibitor effect delay.

Once the model is calibrated for an individual, the *excitation effect intensity* is represented by the individual excitation effect power value divided by the tonic level, *p/b*, and the *inhibitor effect intensity* is represented by the individual inhibitor effect power value multiplied by the tonic level, *b·q*. Thus, both terms, *p/b* and *b·q*, represent the corresponding individual intensities of the excitation effect and the delayed inhibitor effect demanded by the dynamic patterns forecasted in the works [8, 11, 12]. Therefore, the *p/b* and *b·q* terms are interpreted in the following way: the more inclination to the individual alcohol misuse (with higher GFP-T scores), the greater the individual excitation effect intensity value, *p/b*, and the greater the inhibitor effect intensity value, *b·q*, must be held; and oppositely, the lesser inclination to the individual alcohol misuse (with lower GFP-T scores), the lower the individual excitation effect intensity value, *p/b*, and the lower the inhibitor effect intensity value, *b·q*, must also be held.

4. Calibration of the response model

The calibration of the response model for the experimental group (Case 0) and the control group (Control 0), represented by their GFP-S time averages (one point each seven minutes), is provided respectively in Figs. 1 and 2. Both figure's description provides the fitting level by the determination coefficient (R^2). They also provide the residuals' randomness by the p-value of the Anderson-Darling test, which reports if the residuals distribute as a N(0,std), i.e., as a Normal distribution of zero mean and constant standard deviation (std), being std the standard deviation of the residuals.
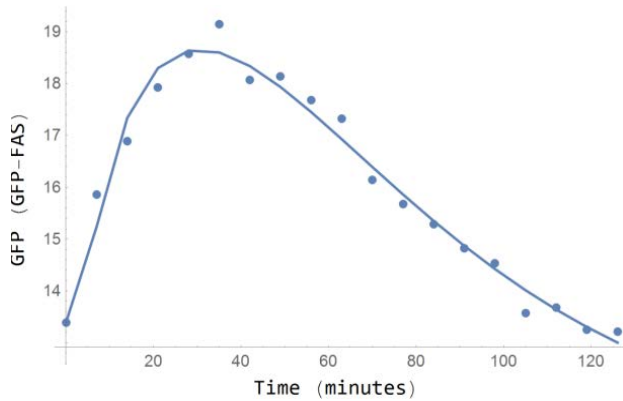
Fig. 1. General Factor of Personality response to the alcohol Intake versus time. Experimental values (dots) and theoretical values (line) for the experimental group (Case 0). R^2=0.97. P-value=0.97.
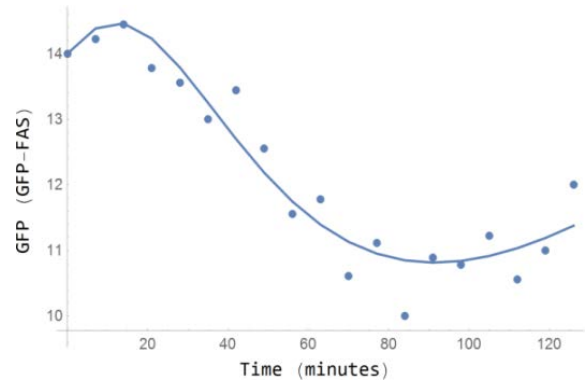
Fig. 2. General Factor of Personality response to the atmosphere stimulus versus time. Experimental values (dots) and theoretical values (line) for the control group (Control 0). R2=0.91. P-value=0.92.

5. The response model and the alcohol misuse

The experimental group (EG) is divided into two subgroups: the introverted group (IEG), or the consumers that have a GFP-T scores lesser than the median (17) (N=12) and the extraverted group (EEG), or the consumers that have a GFP-T scores greater than the median (N=12), also considering their b, p and q values obtained from the model calibration for both groups. Subsequently, a Mann-Whitney test is performed to know if there significant differences between both groups. The results are presented in Table 1.

Table 1. Statistics (U) and p-values of the Mann-Whitney tests to compare the excitation effect intensity values (p/b) and the inhibitor effect intensity values ($b·q$) for extraverts and introverts. EEG: extraverted group; IEG: introverted group.

Intensities	Group	N	Average rang	U	Sig.
p/b	EEG	12	15.75	33	.024
	IEG	12	9.25		
b·q	EEG	12	15.58	35	.033
	IEG	12	9.42		

6. Discussion

Note that the results of the Mann-Whitney test confirm that the more introverted group (those with lesser GFP-T, below median) has a lesser value for p/b and $b·q$ intensities, and thus, lesser stimulant-like and sedative-like effects, as a consequence of the alcohol intake. And, vice versa, the more extraverted group (those with greater GFP-T, above median) has a greater value for p/b and $b·q$ intensities, and thus, greater stimulant-like and sedative-like effects, as a consequence of the alcohol intake. The consequence of this test with the experimental group is that the GFP-T scores, i.e., the GFP-FAS in its trait format, jointly the response model, are good predictors of the alcohol misuse.

References

[1] J.M. Malouff, E.B. Thorsteinsson, S.E. Rooke, N.S. Schutte, Alcohol involvement and the Five-Factor Model of personality: a meta-analysis, J. Drug Educ. 37 (2007) 277-294.

[2] K.J. Sher, T.J. Trull, Personality and disinhibitory psychopathology: alcoholism and Antisocial Personality Disorder, J. Abnorm. Psychol. 1 (1994) 92-102.

[3] G.F. Koob, F. Weiss, Pharmacology of drug self-administration, Alcohol 7 (1990) 1142– 8.

[4] J. Stewart, H. de Wit, R. Eikelboom, Role of unconditioned and conditioned drug effects in the self-administration of opiates and stimulants, Psychol. Rev. 91 (1984) 251– 68.

[5] S. Amigó, A. Caselles, J.C. Micó, The General Factor of Personality Questionnaire (GFPQ): Only one factor to understand the personality?, Span. J. Psychol. (2010) 5–17.

[6] S. Amigó, J.C. Micó, A. Caselles, Five adjectives to explain the whole personality: a brief scale of personality, Rev. Int. Sist. 16 (2009) 41–43.

[7] S. Amigó, J.C. Micó, A. Caselles, Adjective scale of the unique personality trait: measure of personality as
an overall and complete system, in: Proc. 7th Congr. Eur. Syst. Union, Lisboa, 2008.

[8] S. Amigó, La teoría del rasgo único de personalidad. Hacia una teoría unificada del cerebro y la conducta (The unique-trait personality theory. Towards a unified theory of brain and conduct), Ed. Universitat Politècnica de València, 2005.

[9] S. Amigó, A. Caselles, J.C. Micó, A dynamic extraversion model. The brain's response to a single dose of a stimulant drug, Br. J. Math. Stat. Psychol. 61 (2008) 211–231.

[10] A. Caselles, J.C. Micó, S. Amigó, Cocaine addiction and personality: A mathematical model, Br. J. Math.
Stat. Psychol. 63 (2010) 449–480.

[11] R.L. Solomon, J.D. Corbit, An opponent-process theory of motivation: I. Temporal dynamics of affect, Psychol. Rev. 81 (1974) 119–145.

[12] S. Grossberg, The imbalanced brain: from normal behavior to schizophrenia, Biol. Psychiatry. 48 (2000) 81–98.

[13] L. Pohorecky, Biphasic action of ethanol. Bio behavioral Reviews, 1 (1977) 231–240.

[14] T.W. Rall, (1990), Hypnotics and sedatives: ethanol, Goodman and Gilman's the pharmacological basis of therapeutics, 8th ed. New York: Pergamon. pp. 345– 82, Ed. A.G. Gilman T.W. Rall, A.S. Nies P. Taylor.

[15] D.B. Newlin, J.B. Thomson, Alcohol challenge with sons of alcoholics: a critical review and analysis,
Psychol. Bull. 108 (1990) 383–402.

[16] J.C. Micó, S. Amigó, A. Caselles, Changing the General Factor of Personality and the c-fos Gene Expression with Methylphenidate and Self-Regulation Therapy, Span. J. Psychol. 15 (2012) 850–867.

[17] J.C. Micó, A. Caselles, S. Amigó, A. Cotolí, M.T. Sanz, A Mathematical Approach to the Body-Mind Problem from a System Personality Theory (A Systems Approach to the Body-Mind Problem), Syst. Res. Behav. Sci. 30 (2013) 735–749.

[18] A. Caselles, J.C. Micó, S. Amigó, Dynamics of the General Factor of Personality in response to a single dose of caffeine, Span. J. Psychol. 14 (2011), 675-692.

[19] J.C. Micó, S. Amigó, A. Caselles, From the Big Five to the General Factor of Personality: a Dynamic Approach, Span. J. Psychol. 17 (2014) E74 1-18.

An optimal eighth-order scheme for multiple roots applied to some real life problems

Ramandeep Behl [a], Eulalia Martínez [b], Fabricio Cevallos [c], Ali Saleh Alshomrani [a]

[a] *Department of Mathematics, King Abdulaziz University, Jeddah, Saudi Arabia*

[b] *Instituto Universitario de Matemtica Multidisciplinar, Universitat Politècnica de València, Spain*

[c] *Fac. de Ciencias Económicas, Universidad Laica "Eloy Alfaro" de Manabí, Ecuador*

November 30, 2018

1   Introduction

The construction of higher-order optimal multi-point iterative methods for locating zeros of the nonlinear equation $f(x) = 0$ with multiplicity $m \geq 1$ when the involved function $f : D \subset \mathbb{C} \to \mathbb{C}$ is analytic in the region enclosing the required zero is one of the toughest, most challenging and most important tasks in the field of numerical analysis.

In the recent and past years, many researchers have tried to construct an optimal eighth-order iterative scheme for multiple zeros with multiplicity $m \geq 1$. There are few multi-point iterative schemes/families reaching high order of convergence. Our mean to say that some multi-point iterative schemes for multiple zeros who attain maximum sixth-order convergence. All these schemes were proposed in the recent years.

These schemes use four functional evaluations in order to attain sixth-order convergence with the efficiency index $6^{\frac{1}{4}} = 1.5650$. According to the classical Kung-Traub's conjecture [1], all these schemes are non optimal. And not everyone works with multiplicity $m = 1$. So, we need an optimal eighth-order scheme which will work for multiple zeros ($m > 1$) as well as for simple zeros ($m = 1$). The better efficiency index compared to existing methods of lower order and the small number of iterations required in order to obtained the desired accuracy make eight-order shceme an interesting field of study.

Motivated and inspired by this, we present an optimal scheme with eighth-order convergence which will work for multiple zeros with multiplicity $m \geq 1$. The proposed scheme is the extension of Chun and Neta's scheme [4].

2   Development of an optimal eighth-order scheme

An optimal eighth-order scheme proposed by Chun and Neta [4] for simple zeros can be extended for multiple zeros by implementing it in the following form

$$
\begin{aligned}
y_n &= x_n - mu_n, \\
z_n &= x_n - mu_n \left[v_n^2 - \frac{1}{v_n - 1} \right], \\
x_{n+1} &= z_n - mt_n u_n \left[\phi_f(v_n) + \frac{t_n}{v_n - at_n} + 4t_n \right],
\end{aligned} \tag{1}
$$

where the weight function $\phi_f : \mathbb{C} \to \mathbb{C}$ is an analytic function in a neighborhood of zero, with $u_n = \frac{f(x_n)}{f'(x_n)}$, $v_n = \left(\frac{f(y_n)}{f(x_n)} \right)^{\frac{1}{m}}$, $t_n = v_n \left(\frac{f(z_n)}{f(y_n)} \right)^{\frac{1}{m}}$ and the free parameter $a \in \mathbb{R}$. The Chun and Neta's [4] scheme is a special case of the above algorithm for $m = 1$.

Theorem 1. *Let $x = \alpha$ be a multiple zero with multiplicity $m \geq 1$ of $f(x) = 0$, with $f : \mathbb{C} \to \mathbb{C}$ an analytic function in a region enclosing the required zero. Then, the scheme defined by (1) reaches eighth-order convergence if the following expressions hold*

$$
\phi(0) = 1, \quad \phi'(0) = 2, \quad \phi''(0) = 4, \quad \phi'''(0) = -6.
$$

Finally we obtain the optimal asymptotic error constant term, which is given as follows:

$$
\begin{aligned}
e_{n+1} = -\frac{c_1 \left((m+7)c_1^2 - 2mc_2 \right)}{48m^7} &\left[\left(\phi''''(0) + 6a(m+7)^2 - 14m^2 - 192m - 730 \right) c_1^4 \right. \\
&\left. - 24m \left(a(m+7) - 2(m+8) \right) c_1^2 c_2 + 24(a-1)m^2 c_2^2 - 24m^2 c_1 c_3 \right] e_n^8 + O(e_n^9),
\end{aligned}
$$

where $\phi''''(0),\ a \in \mathbb{R}$.

The above expression demonstrate that our proposed scheme reaches eighth-order convergence by using only four functional evaluations (viz. $f(x_n)$, $f'(x_n)$, $f(y_n)$ and $f(z_n)$) per iteration. Therefore, it is an optimal scheme according to Kung-Traub's conjecture.

2.1   Special cases of the proposed scheme

In this section, we will discuss some special cases of our proposed scheme (1) by assigning different weight functions ϕ_f. For example

1. We consider one weight function of the following form:

$$\phi(v) = \frac{1 - v^3}{1 - 2v + 2v^2}.$$

   Then, we find another optimal eighth-order iteration function, which is given as follows:

$$
\begin{aligned}
y_n &= x_n - mu_n, \\
z_n &= x_n - mu_n \left[v_n^2 - \frac{1}{v_n - 1} \right] \\
x_{n+1} &= z_n - mt_n u_n \left[\frac{1 - v_n^3}{1 - 2v_n + 2v_n^2} + \frac{t_n}{v_n - at_n} + 4t_n \right].
\end{aligned}
\tag{2}
$$

2. We consider

$$\phi(v) = \frac{v + 1}{3v^3 - v + 1},$$

   with the above weight function, we will obtain a new optimal eighth-order iteration function, which is given as below:

$$
\begin{aligned}
y_n &= x_n - mu_n \\
z_n &= x_n - mu_n \left[v_n^2 - \frac{1}{v_n - 1} \right] \\
x_{n+1} &= z_n - mt_n u_n \left[\frac{v + 1}{3v^3 - v + 1} + \frac{t_n}{v_n - at_n} + 4t_n \right].
\end{aligned}
\tag{3}
$$

3   Numerical experiments

To conclude, we will check the efficiency and effectiveness of our proposed scheme with the weight functions. Therefore, we choose some of the expressions from our scheme (1), namely, the expression (2) for $\left(a = 1, \frac{2(m+8)}{m+7}, \frac{7m^2 + 96m + 437}{3(m+7)^2} \right)$ and the expression (3) for $\left(a = 1, \frac{2(m+8)}{m+7} \right)$, with what we get the methods $PM1, PM2, PM3, PM4$ and $PM5$, respectively. Here we have an application example:

Example 1. We consider one standard nonlinear test function, which is given as follows:

$$f_1(x) = \frac{\left(x - \sqrt{5}\right)^4}{(x - 1)^2 + 1}.$$

The above function has a multiple zero at $x = \sqrt{5}$ of multiplicity 4. We have chosen the initial approximation $x_0 = 2.5$.

$f(x)$	n	$PM1$	$PM2$	$PM3$	$PM4$	$PM5$
$f_1(x)$	1	$1.1(-6)$	$3.4(-5)$	$4.2(-5)$	$1.1(-6)$	$3.4(-5)$
	2	$1.2(-55)$	$1.4(-45)$	$5.8(-46)$	$1.5(-55)$	$2.2(-45)$
	3	$4.1(-447)$	$2.9(-368)$	$1.8(-372)$	$2.3(-446)$	$1.6(-366)$

Table 1: Comparison based on residual error (i.e. $|f(x_n)|$) of different iteration functions.

Acknowledgements

References

[1] H. T. Kung, J. F. Traub, *Optimal order of one-point and multipoint iteration*, J. Assoc. Comput. Mach. **21** (1974) 643–651.

[2] Y. H. Geum, Y. I. Kim, B. Neta, *A class of two-point sixth-order multiple-zero finders of modified double-Newton type and their dynamics*, Appl. Math. Comput. **270** (2015) 387–400.

[3] Y. H. Geum, Y. I. Kim, B. Neta, *A sixth-order family of three-point modified Newton-like multiple-root finders and the dynamics behind their extraneous fixed points*, Appl. Math. Comput. **283** (2016) 120–140.

[4] C. Chun, B. Neta, *An analysis of a family of Maheshwari-based optimal eighth-order methods*, Appl. Math. Comput. **253** (2015) 294–307.

[5] G. V. Balaji, J. D. Seader, *Application of interval Newton's method to chemical engineering problems*, Rel. Comput. **1 (3)** (1995) 215–223.

[6] M. SHACHAM, *An improved memory method for the solution of a non-linear equation*, Chem. Eng. Sci. **44 (7)** (1989) 1495–1501.

Optimal Control of Plant Virus Propagation

B. Chen-Charpentier[♭] [*], and M. Jackson[†]

(♭)(†) Department of Mathematics, University of Texas at Arlington,

Arlington, TX 76019-0408.

November 30, 2018

1 Introduction

Plants are a food source for man and many species. They also are sources of medicines, fibers for clothes, and are essential for a healthy environment. But plants are subject to diseases many of which are caused by viruses. These viruses often kill the plant. As a result, billions of dollars are lost every year because of virus related crop loss. Most of the time, virus propagation is done by a vector, usually insects that bite infected plants, get themselves infected and then bite susceptible plants. Insect vectors typically have a seasonal behavior. They are very active in the warm months and not very active, almost dormant, in the cool months. To combat the vectors, chemical insecticides are commonly used as a control. Unfortunately, these chemicals not only are expensive but also have toxic effects on humans, animals and the environment in general. An alternative is to introduce a predator species, or just increase the number of a naturally present one, to prey on the insects and limit the spread of the virus. A combination of insecticide and predators can be used to control the vector population. The question is whether there is an optimal combination.

In our study we consider six populations: susceptible, infected and recovered plants, susceptible and infected vectors, and predators. We assume that the susceptible plants can become infected if an infected insect feeds and is

---

[*]e-mail:bmchen@uta.edu

able to transmit the virus to the plant; the infected plant will either die from the virus or recover; a healthy vector will can obtain the virus by feeding on an infected plant; the infected vectors have no ill effects from the virus so they do not fight the virus and therefore they do not recover; and the virus does not affect the predators. Also, the total number of plants is assumed to be constant since the farmers will replace a dead plant with a healthy one. The plant populations can be determined using the total constant plant population.

We first introduce a mathematical model of ordinary differential equations describing the interaction between plants, vectors and predators. This model can be used with constant coefficients or with periodic coefficients such as the infection and birth and death rates. To determine the optimal amount of predators to introduce and insecticide to use, an objective function giving the total cost to the farmer of the disease. This function depends on the number of infected plants and on the cost of the predators and the insecticide. The cost of the insecticide can also include an environmental cost. We find the controls that minimize the objective function subject to the population variables satisfying the differential equation model and initial conditions, together with constraints such that the controls are nonnegative.

There are two main methods of determining the optimum cost. One is using indirect methods. This approach is based on Pontryagin maximum principle.

Because this method can present convergence issues, we also consider a direct method to solve the problem. Direct methods have the advantage over indirect methods in that they are more straightforward to apply and more robust with respect to the initialization. The cost, however, is that some precision is lost [1]. The direct methods transforms the infinite dimensional optimal control problem into a finite dimensional problem. To do this, the direct method constructs approximations of the state and control variables which are substituted into the objective function and dynamics equations to obtain an optimization problem in many variables. The BOCOP software discretizes the equations and the variables giving the user several choices. It then utilizes the IPOPT solver that implements a primal-dual interior point algorithm [7] to solve the discrete nonlinear optimization problem.

2   Assumptions and Mathematical Model

With similar assumptions as in [2], [3], [4], [6], we consider 6 populations: susceptible plants $S(t)$, infected plants $I(t)$, recovered plants $R(t)$, susceptible insect vectors $X(t)$, infected insect vectors $Y(t)$, and predators $P(t)$. Each variable describes it's respective population at time t. Susceptible plants do not have the disease but could contract the disease if infected with the virus. The infected plants have the virus but cannot directly transmit the virus to susceptible plants. Infected plants can either die from the disease or recover. Additionally, since the infected plants can die from the viral infection their death rate is higher than that of plants that do not have the virus. We also assume that as soon as a plant dies either from the infection or from a natural death, it is immediately replaced with a new susceptible plant by a farm worker. Thus it is reasonable to assume that the plant population remains fixed and the total plant population will be denoted by K. This assumption has the modeling advantage that $K = S(t) + I(t) + R(t)$ can be used to eliminate the recovered population from the system of equations. The susceptible insects do not have the virus but can obtain the virus if they come in contact with a infected plant. Infected insects can transmit the virus to susceptible plants upon contact. We assume no vertical transmission of the virus with neither plants nor vectors. Moreover, we assume that the virus does not harm the vector and thus the vector does not defend against the virus and it retains the virus for the rest of its life. Since the insects do not show signs of being infected, the predators cannot differentiate between healthy and infected insects. Thus, we assume that the predators consume both infected and healthy insects at the same rate. The interaction between vector and plant as well as that of predator and vector are assumed to have a limitation of the form of predator-prey Holling type 2. The following table lists the parameters in the model.

The following is system of ordinary differential equations modeling the biological situation:

$$\frac{dS}{dt} = \mu(K - S) + dI - \frac{\beta Y}{1 + \alpha Y}S$$

$$\frac{dI}{dt} = \frac{\beta Y}{1 + \alpha Y}S - (d + \mu + \gamma)I$$

$$\frac{dX}{dt} = \Lambda - \frac{\beta_1 I}{1 + \alpha_1 I}X - \frac{c_1 X}{1 + \alpha_3 X}P - mX - d_{in}(t)X$$

$$\frac{dY}{dt} = \frac{\beta_1 I}{1 + \alpha_1 I}X - \frac{c_2 Y}{1 + \alpha_3 Y}P - mY - d_{in}(t)Y$$

$$\frac{dP}{dt} = \Lambda_p + \frac{\alpha_4 c_1 X}{1 + \alpha_3 X}P + \frac{\alpha_4 c_2 Y}{1 + \alpha_3 Y}P - \delta P - \epsilon P^2$$

Our first goal is to minimize the cost of insecticide and infected plants. To achieve such goals, we use the above equations as constraints to an objective function. We consider the case when insecticide is the only control used in the minimization. To do so, we want to minimize the cost functional

$$\int_0^T AI(t)^2 + Gd_{in}(t)^2 dt.$$

3 Methods for Solving the Optimization Problem

We first tried the indirect method based on Pontryagin's principle [5]. We had convergence difficulties for times greater than a 100 days. Therefore we consider a direct method to solve the problem. The idea is to discretize the control problem, then apply Nonlinear Programming (NLP) techniques to the resulting finite-dimensional optimization problem.

There are several software packages for direct methods.We chose BOCOP [1] that uses C++ and includes a GUI. Solves problems with multiple controls and delay equations, but one has to write several routines [1]. We succesfully solved the problem with no delays, no seasonality and only one control for time up to at least 365 days. Therefore we use BOCOP to solve the full problem with delays, seasonality and two controls.

4 Direct Methods Solving the Delay Optimization Problem with Seasonality

Since it takes time for the virus to spread throughout the plant and insect, we now consider an optimal control problem with delays and seasonality. Let τ_1 be the time it takes a plant to become infected after contagion and τ_2, to be the time it takes a vector to become infected after contagion. Then the problem with the two discrete delays and seasonality is

$$\min_{d_{in}(t),\Lambda_p} \int_0^T AI(t)^2 + Gd_{in}(t)^2 + F\Lambda_p^2 dt$$

subject to

$$\frac{dS}{dt} = \mu(K - S) + dI - \frac{\beta(t)Y(t - \tau_1)}{1 + \alpha Y(t - \tau_1)}S(t - \tau_1)$$

$$\frac{dI}{dt} = \frac{\beta(t)Y(t - \tau_1)}{1 + \alpha Y(t - \tau_1)}S - (d + \mu + \gamma)I$$

$$\frac{dX}{dt} = \Lambda - \frac{\beta_1(t)I(t - \tau_2)}{1 + \alpha_1 I(t - \tau_2)}X(t - \tau_2) - \frac{c_1 X}{1 + \alpha_3 X}P - mX$$

$$\frac{dY}{dt} = \frac{\beta_1(t)I(t - \tau_2)}{1 + \alpha_1 I(t - \tau_2)}X(t - \tau_2) - \frac{c_2 Y}{1 + \alpha_3 Y}P - mY$$

$$\frac{dP}{dt} = \frac{\alpha_4 c_1 X}{1 + \alpha_3 X}P + \frac{\alpha_4 c_2 Y}{1 + \alpha_3 Y}P - \delta P - \epsilon P^2,$$

where

$$\beta(t) = \beta(1 + h\cos(\frac{2\pi t}{365})) \quad \beta_1(t) = \beta_1(1 + h\cos(\frac{2\pi t}{365})). \tag{1}$$

5 Conclusions

For our particular models, the direct optimal control methods implemented in BOCOP are more robust than the indirect methods using Pontryagin maximum principle. The plant virus propagation model presented which includes periodicity and delays to make the model more realistic but at the cost of making the model more complex. However, with the BOCOP software, we

are able to calculate the state values and control functions at the optimal solution for specific parameter values. Even though we know of no measured or calculated values for some of the model parameters, we hope that this work will encourage farmers to measure the parameters necessary to determine an optimal cost for a particular situation. We also showed that relative costs are important and, as the cost for pesticides and predators change, so does the optimal controls. While the simulations only provide specific examples, they can give insight to farmers who want to minimize the cost of virus disease to plants.

References

[1] J.F. Bonnans, V. Grelard, and P. Martinon, Bocop, the optimal control solver, Open source toolbox for optimal control problems., URL http://bocop. org. (2011)

[2] M. Jackson, B.M. Chen-Charpentier, Modeling plant virus propagation with delays, Journal of Computational and Applied Mathematics. 309 611–621. (2016)

[3] M. Jackson, B.M. Chen-Charpentier, A model of biological control of plant virus propagation with delays, Journal of Computational and Applied Mathematics. 330:855–65. (2018)

[4] M. Jackson, B.M. Chen-Charpentier, Modeling Plant Virus Propagation with Seasonality, Journal of Computational and Applied Mathematics. 345: 310–319. (2019)

[5] S. Lenhart, J. Workman, *Optimal Control Applied to Biological Models*, Taylor and Francis Group, LLC. Boca Raton, FL (2007)

[6] R. Shi, H. Zhao, S. Tang, Global Dynamic Analysis of a Vector-Borne Plant Disease Model, Advances in Difference Equations. 59. (2014)

[7] A. Wächter and L.T. Biegler. On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming, Mathematical Programming, 106(1):25–57. (2006)

On the inclusion of memory in Traub-type iterative methods for solving nonlinear equations[*]

F. I. Chicharro[†], A. Cordero, N. Garrido, and J. R. Torregrosa

Institute for Multidisciplinary Mathematics, Universitat Politècnica de València,

Camino de Vera s/n, 46022 València (Spain).

November 30, 2018

1   Introduction

A large amount of problems in Science and Engineering lack of an analytical solution. A way to proceed with this kind of problems is the application of iterative methods. There is a vast literature regarding this topic [1], and so many classifications based on their own features.

The quality of iterative methods can be analyzed in terms of the order of convergence p, the number of functional evaluations per step d or the optimality of the method [2] when $p = 2^{d-1}$. However, these parameters do not guarantee the stability of the method for every initial guess. This study can be performed with a dynamical analysis.

The inclusion of memory in the iterative schemes improves the order of convergence of the method [3, 4], at the expense of increasing the computational effort [5].

From the well-known non-optimal Traub's method of third-order of convergence [6], whose iterative expression is

$$
\begin{aligned}
y_k &= x_k - \frac{f(x_k)}{f'(x_k)}, \\
x_{k+1} &= y_k - \frac{f(y_k)}{f'(x_k)},
\end{aligned}
\tag{1}
$$

two schemes are designed for the inclusion of memory in order to increase the order of convergence.

2   Inclusion of memory in Traub-type methods

The first step for the improvement of the order requires the inclusion of an accelerating parameter in (1), resulting in the method T1:

$$
\begin{aligned}
y_k &= x_k - \frac{f(x_k)}{f'(x_k) + \delta f(x_k)}, \\
x_{k+1} &= y_k - \frac{f(y_k)}{f'(x_k)},
\end{aligned}
$$

whose error equation is

$$
e_{k+1} = 2c_2(c_2 + \delta)e_k^3 + \mathcal{O}(e_k^4), \tag{2}
$$

being $e_k = x_k - \alpha$, α represents the solution of $f(x)$ and $c_j = \frac{f^{(j)}(\alpha)}{j!f'(\alpha)}, j \geq 2$. Focusing on the error equation (2), if $\delta = -c_2$, the order of convergence reaches the value, at least, four. However, since α is unknown, obtaining an approximation of $f'(\alpha)$ and $f''(\alpha)$ is mandatory. This approximation is the key point of the memory.

    Different approximations of δ and, consequently, of $f'(\alpha)$ and $f''(\alpha)$, can be applied. For instance, if a linear approximation is performed,

$$
f'(\alpha) \approx f'(x_k), \quad f''(\alpha) \approx \frac{f'(x_k) - f'(x_{k-1})}{x_k - x_{k-1}},
$$

the accelerating parameter results in

$$
\delta_k = -\frac{1}{2}\frac{f'(x_k) - f'(x_{k-1})}{(x_k - x_{k-1})f'(x_k)},
$$

obtaining an iterative method with $p = 3.30$ when it replaces the parameter δ in T1. Using the Newton's interpolation polynomial of second order

$$
N_2(t) = f(x_k) + f[x_k, x_{k-1}](t - x_k) + f[x_k, x_{k-1}, y_{k-1}](t - x_k)(t - x_{k-1}),
$$

and approximating the derivatives by

$$
f'(\alpha) \approx N_2'(x_k), \quad f''(\alpha) \approx N_2''(x_k),
$$

the accelerating parameter has the expression

$$\delta_k = -\frac{N_2''(x_k)}{2N_2'(x_k)} = -\frac{f[x_k, x_{k-1}, y_{k-1}]}{f[x_k, x_{k-1}] + f[x_k, x_{k-1}, y_{k-1}](x_k - x_{k-1})}.$$

The resulting method, named TM1, has order of convergence 3.30.

Recalling (1), if two accelerating parameters are included in each step, it gives the method T2, whose iterative expression is

$$\begin{aligned}
y_k &= x_k - \frac{f(x_k)}{f'(x_k) + \delta_1 f(x_k)}, \\
x_{k+1} &= y_k - \frac{f(y_k)}{f'(x_k) + \delta_2 f(x_k)},
\end{aligned}$$

whose error equation is

$$e_{k+1} = (\delta_1 + c_2)(\delta_2 + 2c_2)e_k^3 + \mathcal{O}(e_k^4).$$

In a similar way to proceed as in the T1 case, for $\delta_1 = -c_2$ and $\delta_2 = 2\delta_1 = -2c_2$, the method has $p \geq 4$, but the information about α is not available. The linear and the Newton's interpolation polynomial approximations result in a method with order of convergence 3.56. For Newton's case, the expressions of the accelerating parameters are

$$\delta_{1k} = -\frac{N_2''(x_k)}{2N_2'(x_k)} = -\frac{f[x_k, x_{k-1}, y_{k-1}]}{f[x_k, x_{k-1}] + f[x_k, x_{k-1}, y_{k-1}](x_k - x_{k-1})},$$

and $\delta_{2k} = 2\delta_{1k}$. The resulting iterative scheme is called TM2.

3   Dynamical analysis

In order to analyze the stability of methods TM1 and TM2, their dynamical analysis is performed. Some fundamentals about dynamics in a real multidimensional scenario can be found in [4].

Let $\Psi_1 : \mathbb{R}^3 \to \mathbb{R}^3$ be the fixed point function associated to the method TM1, defined as

$$\Psi_1(x_{k-1}, y_{k-1}, x_k) = (x_k, y_k, x_{k+1}) = (x_k, y_k, \psi_1(x_{k-1}, y_{k-1}, x_k)), \qquad k \geq 1, \qquad (3)$$

where ψ_1 represents the method TM1. When TM1 is applied on a generic quadratic polynomial $p_c(x) = x^2 + c$, the fixed point operator results in

$$\Psi_1(z, w, x) = \left(x, y, -\frac{c^3 - 3c^2 x^2 + 23cx^4 - 5x^6}{2x\left(c - 3x^2\right)^2} \right), \qquad k \geq 1,$$

where $z = x_{k-1}$, $w = y_{k-1}$, $y = y_k$ and $x = x_k$.

A fixed point of Ψ_1 must satisfy $z = w = x$ and $x = \Psi_1(z, w, x)$. Therefore, when these conditions are applied to (3) the following one-dimensional operator is obtained:

$$\left[\Psi_1(z, w, x)\right]_{|z=w=x} = \tilde{\Psi}_1(x) = -\frac{c^3 - 3c^2x^2 + 23cx^4 - 5x^6}{2x\left(c - 3x^2\right)^2}.$$

The only real fixed points of $\tilde{\Psi}_1(x)$ are $x^F_{1,2}(c) = \mp i\sqrt{c}$ for $c < 0$, that match with the roots of $p_c(x)$ and their behavior is superattracting. By solving $\tilde{\Psi}'_1(x) = 0$, two free critical points $x^C_{1,2}(c) = \mp\sqrt{\frac{c}{15}}$ are obtained.

Let $\Psi_2 : \mathbb{R}^3 \to \mathbb{R}^3$ be the fixed point function associated to the method TM2. When TM2 is applied on $p_c(x)$ the fixed point operator is

$$\Psi_2(z, w, x) = \left(x, y, -\frac{2x(-2c^3 + 5c^2x^2 - 8cx^4 + x^6)}{(c - 3x^2)^2(c - x^2)}\right), \qquad k \geq 1.$$

Following the same procedure as method TM1, the one-dimensional operator of method TM2 is

$$\Psi_2(z, w, x)|_{z=w=x} = \tilde{\Psi}_2(x) = -\frac{(c - 15x^2)(c + x^2)^3}{2x^2(3x^2 - c)^3}.$$

The fixed points of $\tilde{\Psi}_2$ are the roots of $p_c(x)$, $x^F_{1,2}(c) = \mp i\sqrt{c}$ for $c < 0$, and also $x^F_3 = 0$ is a strange fixed point. By evaluating the fixed points in $|\tilde{\Psi}'_2(x)|$, $x^F_{1,2}(c)$ are superattracting and x^F_3 is a repelling point. The operator $\tilde{\Psi}_2$ has two free critical points: $x^C_{1,2}(c) = \pm\sqrt{\frac{2}{3}c}$.

An interesting representation of the stability of the methods is the dynamical plane. In its origins, it was devoted to complex dynamics [7]. However, it has been adapted to real dynamics [8] to represent dynamical lines. When the rational functions include one parameter, either by the iterative expression, or by the involved polynomial, the best tool is the convergence plane [9]. In any of the three cases, the information is similar. Each attracting fixed point is mapped to a different color. When an initial guess tends to one of the attracting points, it is illustrated with the corresponding color. Below there are some representations of convergence planes in Figure 1 varying the value of the parameter c, and some illustrations of dynamical lines in Figure 2, for specific cases of the value c. The roots $-i\sqrt{c}$ and $i\sqrt{c}$ are mapped with colors blue and orange, respectively. In the convergence planes, black and white lines represent the critical and strange points, respectively.

(a) Method TM1   (b) Method TM2
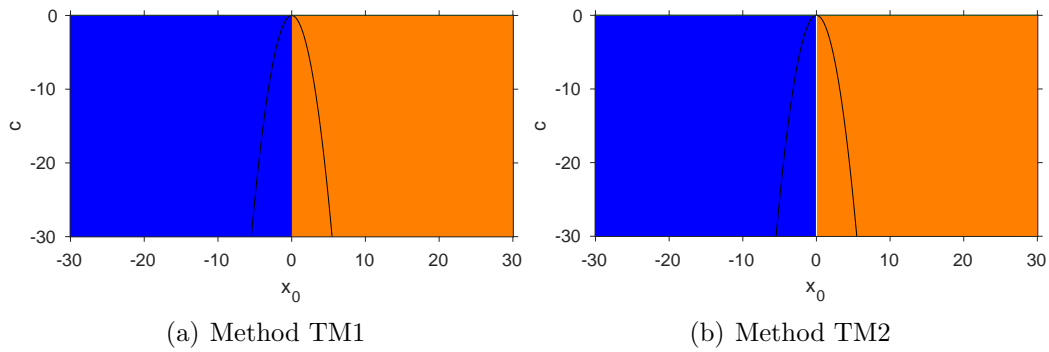
Figure 1: Convergence planes of methods on $p_c(x) = x^2 + c$.

Focusing on Figure 1, every initial guess tends to an attracting point, for both TM1 and TM2 methods, showing the wide stability of this schemes. Note that only the region of $c < 0$ has been represented. In Figure 2 two dynamical lines of both methods are included to capture the effect of $c > 0$. In this cases, every initial guess does not tend to any real root.

(a) TM1   (b) TM1

(c) TM2   (d) TM2
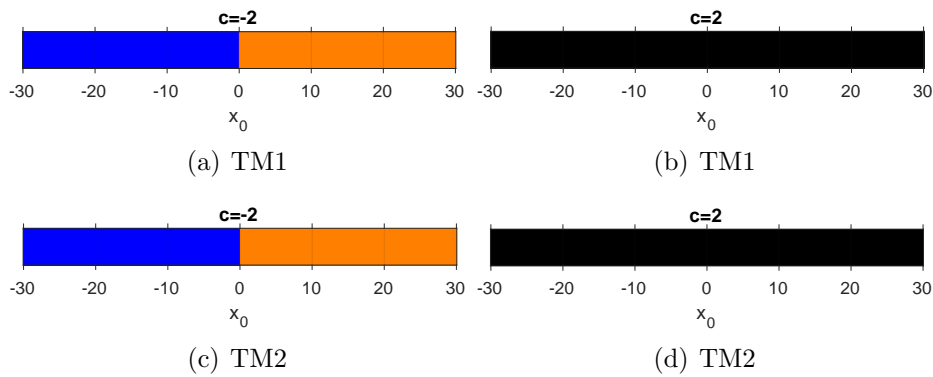
Figure 2: Dynamical lines of methods on $p_c(x) = x^2 + c$, for $c = \{-2, 2\}$.

4   Conclusions

This paper presents two iterative methods with memory based on the well-known Traub's method. In order to increase the order of convergence, the introduced methods include some acceleration parameters that involve the use of memory. The resulting schemes achieve order of convergence 3.30 and 3.56, improving the third

order of the original method. Moreover, an essential feature of the iterative methods is also proved. The stability of the method in terms of the election of an initial guess is checked with the help of dynamical analysis, showing for both methods two wide basins of attraction.

References

[1] A. Cordero and J. R. Torregrosa. Advances in iterative methods for nonlinear equations. Springer, 2016.

[2] H. T. Kung and J. F. Traub. Optimal order of one-point and multipoint iteration. *J. Assoc. Comput. Math.*, 21:643–651, 1974.

[3] M. S. Petković, B. Neta, L. D. Petković, and J. Džunić. Multipoint methods for solving nonlinear equations. Elsevier, 2013.

[4] B. Campos, A. Cordero, J. R. Torregrosa, and P. Vindel. A multidimensional dynamical approach to iterative methods with memory. *Applied Mathematics and Computation*, 318:701–715, 2015.

[5] C. Chun and B. Neta. How good are methods with memory for the solution of nonlinear equations? *SeMA Journal*, 74:613–625, 2017.

[6] J. F. Traub. Iterative methods for the solution of equations. Prentice-Hall, 1964.

[7] F. I. Chicharro, A. Cordero, and J. R. Torregrosa. Drawing dynamical and parameters planes of iterative families and methods. *The Scientific World Journal*, ID 780153:1–11, 2013.

[8] F. I. Chicharro, A. Cordero, J. R. Torregrosa, and M. P. Vassileva. King-type derivative-free iterative families: Real and memory dynamics. *Complexity*, ID 2713145:1–15, 2017.

[9] Á. A. Magreñán. A new tool to study real dynamics: the convergence plane. *Applied Mathematics and Computation*, 248:215–224, 2014.

Mean square analysis of non-autonomous second-order linear differential equations with randomness

J. Calatayud[♭], J.-C. Cortés[♭*], M. Jornet[♭],
L. Villafuerte[†]

(♭) Instituto Universitario de Matemática Multidisciplinar,

Universitat Politècnica de València, Spain,

(†) Department of Mathematics,

University of Texas at Austin, USA.

November 30, 2018

1   Introduction

In this paper, we deal with the random second order linear differential equation

$$\begin{cases} \ddot{X}(t) + A(t)\dot{X}(t) + B(t)X(t) = 0, \ t \in \mathbb{R}, \\ X(t_0) = Y_0, \\ \dot{X}(t_0) = Y_1. \end{cases} \quad (1)$$

The data coefficients $A(t)$ and $B(t)$ are stochastic processes and the initial conditions Y_0 and Y_1 are random variables on an underlying complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The solution of (1), $X(t)$, is a stochastic process as well.

The goals of this paper are the following:

---

*e-mail: jccortes@imm.upv.es

- To specify the meaning of the random differential equation (1) via the $L^p(\Omega)$ random calculus or more concretely using the so-called mean square calculus, that corresponds to $p = 2$.

- To find a proper stochastic process solution to (1).

- To compute its main statistical information (expectation and variance) under mild conditions.

Particular cases of (1) have been used in previous contributions using $L^p(\Omega)$ random calculus. For instance, Airy, Hermite, Legendre and Bessel differential equations have been randomized and rigorously studied in [1], [2], [3] and [4]. A very important case of problem (1) is when its coefficients are random variables rather than stochastic processes. In [5], the authors constructed the first and second PDF of the solution stochastic process.

The novelty of this article is that we solve the general form of a random second order linear differential equation via mean square power series. Some important equations studied in the literature, like Airy, Hermite, etc., will be particular cases of our theory.

2 Solving the random non-autonomous second order linear differential equation

We will assume that the data stochastic process $A(t)$ and $B(t)$ are analytic at t_0: $A(t) = \sum_{n=0}^{\infty} A_n(t - t_0)^n$ and $B(t) = \sum_{n=0}^{\infty} B_n(t - t_0)^n$, for $t \in (t_0 - r, t_0 + r)$, being $r > 0$ fixed, and the sum is understood in the $L^2(\Omega)$ setting. We search for a solution process $X(t)$ of the form $X(t) = \sum_{n=0}^{\infty} X_n(t - t_0)^n$, for $t \in (t_0 - r, t_0 + r)$, where the sum is in $L^2(\Omega)$.

Theorem 2.1 *Let $A(t) = \sum_{n=0}^{\infty} A_n(t - t_0)^n$ be a random power series in the $L^p(\Omega)$ setting ($p \geq 1$), for $t \in (t_0 - r, t_0 + r)$, $r > 0$. Then the random power series $\sum_{n=1}^{\infty} n A_n(t - t_0)^{n-1}$ exists in $L^p(\Omega)$ for $t \in (t_0 - r, t_0 + r)$, and the $L^p(\Omega)$ derivative of $A(t)$ is equal to it: $\dot{A}(t) = \sum_{n=1}^{\infty} n A_n(t - t_0)^{n-1}$, for all $t \in (t_0 - r, t_0 + r)$.*

Theorem 2.2 *Let $U = \sum_{n=0}^{\infty} U_n$ and $V = \sum_{n=0}^{\infty} V_n$ be two random series that converge in $L^2(\Omega)$. Suppose that one of the series converges absolutely,*

say $\sum_{n=0}^{\infty} \|V_n\|_{L^2(\Omega)} < \infty$. *Then*

$$\left(\sum_{n=0}^{\infty} U_n \right) \left(\sum_{n=0}^{\infty} V_n \right) = \sum_{n=0}^{\infty} W_n, \quad W_n = \sum_{m=0}^{n} U_{n-m} V_m,$$

where $\sum_{n=0}^{\infty} W_n$ *is understood in* $L^1(\Omega)$.

Theorem 2.3 *Let* $A(t) = \sum_{n=0}^{\infty} A_n(t-t_0)^n$ *and* $B(t) = \sum_{n=0}^{\infty} B_n(t-t_0)^n$ *be two random series in the* $L^2(\Omega)$ *setting, for* $t \in (t_0 - r, t_0 + r)$, *being* $r > 0$ *finite and fixed. Assume that the initial conditions* Y_0 *and* Y_1 *belong to* $L^2(\Omega)$. *Suppose that there is a constant* $C_r > 0$, *maybe dependent on* r, *such that* $\|A_n\|_{L^\infty(\Omega)} \leq C_r/r^n$ *and* $\|B_n\|_{L^\infty(\Omega)} \leq C_r/r^n$, $n \geq 0$. *Then the stochastic process* $X(t) = \sum_{n=0}^{\infty} X_n(t-t_0)^n$, $t \in (t_0 - r, t_0 + r)$, *where*

$$X_0 = Y_0, \quad X_1 = Y_1, \tag{2}$$

$$X_{n+2} = \frac{-1}{(n+2)(n+1)} \sum_{m=0}^{n} \left[(m+1)A_{n-m}X_{m+1} + B_{n-m}X_m \right], \; n \geq 0, \tag{3}$$

is the unique analytic solution to the random initial value problem (1) in the mean square sense.

The hypotheses concerning the $L^\infty(\Omega)$ growth of the coefficients A_n and B_n, $n \geq 0$, may seem quite restrictive. However, these hypotheses have been necessary to prove the main theorem. Moreover, these $L^\infty(\Omega)$ hypotheses are equivalent to a growth condition on the moments of the random variables $A_0, A_1, \ldots$ and $B_0, B_1, \ldots$: for a given random variable Z, we have that $\mathbb{E}[|Z|^n] \leq HR^n$ for certain $H > 0$ and $R > 0$, if and only if $\|Z\|_{L^\infty(\Omega)} \leq R$. This key fact is a direct consequence of the following result: if Z is a random variable, then $\lim_{n\to\infty} \|Z\|_{L^n(\Omega)} = \|Z\|_{L^\infty(\Omega)}$.

Growth hypotheses of the form $\mathbb{E}[|Z|^n] \leq HR^n$, for certain $H > 0$ and $R > 0$, are common in the literature to find stochastic analytic solutions to particular cases of (1). See for example Airy's random differential equation in [1] and Hermite's random differential equation in [2]. Hence, our main theorem will allow us to generalize the results obtained in the literature.

The hypotheses of our main theorem, besides providing a stochastic solution to our problem (1), also give a pointwise classical solution to (1) under the additional assumption $Y_0, Y_1 \in L^\infty(\Omega)$. This manner of studying random differential equations is referred to as the sample path approach, see [6, Appendix I].

3   Statistical information of the solution stochastic process: mean and variance

The expectation and variance of the stochastic process $X(t) = \sum_{n=0}^{\infty} X_n(t - t_0)^n$ given by (2)–(3) can be approximated. Indeed, first, one has to obtain X_n as a function of Y_0, Y_1, $A_0, \ldots, A_{n-1}$ and $B_0, \ldots, B_{n-1}$ by recursion via (3), for $n = 0, 1, \ldots, N$. After this, we construct a truncation

$$X_N(t) = \sum_{n=0}^{N} X_n(t - t_0)^n \tag{4}$$

of the solution stochastic process $X(t)$. Since $\lim_{N \to \infty} X_N(t) \to X(t)$ in $\mathrm{L}^2(\Omega)$, we have $\lim_{N \to \infty} \mathbb{E}[X_N(t)] = \mathbb{E}[X(t)]$ and $\lim_{N \to \infty} \mathbb{V}[X_N(t)] = \mathbb{V}[X(t)]$.

There are other approaches to approximate these statistics of $X(t)$: dishonest method, Monte Carlo simulations, gPC expansions, etc.

4   Examples

Example 4.1 Airy's random differential equation is the following:

$$\begin{cases} \ddot{X}(t) + AtX(t) = 0, \ t \in \mathbb{R}, \\ X(0) = Y_0, \\ \dot{X}(0) = Y_1, \end{cases} \tag{5}$$

where A, Y_0 and Y_1 are random variables. In [1], the hypothesis used in order to obtain a mean square analytic solution $X(t)$ is $\mathbb{E}[|A|^n] \leq HR^n$, $n \geq n_0$. Notice that this hypothesis is equivalent to $\|A\|_{\mathrm{L}^\infty(\Omega)} \leq R$.

Consider $A \sim \text{Beta}(2, 3)$ and Y_0, Y_1 independent random variables such that $Y_0 \sim \text{Normal}(1, 1)$ and $Y_1 \sim \text{Normal}(2, 1)$. In Table 1 and Table 2, we approximate the expectation and variance of the solution process $X(t)$ at different times t.

Example 4.2 Consider

$$\begin{cases} \ddot{X}(t) + (A_0 + A_1 t)\dot{X}(t) + (B_0 + B_1 t)X(t) = 0, \ t \in \mathbb{R}, \\ X(0) = Y_0, \\ \dot{X}(0) = Y_1, \end{cases} \tag{6}$$

t	$\mathbb{E}[X_{15}(t)]$	$\mathbb{E}[X_{16}(t)]$	dishonest	MC $50,000$	MC $100,000$
0.00	1	1	1	0.99701	1.00138
0.25	1.49870	1.49870	1.49870	1.49519	1.49976
0.50	1.98752	1.98752	1.98752	1.98353	1.98829
0.75	2.45108	2.45108	2.45102	2.44667	2.45160
1.00	2.86856	2.86856	2.86818	2.86383	2.86893
1.25	3.21494	3.21494	3.21339	3.21008	3.21534

Table 1: Approximation of the expectation of the solution stochastic process, Example 4.1.

t	$\mathbb{V}[X_{15}(t)]$	$\mathbb{V}[X_{16}(t)]$	MC $50,000$	MC $100,000$
0.00	1	1	0.99610	0.99530
0.25	1.06035	1.06035	1.05902	1.05642
0.50	1.23142	1.23142	1.23408	1.22793
0.75	1.49261	1.49261	1.50041	1.48944
1.00	1.81392	1.81392	1.82744	1.81127
1.25	2.15870	2.15870	2.17768	2.15721

Table 2: Approximation of the variance of the solution stochastic process, Example 4.1.

where $A_0 = 4$, $A_1 \sim \text{Uniform}(0, 1)$, $B_0 \sim \text{Gamma}(2, 2)$, $B_1 \sim \text{Bernoulli}(0.35)$, $Y_0 = -1$ and $Y_1 \sim \text{Binomial}(2, 0.29)$ are assumed to be independent.

In order for the hypotheses of our main theorem to be satisfied, the Gamma distribution will be truncated. For the Gamma distribution with shape and rate 2, it can straightforwardly be checked that the interval $[0, 4]$ contains approximately 99.7% of the observations. In Table 3 and Table 4, we approximate the main statistics of the solution stochastic process $X(t)$ at different times t.

t	$\mathbb{E}[X_{19}(t)]$	$\mathbb{E}[X_{20}(t)]$	dishonest	MC $50,000$	MC $100,000$
0.00	-1	-1	-1	-1	-1
0.25	-0.886467	-0.886467	-0.886418	-0.886789	-0.886432
0.50	-0.809269	-0.809269	-0.808743	-0.809370	-0.809219
0.75	-0.747589	-0.747589	-0.745742	-0.747321	-0.747526
1.00	-0.693453	-0.693453	-0.689284	-0.692816	-0.693375
1.25	-0.643943	-0.643944	-0.636462	-0.642985	-0.643845

Table 3: Approximation of the expectation of the solution stochastic process, Example 4.2.

t	$\mathbb{V}[X_{15}(t)]$	$\mathbb{V}[X_{16}(t)]$	MC $50,000$	MC $100,000$
0.00	0	0	0	0
0.25	0.0102077	0.0102074	0.0101172	0.0102664
0.50	0.0190996	0.0190999	0.0189214	0.0192053
0.75	0.0237400	0.0237403	0.0235191	0.0238499
1.00	0.0268721	0.0268711	0.0266311	0.0269620
1.25	0.0297852	0.0297465	0.0295049	0.0298201

Table 4: Approximation of the variance of the solution stochastic process, Example 4.2.

References

[1] J. C. Cortés, L. Jódar, F. Camacho, L. Villafuerte. *Random Airy type differential equations: Mean square exact and numerical solutions.* Computers & Mathematics with Applications, 60(5), 1237–1244 (2010).

[2] G. Calbo, J. C. Cortés, L. Jódar. *Random Hermite differential equations: Mean square power series solutions and statistical properties.* Applied Mathematics and Computation, 218(7), 3654–3666 (2011).

[3] G. Calbo, J. C. Cortés, L. Jódar, L. Villafuerte. *Solving the random Legendre differential equation: Mean square power series solution and its statistical functions.* Computers & Mathematics with Applications, 61(9), 2782–2792 (2011).

[4] J. C. Cortés, L. Jódar, L. Villafuerte. *Mean square solution of Bessel differential equation with uncertainties.* Journal of Computational and Applied Mathematics, 309, 383–395 (2017).

[5] M. C. Casabán, J. C. Cortés, J. V. Romero, M. D. Roselló. *Solving random homogeneous linear second-order differential equations: a full probabilistic description.* Mediterranean Journal of Mathematics, 13(6), 3817–3836 (2016).

[6] T. T. Soong. *Random Differential Equations in Science and Engineering.* Academic Press, New York, 1973.

A Statistical Model with a Lotka-Volterra Structure for Microbiota Data

I. Creus-Martí$^{\flat\dagger*}$, A. Moya$^{\flat\ddagger\sharp}$, and F.J. Santonja†

($\flat$) Instituto de Biología de Sistemas (I2Sysbio). Universitat de València-CSIC.

($\dagger$) Departamento de Estadística e Investigacin Operativa, Universitat de València.

($\ddagger$) Fundación para el Fomento de la Investigación Sanitaria y Biomédica
de la Comunidad Valenciana (FISABIO).

($\sharp$) CIBER en Epidemiología y Salud Pública (CIBEResp).

November 30, 2018

1 Introduction

Gut microbiota is the complex community of microorganisms that lives in the digestive tracts of humans and other animals. The composition of human gut microbiota changes over time, and modeling its structure and changes could be of great value in the rational design of microbiota-tailoring diets and therapies.

In this work, we present a statistical model with a Lotka-Volterra structure to predict microbiome dynamics of the growth of intestinal bacteria. This approach allows us to understand the ecological structure of the intestinal microbiota and make predictions.

Our proposal is based on Lotka-Volterra approach, which has been used in several works on microbiota, see for example [1–4]. However, in most of them compositional data conformation has not been considered. In this paper, the

---

*e-mail: icreus@alumni.uv.es

proposed model considers that the proportions (the relative abundances) follow a Dirichlet distribution, and that its time-varying parameters, after a proper transformation, presents a Lotka-Volterra dynamic structure. Therefore, our proposal includes both a Lotka-Volterra structure and compositional data consideration.

2 Data

The dataset includes sequencing counts of a marker gene (16S rRNA) which are put into correspondence to the available taxonomically annotated database of microbial genomes or genes. The result of this classification determinates the relative abundances of each specie. We consider the relative abundance of *Porphyromonadaceae*, *Prevotellaceae* and *Other* for 30 time points in a Spanish sixty-six year old male. The data are extracted from [5].

3 The Model

Let $\mathbf{y}_t = (y_{1t}, y_{2t}, \ldots, y_{Kt})$, where $y_{it} \in (0,1)$, and $i = 1, 2, \ldots, K$, is the relative abundance of specie i at time t, so that $y_{1t} + y_{2t} +, \ldots, + y_{Kt} = 1$. We consider that the vector $\mathbf{y}_{t|\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots \mathbf{y}_1}$, follows a Dirichlet distribution with positive parameters $\boldsymbol{\alpha}_t = (\alpha_{1t}, \alpha_{2t}, \ldots, \alpha_{Kt})$:

$$\mathbf{y}_t | \mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots \mathbf{y}_1 \sim \mathrm{Dir}(\boldsymbol{\alpha}_t)$$

In order to link $\boldsymbol{\alpha}_t$ with $\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots$, we define:

$$g(\alpha_{jt}) = \mu_{jt} := \epsilon_j \ln\left(\frac{y_{jt-1}}{y_{Kt-1}}\right) + \ln\left(\frac{y_{jt-1}}{y_{Kt-1}}\right) \cdot \left(M_{j1} \ln\frac{y_{jt-1}}{y_{Kt-1}} + \cdots + M_{jK-1} \ln\frac{y_{K-1t-1}}{y_{Kt-1}}\right)$$

Taking into account similar proposals presented in [6], we consider that the reparametrized vector of the time-varying parameters, $\boldsymbol{\alpha}_t$, follows a Lotka-Volterra structure. Note that M_{jl} represents the effect that specie j has upon specie l and the parameter ϵ_j is related to birth rate of the specie j. Our microbiota database is a compositional time series where the observation vector at each t sum up to 1. Traditionally, such time series have been modeled considering a log-ratio transformation of the observations.

Note that given $g(\alpha_{jt})$, with $j = 1, 2, \ldots, K - 1$, the vector of parameters is unidentifiable. In order to properly identify these parameters, we assume that $\alpha_{1t} + \alpha_{2t} + \ldots + \alpha_{Kt} = \tau$ for all time-point t. This assumption allows us to estimate $\boldsymbol{\alpha}_t$ with $\boldsymbol{\alpha}_t = g^{-1}(\boldsymbol{\mu}_t)$. In our case, and taking into account a Lotka-Volterra structure and the first-order Taylor approximation, we consider $g(\alpha_{jt}) = \mathrm{E}\big(\ln(y_{jt}/y_{Kt})\big) \approx \ln\big(\mathrm{E}(y_{jt}/y_{Kt})\big)$, and then $\ln\big(\alpha_{jt}/\alpha_{Kt}\big) = \mu_{jt}$. The inverse expressions are then given by:

$$\hat{\alpha}_{jt} = \frac{\tau\, e^{\hat{\mu}_{jt}}}{1 + e^{\hat{\mu}_{1t}} + e^{\hat{\mu}_{2t}} + \cdots + e^{\hat{\mu}_{K-1t}}}$$

$$\hat{\alpha}_{Kt} = \frac{\tau}{1 + e^{\hat{\mu}_{1t}} + e^{\hat{\mu}_{2t}} + \cdots + e^{\hat{\mu}_{K-1t}}}$$

Using the previously presented formulas, we can calculate $\mathrm{E}\big(y_{it}\big)$ and $\mathrm{Var}\big(y_{it}\big)$ as:

$$\mathrm{E}\big(y_{it}\big) = \frac{\hat{\alpha}_{it}}{\hat{\alpha}_{1t} + \hat{\alpha}_{2t} + \cdots + \hat{\alpha}_{Kt}}$$

$$\mathrm{Var}\big(y_{it}\big) = \frac{\alpha_{it}\big(\tau - \alpha_{it}\big)}{\tau^2\big(\tau + 1\big)}$$

3.1 Model estimation

We have considered maximum likelihood estimation. Let β be the vector of the model parameters, $\beta = (\epsilon_1, \epsilon_2, M_{11}, M_{12}, M_{21}, M_{22}, \tau)$. Note that in our scenario, K=3. The log-likelihood function is defined by $L_T(\beta) = \sum_{t=1}^{T} \ell_t(\beta)$, where

$$\ell_t(\beta) = \ln\big(\Gamma(\tau)\big) - \sum_{i=1}^{3} \ln\big(\Gamma(\alpha_{it})\big) + \sum_{i=1}^{3} \big(\alpha_{it} - 1\big)\ln(y_{it}) \qquad t = 1, 2, \ldots, T = 30$$

We maximize $L_T(\beta)$ using the Nelder-Mead simplex method with the function *optim* of R [7]. We obtain the following estimations:

Parameter	Estimation
ϵ_1	1.514990
ϵ_2	1.130968
M_{11}	0.285277
M_{12}	-0.006742
M_{21}	-0.125196
M_{22}	0.0910394
τ	168.21223

Table 1: Paremeter estimates.

4 Results

The values of y_{it} (dots), $\mathrm{E}(y_{it})$ (blue line) and $\mathrm{E}(y_{it}) \pm \sqrt{\mathrm{Var}(y_{it})}$ (green lines) are shown in Figure 1. We can observe that the expected values show a good agreement with the original data.

References

[1] Marino, S. Baxter, N.T., Huffnagle G.B. et al. Mathematical modeling of primary succession of murine intestinal microbiota. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, 111: 439–444, 2014.

[2] Stein, R.R., Bucci, V., Toussaint, N.C. et al. Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota. *PLOS Computational Biology*, 9(12): 1–11, 2013.

[3] Alshawaqfeh, M., Serpedin, E. and Younes, A.B. Inferring microbial interaction networks from metagenomic data using SgLV-EKF algorithm. *BMC Genomics*, 18(3): 228.

[4] Gibson, T.E. and Gerber, G.K. Robust and scalable models of microbiome dynamics. In *arXiv preprint arXiv:1805.04591*, 2018 [`https://arxiv.org/abs/1805.04591`]

[5] Durbn, A., Abelln, J., Jimnez-Hernndez et al. Instability of the faecal microbiota in diarrhoea-predominant irritable bowel syndrome. *FEMS Microbiology Ecology*, 86(3): 581–589, 2013.

[6] Zheng, T., and Chen, R. Dirichlet ARMA models for compositional time series. *Journal of Multivariate Analysis*, 158: 31–46, 2017.

[7] R Core Team, R: A Language and Environment for Statistical Computing. Vienna, R Foundation for Statistical Computing, 2018. [`https://www.R-project.org`]
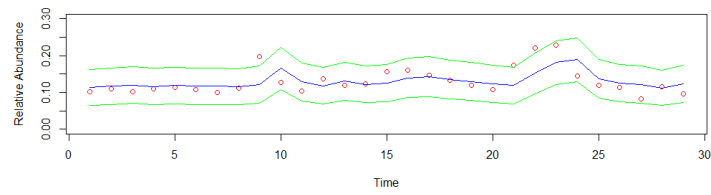
(a) Porphyromonadaceae
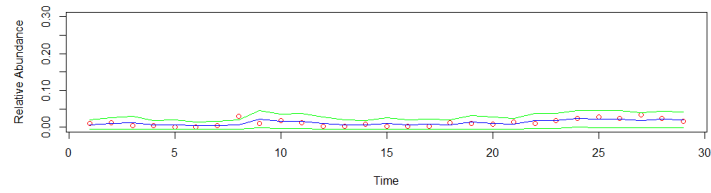
(b) Prevotellaceae
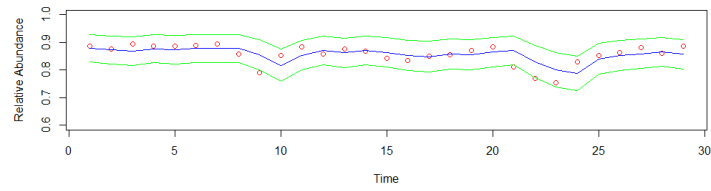
(c) Other

Figure 1: Available data and predictions.

Some new Hermite matrix polynomials series expansions and their applications in hyperbolic matrix sine and cosine approximation [*]

E. Defez[⋆], J. Ibáñez[†], J. Peinado[§], P. Alonso[♮], J. M. Alonso[†], J. Sastre[‡]

[⋆] Instituto de Matemática Multidisciplinar.

[†] Instituto de Instrumentación para Imagen Molecular.

[§] Departamento de Sistemas Informáticos y Computación.

[♮] Grupo Interdisciplinar de Computación y Comunicaciones.

[‡] Instituto de Telecomunicaciones y Aplicaciones Multimedia.

Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia, España.

edefez@imm.upv.es, { jjibanez, jpeinado, palonso, jmalonso }@dsic.upv.es, jorsasma@iteam.upv.es

1 Introduction and notation

Hermite matrix polynomial $H_n(x, A)$ has the generating function, see [1]:

$$e^{xt\sqrt{2A}} = e^{t^2} \sum_{n \geq 0} \frac{H_n(x, A)}{n!} t^n,$$ (1)

from following expressions for the matrix hyperbolic sine and cosine are derived:

$$\left. \begin{array}{rcl} \cosh\left(xt\sqrt{2A}\right) & = & e^{t^2} \sum_{n \geq 0} \dfrac{H_{2n}(x, A)}{(2n)!} t^{2n} \\ \sinh\left(xt\sqrt{2A}\right) & = & e^{t^2} \sum_{n \geq 0} \dfrac{H_{2n+1}(x, A)}{(2n+1)!} t^{2n+1} \end{array} \right\} , \; x \in \mathbb{R}, |t| < \infty.$$ (2)

Recently we have shown the following formulas which are a generalization of formulas (2):

$$\left.\begin{aligned}
\sum_{n \geq 0} \frac{H_{2n+1}(x, A)}{(2n)!} t^{2n} &= e^{-t^2}\left[H_1(x, A) \cosh\left(xt\sqrt{2A}\right) - 2t \sinh\left(xt\sqrt{2A}\right)\right], \\
\sum_{n \geq 0} \frac{H_{2n+2}(x, A)}{(2n+1)!} t^{2n+1} &= e^{-t^2}\left[H_1(x, A) \sinh\left(xt\sqrt{2A}\right) - 2t \cosh\left(xt\sqrt{2A}\right)\right], \\
\sum_{n \geq 0} \frac{H_{2n+3}(x, A)}{(2n+1)!} t^{2n+1} &= e^{-t^2}\left[\left(H_2(x, A) + 4t^2 I\right) \sinh\left(xt\sqrt{2A}\right) - 4t H_1(x, A) \cosh\left(xt\sqrt{2A}\right)\right].
\end{aligned}\right\}$$

(3)

We will use formulas (3) to obtain a new expansion of the hyperbolic matrix sine and cosine in Hermite matrix polynomials series.

Throughout this paper, we denote by $\mathbb{C}^{r \times r}$ the set of all the complex square matrices of size r. We denote by Θ and I, respectively, the zero and the identity matrix in $\mathbb{C}^{r \times r}$. If $A \in \mathbb{C}^{r \times r}$, we denote by $\sigma(A)$ the set of all the eigenvalues of A. For a real number x, $\lfloor x \rfloor$ denotes the lowest integer not less than x and $\lceil x \rceil$ denotes the highest integer not exceeding x.

We recall that for a positive stable matrix $A \in \mathbb{C}^{r \times r}$ the $n-$th Hermite matrix polynomial is defined in [1] by:

$$H_n(x, A) = n! \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} \frac{(-1)^k \left(\sqrt{2A}\right)^{n-2k}}{k!(n-2k)!} x^{n-2k},$$

(4)

which satisfies the three-term matrix recurrence:

$$\left.\begin{aligned}
H_m(x, A) &= x\sqrt{2A} H_{m-1}(x, A) - 2(m-1) H_{m-2}(x, A) \ , \ m \geq 1, \\
H_{-1}(x, A) &= \Theta \ , \ H_0(x, A) = I \ .
\end{aligned}\right\}$$

(5)

2 Some new Hermite matrix series expansions for the hyperbolic matrix cosine and sine

Let $A \in \mathbb{C}^{r \times r}$ be a positive stable matrix, then the matrix polynomial $H_1(x, A) = \sqrt{2A}x$ is invertible if $x \neq 0$. Substituting $\sinh\left(xt\sqrt{2A}\right)$ given in (2) into the first expression of (3) we obtain the following new rational expression for the hyperbolic matrix cosine in terms of Hermite matrix polynomials:

$$\cosh\left(xt\sqrt{2A}\right) = e^{t^2}\left(\sum_{n\geq 0}\frac{H_{2n+1}(x,A)}{(2n)!}\left(1+\frac{2t^2}{2n+1}\right)t^{2n}\right)[H_1(x,A)]^{-1},$$

$$x\in\mathbb{R}\sim\{0\}, |t|<+\infty. \tag{6}$$

Substituting $\sinh\left(xt\sqrt{2A}\right)$ given in (2) into the second expression of (3) and using the three-term matrix recurrence (5) we obtain the expression of $\cosh\left(xt\sqrt{2A}\right)$ given in (2).

On the other hand, replacing the expression of $\sin\left(xt\sqrt{2A}\right)$ given in (2) into the third expression of (3), we obtain another new rational expression for the hyperbolic matrix cosine in terms of Hermite matrix polynomials:

$$\cosh\left(xt\sqrt{2A}\right) =$$

$$= \frac{-e^{t^2}}{4}\left[\sum_{n\geq 0}\frac{H_{2n+3}(x,A)}{(2n+1)!}t^{2n} - \left(H_2(x,A)+4t^2 I\right)\star\left(\sum_{n\geq 0}\frac{H_{2n+1}(x,A)}{(2n+1)!}t^{2n+1}\right)\right][H_1(x,A)]^{-1},$$

$$x\in\mathbb{R}\sim\{0\}, |t|<+\infty. \tag{7}$$

Comparing (7) with (6), we observe that it always has a matrix product more when evaluating (7), the matrix product remarked by symbol "$\star$" in (7). Due to the importance of reducing the number of matrix products, see [2–4] for more details, we will focus mainly on the expansion (6).

From (4), it follows that, for $x\neq 0$:

$$H_{2n+1}(x,A)[H_1(x,A)]^{-1} = \frac{(2n+1)!}{x}\sum_{k=0}^{n}\frac{(-1)^k x^{2(n-k)+1}(2A)^{n-k}}{k!(2(n-k)+1)!}$$

$$= \widetilde{H}_{2n+1}(x,A), \tag{8}$$

where

$$\widetilde{H}_n(x,A) = n!\sum_{k=0}^{\lfloor\frac{n}{2}\rfloor}\frac{(-1)^k\left(\sqrt{2A}\right)^{n-2k-1}}{k!(n-2k)!}x^{n-2k}, \tag{9}$$

so the right side of (8) is still defined in the case where the matrix A is singular. In this way, we can re-write the relation (6) in terms of the matrix polynomial $\widetilde{H}_{2n+1}(x,A)$:

$$\cosh\left(xt\sqrt{2A}\right) \;=\; e^{t^2}\left(\sum_{n\geq 0}\frac{\widetilde{H}_{2n+1}\left(x,A\right)}{(2n)!}\left(1+\frac{2t^2}{2n+1}\right)t^{2n}\right),\tag{10}$$

$$x\in\mathbb{R}, |t|<+\infty.$$

Replacing the matrix A by matrix $A^2/2$ in (10) we can avoid the square roots of matrices, and taking $x=\lambda, \lambda\neq 0, t=1/\lambda$, we finally obtain

$$\cosh\left(A\right)=e^{\frac{1}{\lambda^2}}\left(\sum_{n\geq 0}\frac{\widetilde{H}_{2n+1}\left(\lambda,\frac{1}{2}A^2\right)}{(2n)!\lambda^{2n+1}}\left(1+\frac{2}{(2n+1)\lambda^2}\right)\right), 0<\lambda<+\infty.\tag{11}$$

3 Numerical approximations

Truncating the given series (11) until order m, we obtain the approximation $CH_m\left(\lambda,A\right)\approx\cosh\left(A\right)$ defined by

$$CH_m\left(\lambda,A\right)=e^{\frac{1}{\lambda^2}}\left(\sum_{n=0}^{m}\frac{\widetilde{H}_{2n+1}\left(\lambda,\frac{1}{2}A^2\right)}{(2n)!\lambda^{2n+1}}\left(1+\frac{2}{(2n+1)\lambda^2}\right)\right), 0<\lambda<+\infty.\tag{12}$$

Working analogously to the proof of the formula (3.6) of [5] one gets, for $x\neq 0$ the following bound:

$$\left\|\widetilde{H}_{2n+1}\left(x,\frac{1}{2}A^2\right)\right\|_2\leq (2n+1)!\frac{e\sinh\left(|x|\,\|A^2\|_2^{1/2}\right)}{|x|\,\|A^2\|_2^{1/2}}.\tag{13}$$

Then we can obtain the following expression for the approximation error:

$$\left\|\cosh\left(A\right)-CH_m\left(\lambda,A\right)\right\|_2 \;\leq\; e^{\frac{1}{\lambda^2}}\sum_{n\geq m+1}\frac{\left\|\widetilde{H}_{2n+1}\left(\lambda,\frac{1}{2}A^2\right)\right\|_2}{(2n)!\lambda^{2n+1}}\left(1+\frac{2}{(2n+1)\lambda^2}\right)\tag{14}$$

$$\leq\; \frac{e^{1+\frac{1}{\lambda^2}}\sinh\left(\lambda\,\|A^2\|_2^{1/2}\right)}{\lambda^2\,\|A^2\|_2^{1/2}}\sum_{n\geq m+1}\frac{2n+1}{\lambda^{2n}}\left(1+\frac{2}{(2n+1)\lambda^2}\right).$$

Taking $\lambda>1$ it follows that $\dfrac{2}{(2n+1)\lambda^2}<1$, and one gets

$$\sum_{n\geq m+1}\frac{2n+1}{\lambda^{2n}}\left(1+\frac{2}{(2n+1)\lambda^2}\right) \;\leq\; 2\sum_{n\geq m+1}\frac{2n+1}{\lambda^{2n}}$$

$$=\; \frac{4+(4m+6)(\lambda^2-1)}{\lambda^{2m}\left(\lambda^2-1\right)^2},$$

m	z_m	λ_m
2	0.0020000000061361199	909.39256098888882
4	0.079956209874370632	99.997970988888895
6	0.34561400005673254	39.999499988888893
9	1.1120032200657	17.997896988889799
12	2.2373014291079998	11.882978988901458
16	4.1086396680000004	7.9999999964157498

Table 1: Values of z_m and λ_m for $\cosh(A)$.

	$m_1 = 2$	$m_2 = 4$	$m_3 = 6$	$m_4 = 9$	$m_5 = 12$	$m_6 = 16$
$\bar{m}_k$	1	2	3	5	7	11
$\tilde{m}_k$	1	2	4	10	13	17
$f_{m_k}(\max)$	0	0	$1.9 \cdot 10^{-17}$	$6.0 \cdot 10^{-19}$	$1.4 \cdot 10^{-26}$	$1.3 \cdot 10^{-35}$

Table 2: Values $\bar{m}_k$, $\tilde{m}_k$, and $f_{\max}$.

thus from (14) we finally obtain:

$$\|\cosh(A) - CH_m(\lambda, A)\|_2 \leq \frac{e^{1+\frac{1}{\lambda^2}} \sinh\left(\lambda \|A^2\|_2^{1/2}\right)\left(4 + (4m+6)(\lambda^2 - 1)\right)}{\|A^2\|_2^{1/2} \lambda^{2m+2} (\lambda^2 - 1)^2}.$$

(15)

From this expression (15) we derived the optimal values $(\lambda_m; z_m)$ such that

$$z_m = \max\left\{z = \|A^2\|_2 \, ; \, \frac{e^{1+\frac{1}{\lambda^2}} \sinh\left(\lambda z^{1/2}\right)\left(4 + (4m+6)(\lambda^2 - 1)\right)}{z^{1/2}\lambda^{2m+2} (\lambda^2 - 1)^2} < u\right\}$$

where u is the unit roundoff in IEEE double precision arithmetic, $u = 2^{-53}$. The optimal values of m, z and λ have been obtained with MATLAB. The results are given in the Table 1.

If $\cosh(A)$ is calculated from the Taylor series, then the absolute forward error of the Hermite approximation of $\cosh(A)$, denoted by E_f, can be computed as

$$E_f = \|\cosh(A) - P_{m_k}(B)\| = \left\|\sum_{i \geqslant \bar{m}_k} f_{m_k, i} B^i\right\| \cong \left\|\sum_{i \geqslant \tilde{m}_k} f_{m_k, i} B^i\right\|,$$

where the values of $\bar{m}_k$ and $\tilde{m}_k$ for each $m_k \in \{2, 4, 6, 9, 12, 16\}$ appear in the Table 2.

Scaling factor s and the order of Hermite approximation m_k are obtained by the following:

Theorem 3.1 ([6]) *Let* $h_l(x) = \sum\limits_{i \geq l} p_i x^i$ *be a power series with radius of convergence* w, $\tilde{h}_l(x) = \sum\limits_{i \geq l} |p_i| x^i$, $B \in \mathbb{C}^{n \times n}$ *with* $\rho(B) < w$, $l \in \mathbb{N}$ *and* $t \in \mathbb{N}$ *with* $1 \leqslant t \leqslant l$. *If* t_0 *is the multiple of* t *such that* $l \leqslant t_0 \leqslant l + t - 1$ *and*

$$\beta_t = \max\{d_j^{1/j} : j = t, l, l+1, \ldots, t_0 - 1, t_0 + 1, t_0 + 2, \ldots, l + t - 1\},$$

where d_j *is an upper bound for* $||B^j||$, $d_j \geqslant ||B^j||$, *then*

$$||h_l(B)|| \leqslant \tilde{h}_l(\beta_t).$$

We have empirically verified that by neglecting the coefficients whose absolute value is lower than u, the efficiency results are far superior to the state-of-the-art algorithms, with also excellent accuracy.

4   Numerical experiments

The MATLAB's implementation `coshmtayher` is a modification of the MATLAB's code `coshher` given in [5], replacing the original Hermite approximation `coshher` by the new Hermite matrix polynomial obtained from (11). In this section, we compare the new MATLAB function developed in this paper, `coshmtayher`, with the functions `coshher` and `funmcosh`:

- `coshmtayher`. New code based on the new developments of Hermites matrix polynomials (11).

- `coshher`. Code based on the Hermite series for the hyperbolic matrix cosine [5].

- `funmcosh`. MATLAB function *funm* for compute matrix functions, i. e. the hyperbolic matrix cosine.

The tests have been develop using MATLAB (R2017b), runing on an Apple Macintosh iMac 27" (iMac retina 5K 27" late 2015) with a quadcore INTEL i7-6700K 4 Ghz processor and 16 Gb of RAM.
The following sets of matrices have been used:

a) One hundred diagonalizable matrices of size 128×128. Table 3 show the percentage of cases in which the relative errors of `coshmtayher` (new Hermite code) are lower, greater or equal than the relative errors of `coshher`(Hermite code) and `funmcosh` (funm code). Table 4 shows the matrix products of each method. Graphics with the Normwise relative errors, see [7, p. 253] and Performance Profile, see [7, p. 254], are given in Figure 1.

b) One hundred non diagonalizables matrices of size 128×128 with multiple eigenvalues randomly generated. Table 5 shows the percentage of cases in

which the relative errors of `coshmtayher` are lower, greater or equal than the relative errors of `coshher` and `funmcosh`. Table 6 shows the matrix products of each method. Graphics of the Normwise relative errors and the Performance Profile are given in Figure 2.

c) Ten matrices from the Eigtool MATLAB [8] package with size 128×128, and thirty matrices from the function `matrix` of the Matrix Computation Toolbox [9] with dimensions lower or equal than 128. These matrices have been chosen because they have more varied and significant characteristics. Table 7 shows the percentage of cases in which the relative errors of `coshmtayher` are lower, greater or equal than the relative errors of `coshher` and `funmcosh`. Table 8 shows the matrix products of each method. Graphics of the Normwise relative errors and the Performance Profile are given Figure 3.

$E(coshmtayher) < E(coshher)$	47.50%
$E(coshmtayher) > E(coshher)$	50.00%
$E(coshmtayher) = E(coshher)$	3.00%
$E(coshmtayher) < E(funmcosh)$	100.00%
$E(coshmtayher) > E(funmcosh)$	0.00%
$E(coshmtayher) = E(funmcosh)$	0.00%

Table 3: Comparative between the methods

$cosmtayher$	$coshher$	$funmcosh$
671	973	1500

Table 4: Matrix products

$E(coshmtayher) < E(coshher)$	52.50%
$E(coshmtayher) > E(coshher)$	47.00%
$E(coshmtayher) = E(coshher)$	1.00%
$E(coshmtayher) < E(funmcosh)$	100.00%
$E(coshmtayher) > E(funmcosh)$	0.00%
$E(coshmtayher) = E(funmcosh)$	0.00%

Table 5: Comparative between the methods

(a) Normwise relative errors   (b) Performance Profile
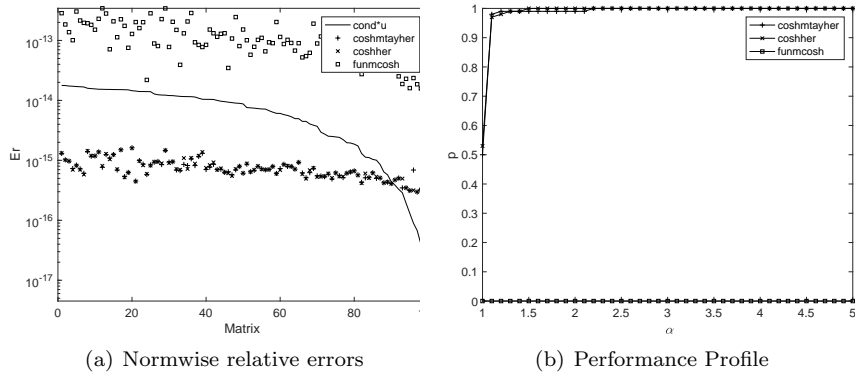
Figure 1: Diagonalizable matrices

$cosmtayher$	$coshher$	$funmcosh$
685	989	1500

Table 6: Matrix products

(a) Normwise relative errors   (b) Performance Profile
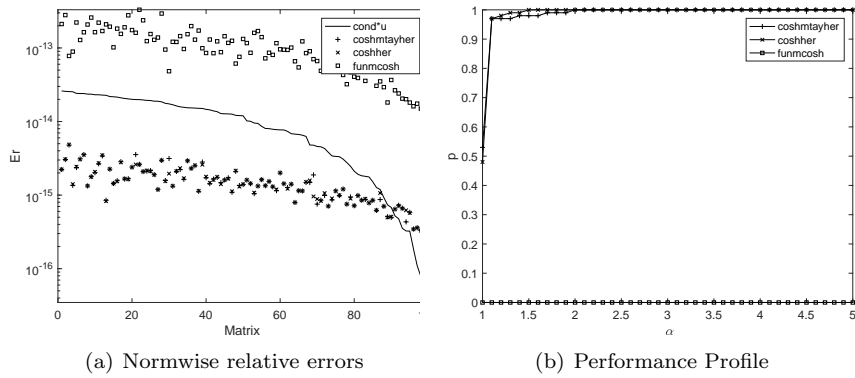
Figure 2: Non diagonalizable matrices

$E(coshmtayher) < E(coshher)$	57.50%
$E(coshmtayher) > E(coshher)$	30.00%
$E(coshmtayher) = E(coshher)$	12.50%
$E(coshmtayher) < E(funmcosh)$	97.50%
$E(coshmtayher) > E(funmcosh)$	2.50%
$E(coshmtayher) = E(funmcosh)$	0.00%

Table 7: Comparative between the methods

cosmtayher	*coshher*	*funmcosh*
191	315	600

Table 8: Matrix products

(a) Normwise relative errors (b) Performance Profile
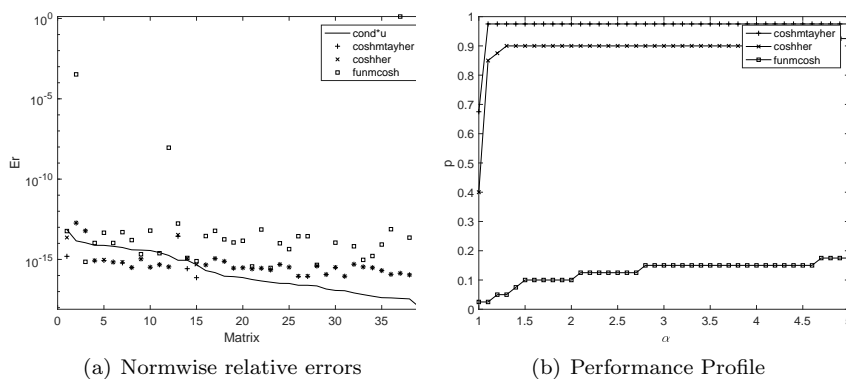
Figure 3: Matrices from the Eigtool and the Matrix Computation Toolbox packages

5 Conclusions

The more accurate are the implementations based on the Hermite series: the initial MATLAB implementation (`coshher`) and the proposed MATLAB implementation based on (11) (`coshmtayher`). Also, the new implementation (`coshmtayher`) have considerably lower computational costs than the other functions.

References

[1] J. Jódar, R. Company, Hermite matrix polynomials and second order matrix differential equations, Approximation Theory and its Applications 12 (2) (1996) 20–30.

[2] J. Sastre, J. Ibáñez, E. Defez, P. Ruiz, New scaling-squaring Taylor algorithms for computing the matrix exponential, SIAM Journal on Scientific Computing 37 (1) (2015) A439–A455.

[3] P. Alonso, J. Peinado, J. Ibáñez, J. Sastre, E. Defez, Computing matrix trigonometric functions with gpus through matlab, The Journal of Supercomputing (2018) 1–14.

[4] J. Sastre, Efficient evaluation of matrix polynomials, Linear Algebra and its Applications 539 (2018) 229–250.

[5] E. Defez, J. Sastre, J. Ibáñez, J. Peinado, Solving engineering models using hyperbolic matrix functions, Applied Mathematical Modelling 40 (4) (2016) 2837–2844.

[6] J. Sastre, J. Ibáñez, P. Ruiz, E. Defez, Efficient computation of the matrix cosine, Applied Mathematics and Computation 219 (14) (2013) 7575–7585.

[7] N. J. Higham, Functions of Matrices: Theory and Computation, SIAM, Philadelphia, PA, USA, 2008.

[8] T. Wright, Eigtool, version 2.1, URL: web. comlab. ox. ac. uk/pseudospectra/eigtool.

[9] N. J. Higham, The test matrix toolbox for MATLAB (Version 3.0), University of Manchester Manchester, 1995.

A novel optimization technique for railway wheel rolling noise reduction

J. Gutiérrez-Gil[♭][*], X. Garcia-Andrés[♭], J. Martínez-Casas[♭],
E. Nadal[♭], F. D. Denia[♭]

(♭)Centro de Investigación en Ingeniería Mecánica, Universitat Politècnica de València,

Camino de Vera s/n, 46022 Valencia, España,

November 30, 2018

1   Introduction

In this work, a novel methodology towards the minimization of rolling noise through changes of the Railway Wheel (RW) cross-sectional geometry is presented. The approach is based on shifting the Natural Frequencies (NF) of the acoustically-relevant modes out of the excitation range or at frequencies where it has a lower content. The presented procedure permits a deep exploration of the solution space with a reduced computational expense.

It is assumed that the maximization of the NF corresponding to the most relevant vibration modes leads to a reduction of the radiated sound power, given the lower frequency content of the excitation force in the high frequency domain [1]. Hence, the range of frequencies in which the optimization is performed is chosen according to the spectrum of the combined excitation roughness defined in the standard EN 13979-1:2011 [2].

This approach has the advantage of a reduced computation cost compared to other similar approaches in the field [3],[4]. The common procedures for

---

[*]e-mail: jorgugi1@upv.es

the calculation of acoustic power radiation are very expensive computation-
ally (Boundary Element Method - BEM) in an iterative procedure or require
simplified and/or commercial models (TWINS software). Instead, the im-
plementation developed in this work, which avoids the need of deriving noise
radiation, intends to reduce greatly the time required to perform an iteration
while achieving an adequate FEM mesh accuracy in the process.

A Genetic Algorithm (GA) is used as optimization technique, whereas
the Objective Function (OF) has been defined as an expression depending
on the NF of a set of selected modeshapes which are more likely to contribute
to sound radiation.

2 Methodology
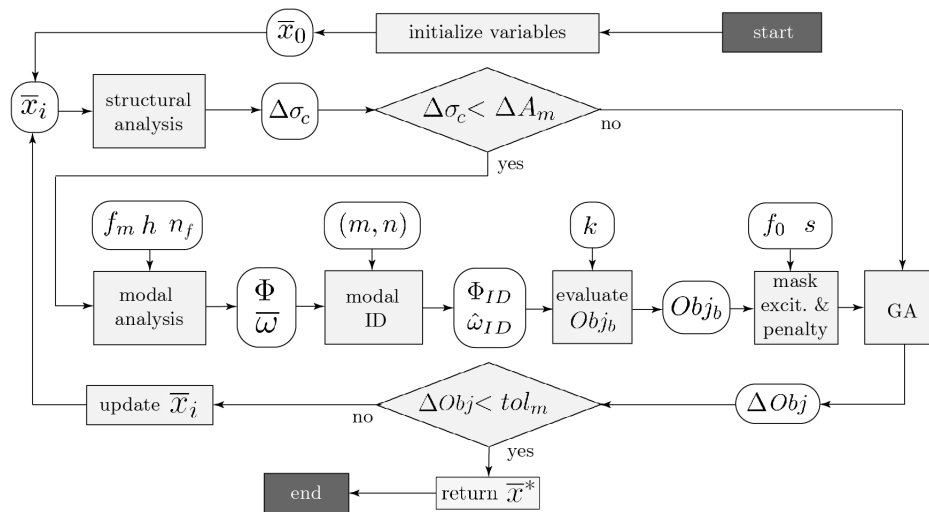
Figure 1: Main components of the optimization procedure.

The optimization algorithm implements a data flow as illustrated in Fig.
1. At the start of the procedure, an initial design is proposed and introduced
in the main search loop. At each iteration, first a high-cycle fatigue analysis is
carried out according to standard [2] using a Finite Element Method (FEM)
model, and only structurally acceptable candidates whose critical stress dif-
ference $\Delta\sigma_c$ is lower than the admissible ΔA_m, are further considered. In
such case, the modeshapes Φ and undamped NF $\overline{\omega}$ of design candidate i
are obtained through a modal analysis. Such analysis is configured by the

maximum frequency $f_m = 5000$ Hz up to which modes are obtained, the maximum element size $h = 0.015$ m of the FE mesh, and the number of Fourier terms $n_f = 12$ used in the axisymmetric computation.

The modeshapes that have a greater sound contribution are selected and identified through their nodal diameters n and nodal circumferences m [1]. The average NF of those selected modes $\hat{\omega}_{m,ID}$ is then used to generate the basis OF value Obj_b through Eq. 1. This value is then filtered by two modifying masks or transfer functions, resulting in the effective penalty value Obj which is to be minimized,

$$Obj = Obj_b \cdot m_e \cdot m_p =$$
$$\left[\frac{1}{(\hat{\omega}_{m,ID})^k}\right] \cdot \left[\frac{-\tanh\left(\frac{\hat{\omega}_{m,ID}-f_0}{s}+1\right)}{2}\right] \cdot \left[1+|\frac{\hat{\omega}_{p,ID}-\hat{\omega}_{p,ID}^0}{\hat{\omega}_{p,ID}}|\right]. \quad (1)$$

The so-called excitation mask term m_e considers the frequency content of the combined excitation roughness spectrum, derived from the standard [2] and is implemented in the form of a hyperbolic tangent function, controlled by parameter s. It acts as a filter favouring the shift of the modal frequencies towards the region considered as having less significant amplitude content, determined by f_0. The penalization mask term m_p considers the possible negative effect of shifting certain modes into the range of important frequency content of the excitation, in case this spectrum is not monotonically decreasing. Such a penalization is computed through the difference on the average natural frequency value of the identified modes to penalise $\hat{\omega}_{p,ID}$ with respect to their initial value, $\hat{\omega}_{p,ID}^0$. The processed candidate OF value Obj is transmitted to the GA, which will derive the design variables for the next iteration and compute the OF value variation between previous candidates ΔObj. If such difference is lower than the specified tolerance tol_m, the best solution found $\overline{x}^*$ is then returned. Otherwise, the update variables are introduced again in the beginning of the main loop.

In order to obtain a geometry to generate a FE model used within the optimization algorithm, a geometric parametrization of the transversal section of the wheel is defined according to its relevance in the noise radiation problem [4],[5]. The details are shown in Figure 2.

While x_1 in Fig. 2 is directly related to its corresponding design variable, x_2, x_3 and x_4 are defined in terms of a scalar multiplier to a base value.
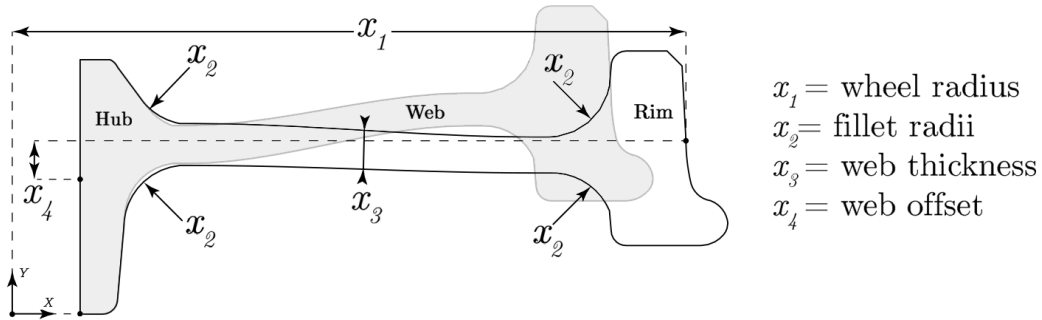
Figure 2: Geometric design variables and an explanatory design (shaded) with reduced x_1, x_2, x_3 and x_4 parameters. Note that x_4 is the axial distance between the middle point of the hub surface and the contact point.

Regarding the discretization of the geometry stated above, general axisymmetric elements have been used for the wheel FEM model presented in this work. These consider Fourier series in their shape functions with the purpose of describing the change of the displacement field in the circumferential direction θ. For the calculation of the full response, n_f nodal planes are generated from the transverse sections, which act as the master plane. The displacements are interpolated taking into account the Fourier series expansion as:

$$u = \sum_{i=1} N_i \, u_i \cdot \left(c_1 + \sum_{n_f} \left(a_{n_f} \cos(n_f \, \theta) + b_{n_f} \sin(n_f \, \theta) \right) \right), \qquad (2)$$

where c_1, a_n and b_n are the Fourier constants, u_i the displacements in each node and N_i the 2D shape functions of linear quadrilateral elements.

3   Results

Fig. 3 compares the obtained NF distributions and cross-sectional geometries of the initial and Best Found Solution (BFS) using the methodology proposed. It is seen that those modeshapes selected for NF maximization, with $n = 2, 3, 4$ and $m = 1$, have been shifted to a higher frequency region where the excitation is significantly lower. As expected, the optimized geometry presents features associated by literature [1],[6] with reduced wheel

rolling noise, namely smaller wheel radius (x_1), larger transition radii (x_2) and a thicker web (x_3) with a straight shape (x_4), showing that low noise-radiation wheel designs can be obtained with a NF maximization approach.
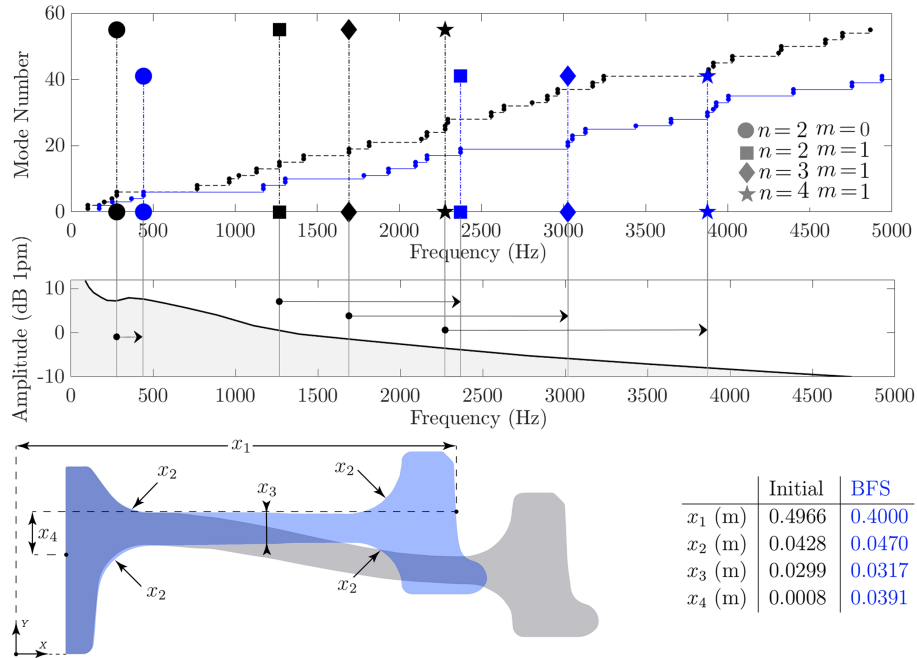
Figure 3: Top: NF distribution of the wheel before (dashed line) and of the BFS (solid blue line). Middle: Roughness spectrum under which the RW is excited as a function of frequency from [2]. Bottom: Cross-sectional designs of the initial wheel (shaded), and the BFS (blue).

Considering only those modeshapes which are most relevant for the generation of rolling noise in Eq. 1, provides slightly better results than using an OF without penalization and excitation masks, which simply consists in the average value of the NF of *all* the set of modeshapes. Improvements in convergence times have been observed with the application of the excitation mask m_e term using $s = 500$ and $f_0 = 1388$ Hz. The use of a penalization mask m_p term has no considerable shift restriction effect on the modeshape targeted for this, with $n = 2, 3, 4; m = 0$. In Fig. 3 it can be seen how the $n = 2; m = 0$ modeshape has been undesirably shifted to a local maximum of the excitation spectrum with increased amplitude, around 400 Hz. The results suggest an important degree of coupling between modeshapes in a RW.

4   Conclusions

A RW geometric optimization maximizing the NF of those modeshapes which contribute more to sound radiation is performed. It is seen that the use of a modeshape identification based OF slightly improves performance, and the so-called excitation mask term produces a faster convergence. Generally, results reflect that a significant maximization of the targeted modeshapes can be achieved using the methodology proposed, leading to wheel designs with characteristics associated in literature with lower sound radiation.

5   Acknowledgments

References

[1] D.J. Thompson, "Railway Noise and Vibration Mechanisms, Modelling and Means of Control", 1st ed. Elsevier, 2009.

[2] UNE-EN 13979-1:2006+A2:2011, "Railway applications - Wheelsets and bogies", 2011.

[3] G.A. Efthimeros et al., "Vibration/noise optimization of a GEM RW model", Engineering Computations, vol. 19, pp. 922-931, 2002.

[4] J.C.O. Nielsen et al., "Multi-disciplinary optimization of RWs", Journal of Sound and Vibration, vol. 293, pp. 510-521, 2006.

[5] D.J. Thompson, "But are the trains getting any quieter?", ICSV14, 2007.

[6] G. Hölz. "A quiet railway by noise optimized wheels (in German)", ZEV+DET Glas. Ann., 188(1), 20-23,1994.

Improving the order of convergence of Traub-type derivative-free methods*

F. I. Chicharro, A. Cordero, N. Garrido,† and J. R. Torregrosa

Institute for Multidisciplinary Mathematics, Universitat Politècnica de València,

Camino de Vera s/n, 46022 València (Spain).

November 30, 2018

1   Introduction

The use of iterative methods for solving nonlinear equations $f(x) = 0$ goes back to several centuries, as stated in [1]. The digital revolution of last decades has given a boost to the development of iterative methods, as can be observed in the amount of publications regarding this issue.

These iterative methods can be classified depending on several aspects. On the one hand, the absence of derivatives results in derivative-free methods. On the other hand, the inclusion of previous iterates to obtain the current one gives rise to methods with memory. The former schemes improve the computational effort and the amount of problems that can be solved, since the obtention of derivatives is avoided. The latter ones increases the order of convergence of the methods without adding new functional evaluations at the expense of computational efficiency [2].

From the third-order Traub's iterative method [3],

$$
\begin{aligned}
y_k &= x_k - \frac{f(x_k)}{f'(x_k)}, \\
x_{k+1} &= y_k - \frac{f(y_k)}{f'(x_k)},
\end{aligned}
\tag{1}
$$

that includes derivatives and has not memory, two derivative-free methods with memory are introduced.

2   Traub-type methods with memory

Taking (1) as a reference, the replacement of the derivatives by divided differences holds the order of convergence in 3. For memory purposes, an accelerating parameter is also included. This gives rise to the iterative expression of D1, whose expression is

$$
\begin{aligned}
w_k &= x_k + \rho f(x_k), \\
y_k &= x_k - \frac{f(x_k)}{f[x_k, w_k]}, \\
x_{k+1} &= y_k - \frac{f(y_k)}{f[x_k, w_k]},
\end{aligned}
\tag{2}
$$

and its error equation is

$$
e_{k+1} = (1 + f'(\alpha)\rho)\,(2 + f'(\alpha)\rho)\,c_2^2 e_k^3 + \mathcal{O}(e_k^4),
\tag{3}
$$

where α is the root of $f(x)$, $e_k = x_k - \alpha$ and $c_j = \frac{f^{(j)}(\alpha)}{j!f'(\alpha)}, j \geq 2$. It is clear that, for $\rho = -\frac{1}{f'(\alpha)}$ or $\rho = -\frac{2}{f'(\alpha)}$ the method has, at least, order of convergence 4. The unknown value of α demands the obtention of an approximation of $f'(\alpha)$ and, therefore, the resulting method will not reach fourth order of convergence.

Below, two approximations are introduced. On the one hand, by applying a linear approximation, $f'(\alpha)$ can be obtained as a divided difference $f[x_k, x_{k-1}]$, where x_{k-1} includes memory. On the other hand, a Newton's interpolation polynomial of second order $N_2(t) = f(x_k) + f[x_k, x_{k-1}](t - x_k) + f[x_k, x_{k-1}, y_{k-1}](t - x_k)(t - x_{k-1})$, whose derivative is obtained in the point x_k, can be applied as a substitute of $f'(\alpha)$. The first approximation results in an iterative method with memory with order of convergence of 3.30, while the second one gives rise to a value of 3.73. The dynamical analysis will be performed over the second method, denoted as DM1 from now on. Its final iterative expression is

$$
\begin{aligned}
w_k &= x_k - \frac{f(x_k)}{f[x_k, x_{k-1}] + f[x_k, x_{k-1}, y_{k-1}](x_k - x_{k-1})}, \\
y_k &= x_k - \frac{f(x_k)}{f[x_k, w_k]}, \\
x_{k+1} &= y_k - \frac{f(y_k)}{f[x_k, w_k]}.
\end{aligned}
\tag{4}
$$

In a similar way to proceed, we include in (2) a new step as follows:

$$
\begin{aligned}
w_k &= x_k + \rho f(x_k), \\
q_k &= x_k + \theta f(x_k), \\
y_k &= x_k - \frac{f(x_k)}{f[x_k, w_k]}, \\
x_{k+1} &= y_k - \frac{f(y_k)}{f[x_k, q_k]}.
\end{aligned}
\tag{5}
$$

It can be proved that (5) has order of convergence 3 with independence of parameters, and its error equation is

$$
e_{k+1} = (1 + \rho f'(\alpha))(2 + \theta f'(\alpha))c_2^2 e_k^3 + \mathcal{O}(e_k^4).
\tag{6}
$$

Needless to say that the replacement $\rho = -\frac{1}{f'(\alpha)}$ or $\theta = -\frac{2}{f'(\alpha)}$ makes (5), at least, fourth-order convergent. Since the value of $f'(\alpha)$ is unknown, the estimation of its value is mandatory. For a lineal approximation such as $\rho = \frac{\theta}{2} = -\frac{1}{f[x_k, x_{k-1}]}$ the order of convergence gets the value 3.56. If a Newton's polynomial approximation is applied,

$$
\rho = \frac{\theta}{2} = -\frac{1}{N_2'(x_k)},
\tag{7}
$$

the order of convergence reaches the value 4.23. The method DM2, wherein Newton's approximation is applied, has the iterative expression

$$
\begin{aligned}
w_k &= x_k - \frac{f(x_k)}{f[x_k, x_{k-1}] + f[x_k, x_{k-1}, y_{k-1}](x_k - x_{k-1})}, \\
q_k &= x_k - \frac{2f(x_k)}{f[x_k, x_{k-1}] + f[x_k, x_{k-1}, y_{k-1}](x_k - x_{k-1})}, \\
y_k &= x_k - \frac{f(x_k)}{f[x_k, w_k]}, \\
x_{k+1} &= y_k - \frac{f(y_k)}{f[x_k, q_k]}.
\end{aligned}
\tag{8}
$$

3   Dynamical analysis

As DM1 is a method with memory which uses Newton's interpolation polynomial through the points x_k, x_{k-1} and y_{k-1}, its fixed point function $\Phi_1 : \mathbb{R}^3 \to \mathbb{R}^3$ has the expression

$$
\Phi_1(x_{k-1}, y_{k-1}, x_k) = (x_k, y_k, x_{k+1}) = (x_k, y_k, \phi_1(x_{k-1}, y_{k-1}, x_k)), \quad k \geq 1,
\tag{9}
$$

being x_0, y_0 and x_1 three initial estimations of the method. When Φ_1 is applied on a quadratic polynomial $p_c(x) = x^2 + c$, $c \in \mathbb{R}$, the result is the fixed point operator:

$$\Phi_1(z, u, x) = \left(x, y, \frac{5c^3x - 13c^2x^3 + 39cx^5 - 7x^7}{(c - 3x^2)^3} \right), \qquad k \geq 1, \qquad (10)$$

where z, u, y and x denote x_{k-1}, y_{k-1}, y_k and x_k, respectively.

Fixed points of (9) must satisfy conditions $z = u = x$ and $x = \phi_1(z, u, x)$, and then real dynamics of Φ_1 becomes in the dynamical study of a one-dimensional operator:

$$\Phi_1(z, u, x)|_{z=u=x} = \tilde{\Phi}_1(x) = \frac{5c^3x - 13c^2x^3 + 39cx^5 - 7x^7}{(c - 3x^2)^3}.$$

Let us recall that the calculation of the fixed points of Φ_1 is equivalent to the calculation of those of $\tilde{\Phi}_1$. Real fixed points of the operator are $x_{1,2}^F(c) = \mp i\sqrt{c}$ for $c < 0$, and also the strange fixed point $x_3^F = 0$. Evaluating the fixed points in $|\tilde{\Phi}_1'(x)|$ it is proven that $x_{1,2}^F(c)$ are superattracting and x_3^F is a repelling point.

The critical points of $\tilde{\Phi}_1(x)$, which are the solutions of the equation $\tilde{\Phi}_1'(x) = 0$, are the roots of $p_c(x)$ and the free critical points $x_{1,2}^C(c) = \mp i\sqrt{\frac{5c}{21}}$.

The dynamical approach of DM2 method follows a similar structure as in method DM1. The real multidimensional fixed point function associated to DM2 is

$$\Phi_2(z, u, x) = (x, y, \phi_2(z, u, x)), \qquad k \geq 1.$$

By applying method DM2 on the polynomial $p_c(x)$, the fixed point operator is of the form

$$\Phi_2(z, u, x) = \left(x, y, -\frac{2x\left(-2c^3 + 5c^2x^2 - 8cx^4 + x^6\right)}{(c - 3x^2)^2 (c - x^2)} \right), \qquad k \geq 1. \qquad (11)$$

In order to obtain the fixed points of Φ_2, the one-dimensional operator obtained imposing conditions $z = u = x$ to (11) is

$$\Phi_2(z, u, x)|_{z=u=x} = \tilde{\Phi}_2(x) = -\frac{2x\left(-2c^3 + 5c^2x^2 - 8cx^4 + x^6\right)}{(c - 3x^2)^2 (c - x^2)}.$$

By solving the equation $\tilde{\Phi}_2(x) = x$, we get that method DM2 has the same fixed points as method DM1. In addition, $x_{1,2}^F(c)$ are superattracting points and x_3^F is a repelling point. The free critical points of method DM2 are $x_{1,2}^C(x) = \mp\sqrt{\frac{2c}{3}}$.

A common use in the complex analysis is the representation of the basins of attraction [4, 5]. When the methods include memory, a real analysis is performed instead [6, 7]. The key point of this representation is the mapping of a root of the polynomial with a color. In this sense, if an initial guess x_0 tends to a root, then x_0 is painted with the root corresponding color; otherwise, the initial guess is represented in black.

Figure 1 shows two particular cases of the dynamical lines for both DM1 and DM2 methods. For $c < 0$ cases, two basins of attraction - one per attracting fixed point - can be found. For $c > 0$, the polynomial has complex roots.

(a) DM1   (b) DM1   (c) DM2   (d) DM2
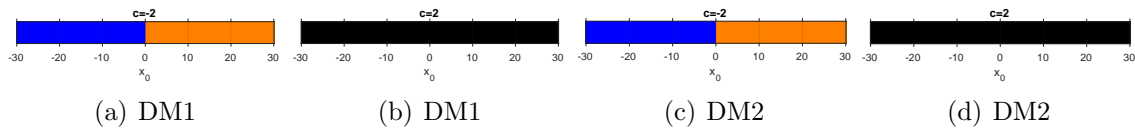
Figure 1: Dynamical lines of methods on $p_c(x) = x^2 + c$.

The convergence plane [6] summarizes in one figure the behavior of every value of c. In this way, Figure 2 represents the convergence planes of DM1 and DM2. The critical points are represented in a black line, while the white line is the strange fixed point.

(a) DM1   (b) DM2
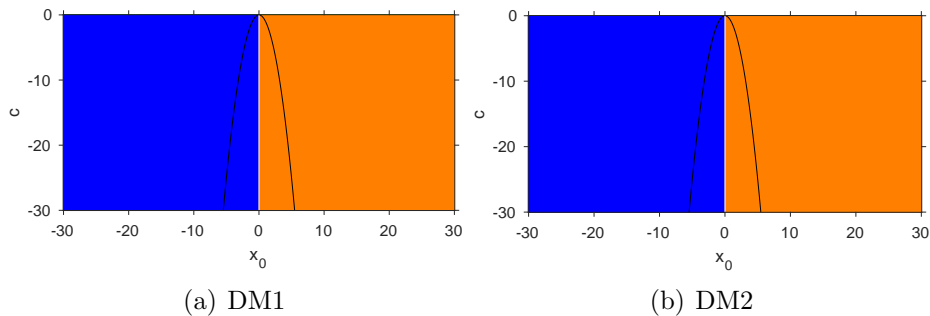
Figure 2: Convergence planes of the methods on $p_c(x) = x^2 + c$.

4   Conclusions

Two iterative methods have been introduced. On the one hand, they include memory for increasing the order of convergence. On the other hand, both are derivative-free. In terms of stability, both convergence planes of Figure 2 show wide basins of attraction, where the stability for every initial guess is guaranteed.

References

[1] S. Plaza and J. M. Gutiérrez. Dinámica del método de Newton. Servicio de Publicaciones Universidad de La Rioja, 2013.

[2] M. S. Petković, B. Neta, L. D. Petković, and J. Džunić. Multipoint methods for solving nonlinear equations. Elsevier, 2013.

[3] J. F. Traub. Iterative methods for the solution of equations. Prentice-Hall, 1964.

[4] F. I. Chicharro, A. Cordero, and J. R. Torregrosa. Drawing dynamical and parameters planes of iterative families and methods. *The Scientific World Journal*, ID 780153:1–11, 2013.

[5] J. L. Varona. Graphic and numerical comparison between iterative methods. *Mathematical Intelligencer*, 24:37–46, 2002.

[6] Á. A. Magreñán. A new tool to study real dynamics: the convergence plane. *Applied Mathematics and Computation*, 248:215–224, 2014.

[7] F. I. Chicharro, A. Cordero, J. R. Torregrosa, and M. P. Vassileva. King-type derivative-free iterative families: Real and memory dynamics. *Complexity*, ID 2713145:1–15, 2017.

Efficient decoupling technique applied to the numerical time integration of advanced interaction models for railway dynamics

J. Giner-Navarro[1], J. Martínez-Casas[1], F. D. Denia[1], L. Baeza[2]

[1] Centro de Investigación en Ingeniería Mecánica, Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain. E-mail: juanginer@upv.es

[2] Institute of Sound and Vibration Research, University of Southampton - Highfield, B13/R3075, Southampton, United Kingdom

ABSTRACT

1. Introduction

Railway interaction is characterised by the coupling between the train and the track introduced through the forces appearing in the wheel/rail contact area. This work presents models for the wheelset and the rail, both in contact, which adopt an Eulerian-modal approach, leading to linear differential equations in modal coordinates that drastically reduce the number of state variables of the dynamic system and thus the associated computational cost. Since these equations of motion obtained in the formulation are still coupled, this paper develops a formulation that decouples them and allows solving each one independently for each time step. The decoupling integration method proposed is compared in terms of computational performance with two time integration schemes commonly used in vehicle dynamics: Newmark algorithm and Matlab's ode45.

2. Wheelset/track interaction model

Both wheelset and track substructures are coupled in the contact area between the wheel and the rail through the contact force. In this work, the contact force is applied to a point corresponding to the contact node in the wheel and the rail. The normal contact force is uncoupled from the tangential one if wheel and rail are made of the same material [1] and is obtained from the Hertzian formulation [2]; the tangential one is computed from the normal one through CONTACT algorithm [3]. Homogenous equations of motion for the coupled wheelset/track system are assembled by considering the equations for the rotating wheelset [4] and for both inner and outer rails supported by a uniform viscoelastic Winkler bedding [5]:

$$
\begin{pmatrix} \ddot{\mathbf{q}}^w \\ \ddot{\mathbf{q}}^r_{inn} \\ \ddot{\mathbf{q}}^r_{out} \end{pmatrix} + \begin{pmatrix} 2\left(\Omega\,\tilde{\mathbf{V}}^w + \tilde{\mathbf{P}}^w\right) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -2V\tilde{\mathbf{C}}^r + \tilde{\mathbf{C}}^r_{wink} + \tilde{\mathbf{C}}^r_{\zeta} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -2V\tilde{\mathbf{C}}^r + \tilde{\mathbf{C}}^r_{wink} + \tilde{\mathbf{C}}^r_{\zeta} \end{pmatrix} \begin{pmatrix} \dot{\mathbf{q}}^w \\ \dot{\mathbf{q}}^r_{inn} \\ \dot{\mathbf{q}}^r_{out} \end{pmatrix}
$$

$$
+ \begin{pmatrix} \Omega^2\left(\tilde{\mathbf{A}}^w - \tilde{\mathbf{C}}^w\right) + 2\,\Omega\,\tilde{\mathbf{S}}^w + \tilde{\mathbf{R}}^w - \tilde{\mathbf{B}}^w + \tilde{\mathbf{D}}^w & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{K}}^r - V^2\tilde{\mathbf{A}}^r + \tilde{\mathbf{K}}^r_{wink} + V\,\mathbf{K}\tilde{\mathbf{C}}^r_{wink} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \tilde{\mathbf{K}}^r - V^2\tilde{\mathbf{A}}^r + \tilde{\mathbf{K}}^r_{wink} + V\,\mathbf{K}\tilde{\mathbf{C}}^r_{wink} \end{pmatrix} \begin{pmatrix} \mathbf{q}^w \\ \mathbf{q}^r_{inn} \\ \mathbf{q}^r_{out} \end{pmatrix} = \tag{1}
$$

$$
\begin{pmatrix} \ddot{\mathbf{q}}^w \\ \ddot{\mathbf{q}}^r_{inn} \\ \ddot{\mathbf{q}}^r_{out} \end{pmatrix} + \begin{pmatrix} \tilde{\mathbf{C}}^w_{eq} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{C}}^r_{eq} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \tilde{\mathbf{C}}^r_{eq} \end{pmatrix} \begin{pmatrix} \dot{\mathbf{q}}^w \\ \dot{\mathbf{q}}^r_{inn} \\ \dot{\mathbf{q}}^r_{out} \end{pmatrix} + \begin{pmatrix} \tilde{\mathbf{K}}^w_{eq} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{K}}^r_{eq} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \tilde{\mathbf{K}}^r_{eq} \end{pmatrix} \begin{pmatrix} \mathbf{q}^w \\ \mathbf{q}^r_{inn} \\ \mathbf{q}^r_{out} \end{pmatrix}
$$

$$
= \ddot{\mathbf{q}}^{wr} + \tilde{\mathbf{C}}^{wr}\dot{\mathbf{q}}^{wr} + \tilde{\mathbf{K}}^{wr}\mathbf{q}^{wr} = \tilde{\mathbf{Q}}^{wr}_c.
$$

Eq. (1) is a linear differential system coupled by the generalised contact forces $\tilde{\mathbf{Q}}^{wr}_c$, in which its matrices are in general not diagonal. This work proposes a strategy to uncouple the previous system through an efficient modal approach based on two variable transformations applied during pre-processing. The resulting formulation consists of $2m$ independent equations that can be expressed as:

$$
\left. \begin{aligned} \dot{s}_\iota + \lambda_\iota s_\iota &= \tilde{G}_\iota \\ \dot{s}_\iota + \lambda^*_\iota s_\iota &= \tilde{G}^*_\iota \end{aligned} \right\} , \qquad \iota = 1,...m, \tag{20}
$$

where m is the number of modal coordinates considered for the wheelset/track system.

3. Results

The computational performance of each integration scheme used is addressed through a parameter study that evaluates the time consumption and error of the computed physical response dependent of the number modal coordinates used for the modal approach. The error is evaluated in the node in which the contact force is applied; the reference solution will be the norm of its displacement solution along the simulation. Since the largest deformation field will be registered at this point, the registered error will be more appropriate for the numerical evaluation.

As first approximation, the 3D MEM rail model is simplified to 1D one based on Koh *et al.* formulation [6] uniformly meshed along the longitudinal direction using 500 bar elements. The contact force has been previously calculated from a simulation in curved and randomly corrugated rails, assuming a corrugation spectrum corresponding to the ISO 3095 limit [7], which establishes a third-octave band spectrum of the rail roughness. The contact node is located in the middle of the beam. The vertical displacement of the contact node along the time simulation is computed and plotted in Fig. 1 using the three

integration schemes mentioned. It indicates that the fixed time step of 5×10^{-6} s for the decoupling and Newmark schemes has been properly selected to match the reference solution. As observed, the decoupling method fits better the ode45 reference solution than the Newmark algorithm, a first indicator of the computational advantages of the proposed method.
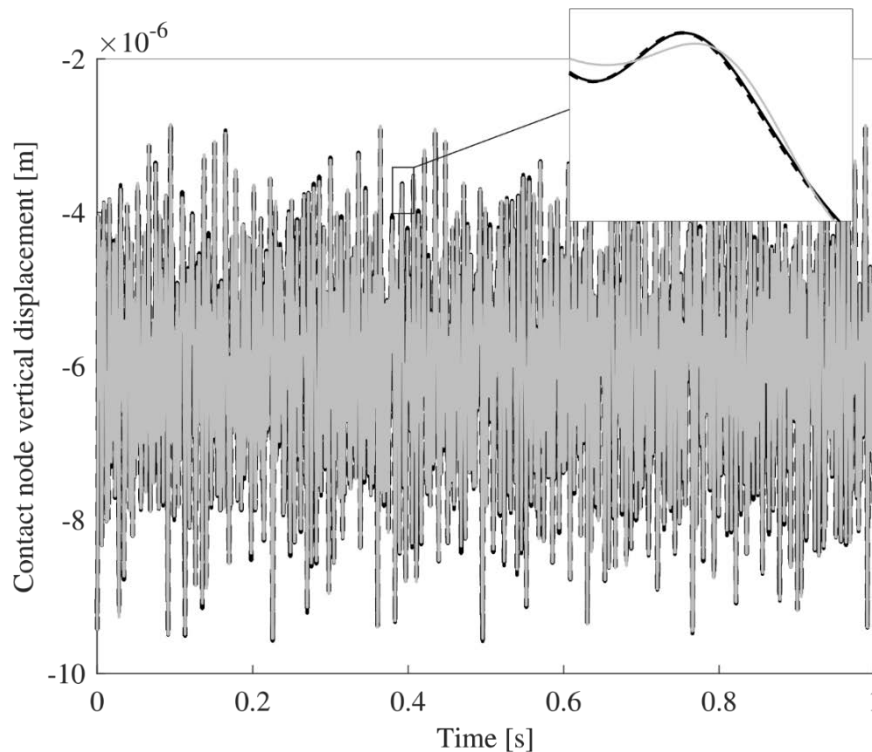
Figure 1: Time series for the vertical displacement of the contact node using three schemes of integration: decoupling technique (━━━), Newmark (━━━) and ode45 (━ ━ ━).

The simulations are run for different number of modes by truncating the mode shape function matrix of dimension 500. The computational time required for a time simulation of 1 s and the error computed from the reference solution are gathered for each simulation and plotted in Figs. 2. Fig. 2 (a) shows that both constant-step decoupling and Newmark schemes reduce drastically the time consumption required for the simulation. The decoupling method requires much lower computational time than Newmark (114 vs. 1964 s for 500 modes). As seen in Fig. 2 (b), this increment of the computational velocity does not compromise the error, which follows the same decreasing curve with the number of nodes than the other two algorithms. Fig. 2 (c) synthesises both figures, indicating that the decoupling technique permits to reach an accurate solution for much lower computational times, then strongly enhancing the numerical efficiency of the time integration.
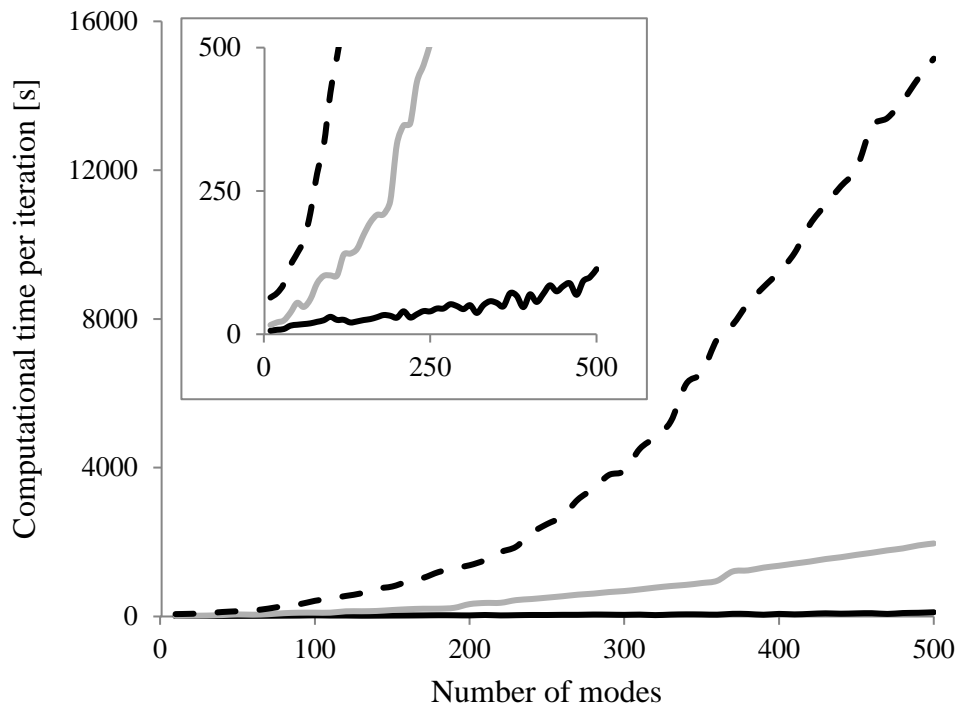
(a)
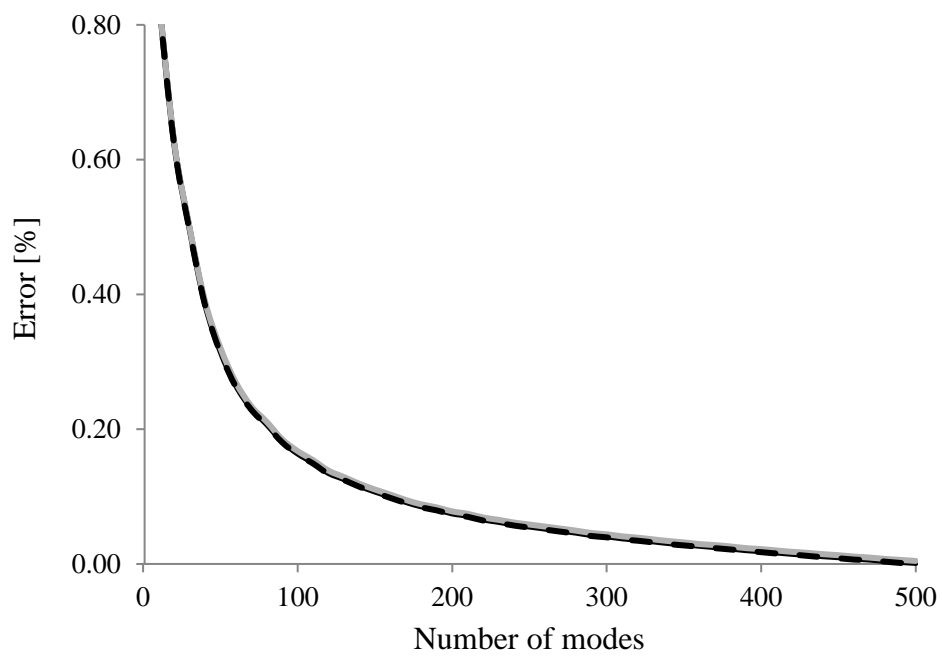
(b)
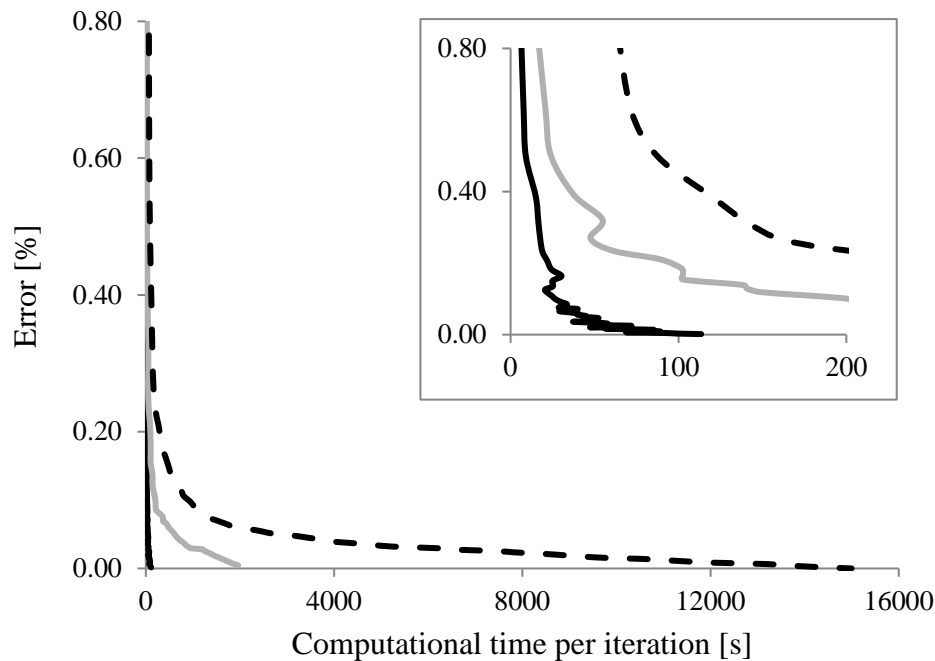
(c)

Figure 2: Comparison of the computational performance of the numerical integration of a 1D MEM track subjected to a precalculated contact force applied in the contact node through the decoupling (━━━), Newmark (━━━) and ode45 (▬ ▬ ▬) schemes. (a) Number of modal coordinates vs. computational time required for a time simulation of 1 s; (b) number of modal coordinates vs. error with respect to the reference solution; (c) computational time vs. error.

ACKNOWLEDGEMENTS

REFERENCES

[1] D. J. Thompson, Railway Noise and Vibration: Mechanisms, Modelling and Means of Control, Elsevier, Oxford, UK, 2009.

[2] H. Hertz, Ueber die Berührung fester elastischer Körper, Journal für reine und angewandte Mathematik 92 (1882) 156-171.

[3] J. J. Kalker, Three-Dimensional Elastic Bodies in Rolling Contact. Kluwer, Dordrecht (1990).

[4] J. Martínez-Casas, E. Di Gialleonardo, S. Bruni, L. Baeza, A comprehensive model of the railway wheelset-track interaction in curves, Journal of Sound and Vibration 333 (2014) 4152-4169.

[5] J. Martínez-Casas, J. Giner-Navarro, L. Baeza, F. D. Denia, Improved railway wheelset-track interaction model in the high–frequency domain, Journal of Computational and Applied Mathematics 309 (2017) 642-653.

[6] C. G. Koh, J. S. Y. Ong, D. K. H. Chua, J. Feng, Moving Element Method for train-track dynamics, International Journal for Numerical Methods in Engineering 56 (2003) 1549-1567.

[7] ISO 3095:2005. Railway applications. Acoustics. Measurement of noise emitted by railbound vehicles, CEN, Brussels, August 2005.

Matrix-free block Newton method to compute the dominant λ-modes of a nuclear power reactor

A. Carreño[♭], L. Bergamaschi[†], A. Martínez[‡], A. Vidal-Ferrándiz[♭],
D. Ginestar[§][*], G. Verdú[♭]

(♭) Instituto de Seguridad Industrial: Radiofísica y Medioambiental,
Universitat Politècnica de València, Spain.
(†) Department of Civil Environmental and Architectural Engineering,
University of Padua, Italy.
(‡) Department of Mathematics "Tullio Levi-Civita",
University of Padua, Italy.
(§) Instituto Universitario de Matemática Multidisciplinar,
Universitat Politècnica de València, Spain.

November 30, 2018

1 Introduction

The neutron diffusion equation is an approximation of the neutron transport equation relying on the assumption that the neutron current is proportional to the gradient of the neutron flux by means of a diffusion coefficient.

For a given a configuration, the criticality of a nuclear reactor core can be forced dividing the fission operator in the neutron diffusion equation by a positive number, λ, obtaining a neutron balance equation: the λ-modes problem. For the two energy groups approximation and without considering up-scattering, this equation can be written as [5]

$$
\begin{bmatrix} -\vec{\nabla}(D_1\vec{\nabla}) + \Sigma_{a_1} + \Sigma_{12} & 0 \\ -\Sigma_{12} & -\vec{\nabla}(D_2\vec{\nabla}) + \Sigma_{a_2} \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} = \frac{1}{\lambda} \begin{bmatrix} \nu\Sigma_{f1} & \nu\Sigma_{f2} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix}, \quad (1)
$$

where ϕ_1 and ϕ_2 denote the fast and thermal flux, respectively. The macroscopic cross sections D_g, Σ_{ag}, $\nu\Sigma_{fg}$, with $g = 1, 2$, and $\Sigma_{1,2}$ are constants that depend on the position.

The dominant eigenvalue indicates a measure of the criticality of the reactor and its corresponding eigenfunction describes the steady-state neutron distribution in the core. Next eigenvalues and their corresponding

---

[*]e-mail:dginesta@mat.upv.es

eigenfunctions are useful to develop modal methods to integrate the time-dependent neutron diffusion equation and to classify BWR (boiling water reactor) instabilities.

To discretize the problem (1), a high order continuous Galerkin finite element method is used leading to a generalized algebraic eigenvalue problem

$$\begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{bmatrix} \begin{bmatrix} \tilde{\phi}_1 \\ \tilde{\phi}_2 \end{bmatrix} = \frac{1}{\lambda} \begin{bmatrix} M_{11} & M_{12} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \tilde{\phi}_1 \\ \tilde{\phi}_2 \end{bmatrix}, \tag{2}$$

where $\tilde{\phi} = \left(\tilde{\phi}_1, \tilde{\phi}_2 \right)^{\mathsf{T}}$ is the algebraic vector of weights corresponding to the neutron flux in terms of the Lagrange polynomials (more details can be found in [6]). The open source finite elements library Deal.II [1] has been used for the implementation of the finite element method.

Different iterative methods have been successfully used to compute a set of dominant eigenvalues and their corresponding eigenvectors of the problem (2). For this computation, we propose to use the modified generalized block Newton method ([2]), where the eigenvectors converge in block. This method requires to solve many linear systems by preconditioned iterative solvers. In this work, we propose several ways to precondition these methods efficiently.

2 The modified generalized block Newton method

The modified generalized block Newton method (MGBNM) was introduced by Lösche *et al* in 1998 [4] for ordinary eigenvalue problems and an extension to generalized eigenvalue problems was studied in [2]. Given the partial generalized eigenvalue problem

$$MX = LX\Lambda, \tag{3}$$

where $X \in \mathbb{R}^{n \times q}$ is a matrix with q eigenvectors and $\Lambda \in \mathbb{R}^{q \times q}$ is a diagonal matrix with the q eigenvalues associated. We suppose that the eigenvectors can be factorized as $X = ZS$, where Z is an orthogonal matrix. Moreover, the biorthogonality condition $W^{\mathsf{T}}Z = I$, is introduced, where W is a fixed matrix. Thus, the problem (3) can be rewritten as

$$MX = LX\Lambda \Leftrightarrow MZ = LZS\Lambda S^{-1} \Leftrightarrow MZ = LZK.$$

Then, the solution can be obtained by solving the non-linear problem

$$F(Z, \Lambda) := \begin{bmatrix} MZ - LZK \\ W^T Z - I_q \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Using the Newton's method, a new iterated solution arises as

$$Z^{(k+1)} = Z^{(k)} - \Delta Z^{(k)}, \qquad K^{(k+1)} = K^{(k)} - \Delta K^{(k)},$$

where $\Delta Z^{(k)} = (\Delta z_1^{(k)}, \dots, \Delta z_q^{(k)})$ and $\Delta K^{(k)} = (\Delta k_1^{(k)}, \dots, \Delta k_q^{(k)})$ are obtained from the solutions of the linear systems

$$\begin{bmatrix} M - L\lambda_i^{(k)} & LZ^{(k)} \\ Z^{(k)^{\mathsf{T}}} & 0 \end{bmatrix} \begin{bmatrix} \Delta z_i^{(k)} \\ -\Delta k_i^{(k)} \end{bmatrix} = \begin{bmatrix} Mz_i^{(k)} - Lz_i^{(k)}\lambda_i^{(k)} \\ 0 \end{bmatrix}, \quad i = 1, \dots, q.$$

The solution of these systems is computed by using the Generalized minimal residual method (GMRES). However, these systems need to be preconditioned (in each iteration and for each eigenvalue) to reduce the condition number of the matrix and to obtain a faster convergence.

2.1 Preconditioning

The first choice for a preconditioner is assembling the matrix

$$A = \begin{bmatrix} M - \lambda_i^{(k)}L & LZ^{(k)} \\ Z^{(k)^{\mathsf{T}}} & 0 \end{bmatrix},$$

and constructing the full preconditioner associated with the matrix. We use the ILU(0) preconditioner since A is a non-symmetric matrix. In other works, it was shown that there are no significant differences if the preconditioner obtained for the matrix associated with the first eigenvalue is used for all eigenvalues in the same iteration. This preconditioner is denoted by P.

To devise an alternative preconditioner without the necessity of assembling the matrix A, we write the explicit inverse of A,

$$A^{-1} = \begin{bmatrix} J^{-1}(I - C_1(C_2^{\mathsf{T}}C_1)^{-1}C_2^{\mathsf{T}}) & J^{-1}C_1(C_2^{\mathsf{T}}C_1) \\ (C_2^{\mathsf{T}}C_1)^{-1}C_2^{\mathsf{T}} & -(C_2^{\mathsf{T}}C_1)^{-1} \end{bmatrix},$$

where

$$J = M - \lambda_i L, \quad C_1 = LZ, \quad C_2^{\mathsf{T}} = Z^{\mathsf{T}}J^{-1}.$$

We desire a preconditioner for A by suitably approximating A^{-1}. Let us call P_J a preconditioner for J. Then, we can define after setting $\tilde{C}_2^{\mathsf{T}} = Z^{\mathsf{T}}P_J$,

$$\hat{P} = \begin{bmatrix} P_J(I - C_1(\tilde{C}_2^{\mathsf{T}}C_1)^{-1}\tilde{C}_2^{\mathsf{T}}) & P_JC_1(\tilde{C}_2^{\mathsf{T}}C_1) \\ (\tilde{C}_2^{\mathsf{T}}C_1)^{-1}\tilde{C}_2^{\mathsf{T}} & -(\tilde{C}_2^{\mathsf{T}}C_1)^{-1} \end{bmatrix}. \tag{4}$$

For instance, $P_J = (LU)^{-1}$, where L, U are the incomplete L and U factors of J. By using the ILU(0) preconditioner of J for P_J, the preconditioner is called $\hat{P}_J$.

The previous preconditioner does not need to assemble the entire matrix A, but it needs to assemble the matrix J to build the ILU preconditioner. Therefore, the next alternative that we propose is using a preconditioner of $-L$ instead of the $J = M - \lambda_1 L$. This preconditioner works well because in

the discretization process, the L matrix comes from the discretization of the differential matrix that has the gradient operators and the diffusion terms. In addition, in nuclear calculations, λ_1 is near to 1.0. We denote by $\hat{P}_L$ the preconditioner $\hat{P}$ where the preconditioner of $-L$ is used to precondition the block J.

Finally, the last alternative to avoid assembling the matrix L is to take advantage of the block structure of this matrix. For that purpose, we carry out a similar process as the one used for matrix A. We write the explicit form of the inverse of L and substitute the inverses by preconditioners. Thus, the preconditioner of L has the following structure

$$
P_L = \begin{bmatrix} P_{11} & 0 \\ -P_{22}L_{21}P_{11} & P_{22} \end{bmatrix},
$$

where P_{11}, P_{22} denote a preconditioner of L_{11} and L_{22}, respectively. The block matrices L_{11}, L_{22} are symmetric and positive definite. Then, we can use as preconditioner the Incomplete Cholesky decomposition. The application of $\hat{P}$ with $P_J = P_L$ is called as $\hat{P}_{\hat{L}}$. However, the main advantage of this preconditioner is that it permits to use a matrix-free implementation that does not require to allocate all matrices. Only, we need to assemble the blocks L_{11} and L_{22} to construct the ILU preconditioners associated.

3   Numerical results

The performance of the preconditioners is studied considering the NEACRP reactor [3]. The four dominant eigenvalues computed have been 1.00200, 0.98862, 0.985406 and 0.985406. The initialization of the MGBNM has been computed using a multilevel technique. The modified block Newton method has been implemented using a dynamic tolerance in the solution of the linear systems. In this case, these values have been $\{10^{-2}, 10^{-3}, 10^{-5}, 10^{-8}, 10^{-8}, \dots\}$. First, we show the results obtained applying directly the ILU preconditioner of A. Table 1 collects the average number of iterations and the total time required by GMRES to reach the residual error of the linear systems given in Tol($\|b - Ax\|$). It is also displayed the time spent to assemble the matrices and to build the preconditioner (Setup time (s)). These data are presented for each iteration and in a total sum. This Table shows that the number of iterations is not very high, but the time spent to assemble the matrix and to construct the preconditioner increases the total CPU time considerably. It is necessary to build in each iteration a new preconditioner for A because of the columns related to the block Z change considerably in each updating.

Table 2 displays these data for the proposed block preconditioner $\hat{P}$ (4) that uses the ILU preconditioner for approximating the inverse of $M - \lambda_1 L$. It is observed that we only need to assemble matrix $M - \lambda_1 L$ and build the preconditioner in the first iteration since we only need to preconditioner

Table 1: Summary of results for the preconditioner P.

n^o it. MGBNM	Tol ($\|b - Ax\|$)	GMRES avg its.	Setup time (s)	Total time (s)
1	10^{-2}	4.5	12.0	18.0
2	10^{-3}	9.75	12.0	20.4
3	10^{-5}	20.75	12.0	25.2
4	10^{-8}	37.5	12.0	33.2
Total		72.5	48.0	96.8

$M - \lambda_1 L$ and the value of λ_1 is very similar for all iterations. The total CPU time of using this block preconditioner has been reduced by more than 26s with respect to the full preconditioner in spite of a (slight) increasing of the average number of the GMRES iterations. This is mainly due to the time saved in the setup stage which goes from 48s (full preconditioner) to 6.6s (block preconditioner).

Table 2: Summary of results for the preconditioner $\hat{P}_J$.

n^o it. MGBNM	Tol ($\|b - Ax\|$)	GMRES avg. its	Setup time (s)	Total time (s)
1	10^{-2}	8.25	6.6	12.9
2	10^{-3}	13.25	–	9.5
3	10^{-5}	23.25	–	16.6
4	10^{-8}	41.25	–	30.0
Total		86.0	6.6	70.0

The next computations are obtained by using the block preconditioner, $\hat{P}$, but, in these cases, approximating the $(M - \lambda_1 L)^{-1}$ by the ILU preconditioner of $-L$ ($\hat{P}_L$) and by a block preconditioner of $-L$ ($\hat{P}_{\hat{L}}$). The most relevant data to compare the preconditioners considered in this work are exposed in Table 3. These are the total iterations of the GMRES, the total setup time, the total time to compute the solution and the maximum computational memory occupied by the matrices. We observe that the number of iterations increases when worse approximations of the inverse of A are considered, but the setup time that needs each preconditioner becomes smaller. Moreover, the maximum memory occupation (max RAM) is also reduced significantly. In the total CPU times, we observed that the block preconditioner ($\hat{P}$), in all its versions, improves the times obtained of applying directly the ILU preconditioner of A. Among the possibilities for obtaining a preconditioner of $M - \lambda_1 L$, there are not big differences in the computational times but

there is an important saving up in the computational memory. The best results are obtained by the $\hat{P}_{\hat{L}}$ preconditioner if the computational memory consumption is taken into account.

Table 3: Summary of results obtained of using different preconditioners.

Prec.	GMRES Its	Setup time (s)	Total time (s)	Max RAM (Mb)
P	72.5	48.0	96.8	2062
$\hat{P}_J$	86.0	6.6	70.0	1418
$\hat{P}_L$	98.0	4.4	77.2	787
$\hat{P}_{\hat{L}}$	100.25	1.8	76.3	319

4   Conclusions

Different implementations for preconditioning the linear systems to be solved at each iteration of the modified block Newton method, have been studied as an alternative to assemble the full matrix and construct a preconditioner in each iteration. These new implementations for the preconditioner break down the setup cost at the price of a slight increasing of the number of iterations. The result is a significant reduction of the total CPU time needed to reach the convergence and the memory occupancy. Among the implementations studied in this work, it is shown that the best option is the one that takes into account the structure of the matrix L. This implementation allows to implement the MGBNM using a matrix-free technique thus greatly reducing the memory consumption.

References

[1] W Bangerth, R Hartmann, and G Kanschat. deal.II – a general purpose object oriented finite element library. *ACM Trans. Math. Softw.*, 33(4):24/1–24/27, 2007.

[2] A Carreño, Antoni Vidal-Ferràndiz, D Ginestar, and G Verdú. Spatial modes for the neutron diffusion equation and their computation. *Annals of Nuclear Energy*, 110:1010–1022, 2017.

[3] H Finnemann and A Galati. *NEACRP 3-D LWR core transient benchmark, final specification*. 1991.

[4] R. Lösche, H. Schwetlick, and G. Timmermann. A modified block Newton iteration for approximating an invariant subspace of a symmetric matrix. *Linear Algebra and its Applications*, 275:381 – 400, 1998.

[5] Weston M Stacey. *Nuclear reactor physics.* John Wiley & Sons, 2018.

[6] A Vidal-Ferrandiz, R Fayez, D Ginestar, and G Verdú. Solution of the lambda modes problem of a nuclear power reactor using an h–p finite element method. *Annals of Nuclear Energy*, 72:338–349, 2014.

A new automatic gonad differentiation for salmon gender identification based on Echography image treatment

Ana Sancho [1], Laura Andrés López[2], Beatriz Baydal Giner[3], Julia Real Herráiz[4]*

[1,2, 3, 4] Institute of Multidisciplinary Mathematics, Polytechnic University of Valencia, Camino de Vera, 46022 Valencia, Spain

*Corresponding author. E-mail: ansanbru@upvnet.upv.es. Telephone: +34 650528210

1. Introduction

Female salmon production is advantageous for salmon farming industry since this gender is less susceptible to illnesses, results in a better quality product and also it allows to increase the farm productivity by needing less resources and attentions than male specimens.

That is the reason why a low cost, portable and precise gonad differentiation system is claimed by the fishing industry. The solution proposed is an automatic and intelligent diagnosis system able to distinguish the salmon gender on early stages of its life (juvenile). The aim of this research is to design and develop an Echography reconstruction algorithm based on morphological mathematical operators to remove the noise component and increase the image resolution for its later analysis.

2. Ultrasound Scan

During an ultrasound scan, a piezoelectric glass generates ultrasound pulses that are partially reflected when arriving to a two different materials interface. When these echoes return to the piezoelectric glass, depending on their intensity, numerical register vectors are created.

For each sensor the evolution in time of the echo's intensity is registered. After a discretization process, in which the time is divided into 'N' sections, the data is stored in a vector.
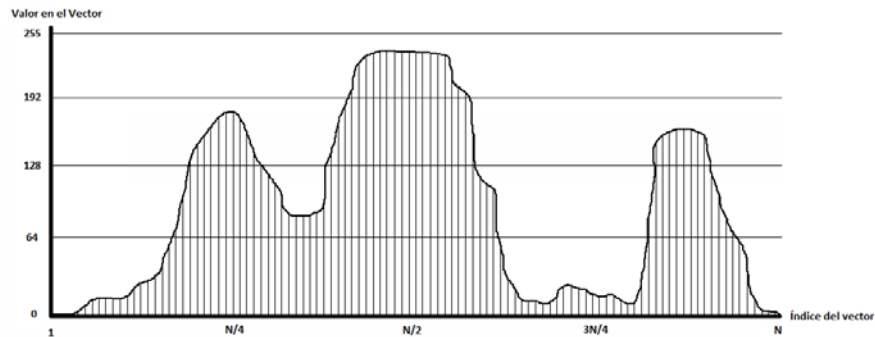
Fig. 1: Echo intensity evolution.

For an ultrasound scan system formed by 'M' sensors, a [N x M] matrix will contain all of the registers. Every matrix element is ranged from 0 (black) to 255 (white), composing an 8 bit image. This is called 'rough picture'.
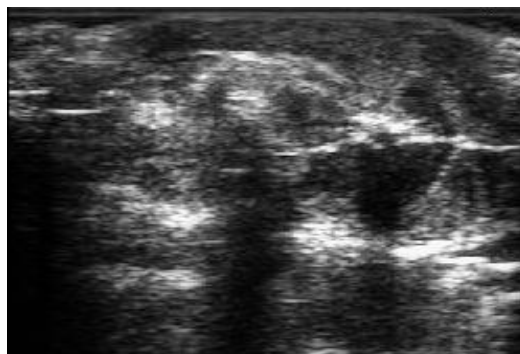
Fig. 2: Rough picture obtained during test.

3. Image treatment

Rough picture contains irrelevant elements known as *artifacts*. To remove these artifacts, the first step is to create a binary image. This filtering procedure is known as *thresholding*. It consists on replacing each pixel in an image comparing to a threshold. Intensity values higher than the threshold become white, and lower values become black. Only white areas are of interest.

Fig. 3: *Binary images. From left to right: Threshold 100, 135 and 170.*

It was determined, by trial-error procedure, that the most useful results were obtained for a threshold value of 135.

Next step in image treatment is known as 'segmentation'. It consists of using morphological algorithms in order to emphasise on picture elements of interest – gonads -. This is achieved by consecutive application of erode-dilation processes.

Eroding is a mathematical process where elements whose size is less than a determinate resolution become removed and bigger ones, become reduced. With this, every element of the picture smaller than supposed size of gonads is removed. Its expression is:

$$A \ominus B = \{x / B_x \subseteq A\} \tag{1}$$

Where 'A' is the original image and 'B' is the structural element.

Later recovering of the main elements size and shape is necessary. That is the reason why a dilation process is required.

$$A \oplus B = \{x \mid (\hat{B})_x \cap A \neq \emptyset\} \tag{2}$$
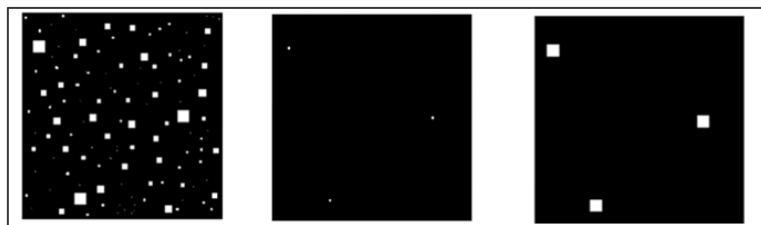
Fig. 4: *Original (left), eroded (mid) and dilated (right) images.*

After three consecutive eroding-dilation processes, images are prepared for abdomen identification.

Fig. 5: Binary and segmented image of salmon abdomen.

4. **Abdomen identification**

A characterization algorithm has been programmed in MATLAB in order to establish a classification criterion of size and shape patterns of stomach. These characteristics are based on its width, height and centroid coordinates. The result of this phase is a database with characteristics that allows later classification.

Attending to those 'patterns', by means of Simple Bayesian Classifiers, it was possible to focus the area of interest closer to the stomach – where gonads are most probably to be founded -.
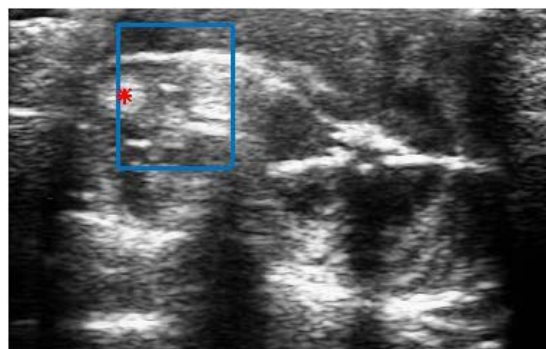
Fig. 6: Gonads area location (centre of the stomach marked on red)

Finally, for salmon gender identification, Histogram of Oriented Gradients (HOG) technique is chosen for gonad identification and classification.
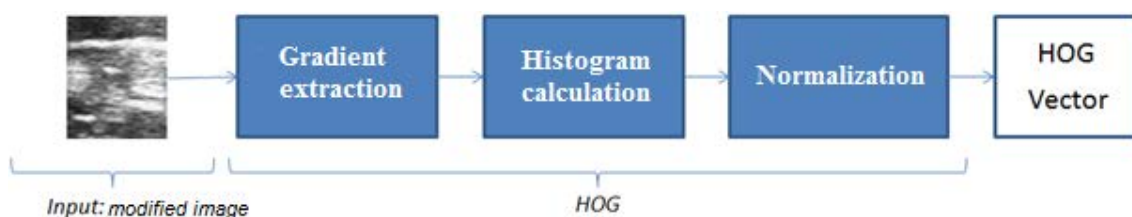
Fig. 7: HOG descriptor phases.

This procedure divides the image in 'n' small regions – cells – and obtains, for each one, a histogram from its pixels gradient orientation. In addition, for a better response, contrast should be normalized in larger areas, which are called blocks.

The size of cells is determined as:

$$C = C_x \times C_y \tag{3}$$

Where C_x and C_y are the numbers of pixels in each axis.

The directional gradients of the image (∇I_x y ∇I_y) are calculated through discrete differential operators, such as Sobel, Prewitt or Laplace. From these, the gradient's magnitude and phase can be obtained:

$$|\nabla I| = \sqrt{\nabla I_x^2 + \nabla I_y^2} \tag{4}$$

$$\varphi = \arctan 2 \left(\frac{\nabla I_y}{\nabla I_x} \right) \tag{5}$$

The histogram should be robust to differences in image contrast. For that reason, the magnitude of the gradient $|\nabla I|$ is normalized as $|\nabla I|_N$:

$$|\nabla I|_N = \frac{|\nabla I|}{\sum_{cell} |\nabla I|} \tag{6}$$

Where:

$$\sum_{cell} |\nabla I|_N = 1 \tag{7}$$

The value of the parameter in each pixel is calculated by multiplying the value of the normalized magnitude of the gradient by the corrected angle:

$$g = \varphi_c \cdot |\nabla I|_N \tag{8}$$

Finally, after obtaining all cell histograms of the image, the image is then divided into blocks. For each block k, the histograms of the cells h_{cij} that are part of it are

concatenated, where the subscripts i and j indicate the row and column of the cells in the block. The resultant vector is:

$$\vec{v_k} \tag{9}$$

$$= \left[[h_{11}][h_{12}]\ldots[h_{1j}]\ldots[h_{1b_x}][h_{21}]\ldots[h_{2j}]\ldots[h_{2b_x}]\ldots[h_{ij}]\ldots\left[h_{b_y1}\right]\left[h_{b_y2}\right]\ldots\left[h_{b_yb_.}\right] \right)$$

This vector is inserted into a Support Vector Machine classifier, where the detection criteria for gonad identification is established as finding in, at least, four of the cells, a HOG pattern similar to obtained by a true gonad containing cell. For this purpose, previous SVM algorithm train period is required.

All of these functions have been already implemented in MATLAB: 'svmtrain' for previous training phase and 'svmclassify' for gonad existence positive or negative classification.

In these early stages, gonads are not completely formed yet. Since male salmons take longer to develop their reproductive organs, only female ones are appreciable.
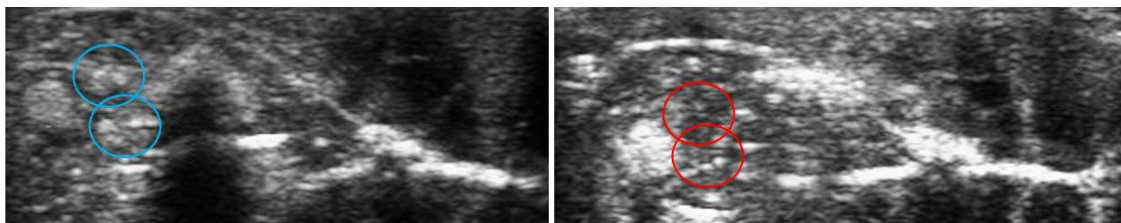
Fig. 8: Female salmon scan (left) and male salmon scan (right). Gonads location marked.

References

[1]   C. Pineda, A. Bernal, R. Espinosa, C. Hernández, N. Marín y A. Peña, «Principios físicos básicos del ultrasonido,» Rev. chil. reumatol., vol. 25, nº 2, pp. 60-66., 2009.

[2]   US, Apuntes de Ingeniería Informática. Tema 4: Morfología binaria (Parte I), Universidad de Sevilla, 2016.

A New Earthwork Measurement System based on Stereoscopic Vision by Unmanned Aerial System flights

Victor Espert[1], Piter Moscoso Godoy[2], Teresa Real Herraiz[3], Manuel Martinez Grau[4], Julia Real Herráiz[5]*

[2, 3, 4 ,5] Institute of Multidisciplinary Mathematics, Polytechnic University of Valencia, Camino de Vera, 46022 Valencia, Spain

[2] DGOP, Ministerio de Obras Públicas, Morandé 59, Santiago de Chile, Chile

*Corresponding author. E-mail:jvictorespertball@gmail.com. Telephone: +34 650528210

Introduction

Almost any civil construction has an important proportion of its budget earmarked for earthworks. In addition, due to the fact that the payment of these tasks is carried out by the same contractor, it is very usual to find a lack of confidence on the actual realized consignment. That shows the necessity of applying measures for real control of the actual stocked or moved earth volume to really know the development of the execution.

In this way, the purpose of this research project is to develop a new economical earthwork control system, based on the 3D reconstruction of the work area by using stereoscopic vision techniques applied to a set of pictures taken from a HD camera assembled in an Unmanned Aerial System (UAS or 'drone') equipped with GPS.

1. Stereoscopic vision

From two different pictures (2D) taken in two adjacent points, it is possible to obtain the depth or the third dimension (3D) by means of a triangulation process, as humans and animals do with their eyes [1].
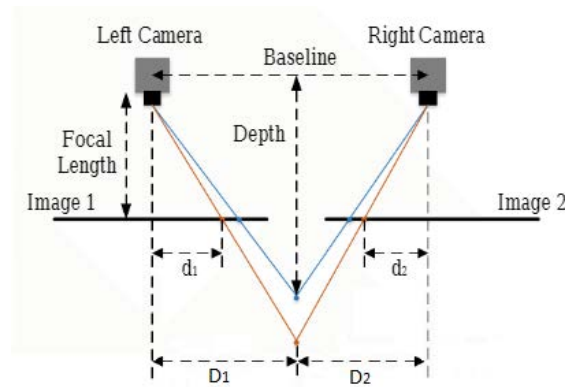
Fig. 1: Stereoscopic vision model with two different cameras or vision points. [2]

So, for a specific point it is possible to calculate the distance to the plane where the images have been extracted by similarity of triangles.

$$\left.\begin{aligned} d_1 &= f \cdot \frac{D_1}{Z} = f \cdot \frac{D}{Z} \\ d_2 &= f \cdot \frac{D_2}{Z} = f \cdot \frac{D-b}{Z} \end{aligned}\right\} \rightarrow D = \frac{d_1 \cdot b}{d_1 - d_2} \rightarrow Z = \frac{b \cdot f}{\Delta d} \tag{1}$$

For this purpose, it is necessary to take various pictures of the same area with at least a superposition of 30-50% in each pair of sequential images.

2. Description of the KTL algorithm

When a camera captures an image, a slight distortion is introduced. So, a calibration of the camera is required, in order to obtain its focal characteristics and the distortion produced. The distortion coefficients of the lens can be used to obtain the undistorted pixel locations.

The identification of coincident points is automatically carried out by applying the 'Matching based on properties' technique, which consists on the identification and match of singular elements (edges, apexes, points, etc.) [3].

After that, those points become tracked and a Bundle Adjustment- by means of KLT algorithm (Kanade-Lucas-Tomasi) - allows to know the relative location between both

pictures, which is transformed to global {X,Y,Z} coordinates using the position of the previous photography. The relative distance is calculated as:

$$L = \sqrt{\sum_{x \in R} \left(F_{(x+h)} - G_{(x)} \right)^2}$$
(2)

Where F(x) and G(x) are both functions that represent the location 'x' of each respective picture – 'x' is a vector -.

The KLT algorithm is based on the idea of a local search using gradients weighted by an approximation to the second derivative of the image. So, the disparity vector, *h*, can be estimated as:

$$h \approx \left[\sum_x [G_{(x)} - F_{(x)}] \cdot \left(\frac{\partial F}{\partial x} \right) \right] \left[\sum_x \left(\frac{\partial F}{\partial x} \right)^T \cdot \left(\frac{\partial F}{\partial x} \right) \right]^{-1}$$
(3)

The KLT algorithm is an iterative process that performs a pyramidal search. It generates a pyramid of images. In each level of this pyramid, the resolution of the image is reduced compared to the inferior image. The algorithm initially performs the search at the upper level until it converges, and then transmits the information to the lower level, where the search is repeated, up to the last level.
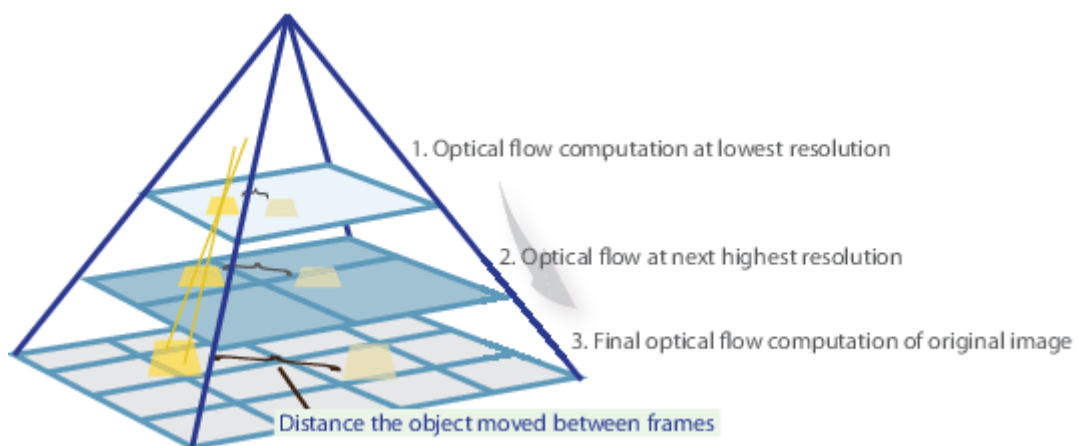
Fig. 2: *Scheme of pyramidal operation of the KLT method. Source: Mathworks.com*

3. Reconstruction process

Although the KLT algorithm has its own error detection, it could be possible that a small number of points are detected incorrectly, generating disparities which fall away from the global average of the image plane.

The way to detect them is by means of a histogram of disparities, defining minimum thresholds of repetition, so that the points that are not around the image plane are eliminated.
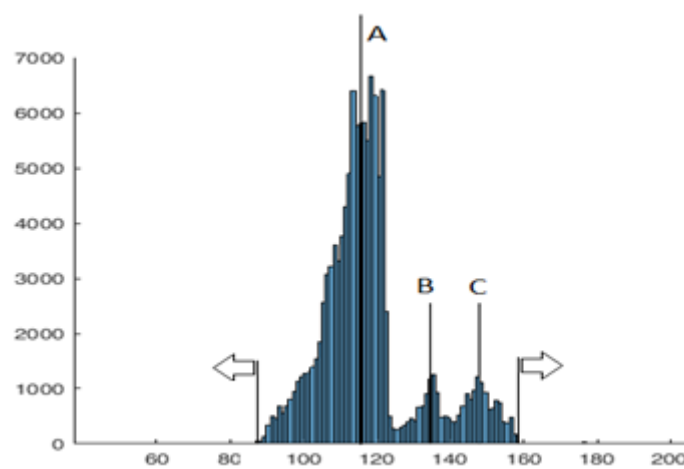
Fig. 3: Detection of erroneous points based on disparities histogram.

Due to the characteristics of the method based on characteristic points, only the disparity of these points is known. So, only a cloud of points is available from the scene to be reconstructed. However, the procedure has been adjusted in order to obtain enough points to make the reconstruction feasible (estimating the disparity of the points that are not known).

In order to obtain full images with high resolution, a moving average filter has been implemented in several iterations so that the window size is increased between consecutive iterations.
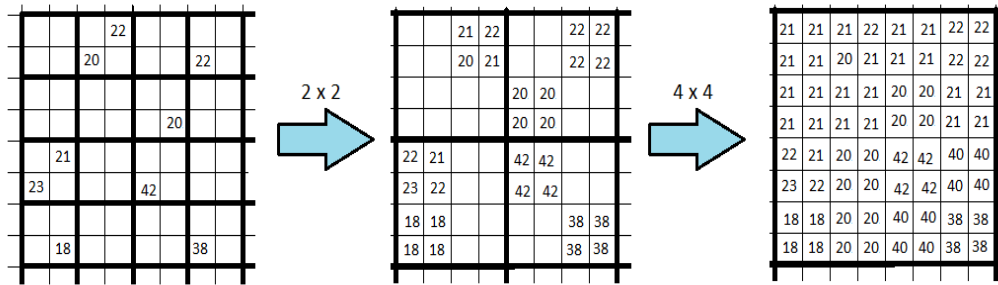
Fig. 4: *Solid image reconstruction process.*

4. Results and discussion

Once the reconstruction is finished, a low-pass filter is applied to smooth the edges and homogenize the result. This filter was calibrated to not miss ground relief peaks. Finally, for each pixel/cell, the volume is obtained as:

$$V = L \cdot W \cdot (Zt - Zb) \tag{4}$$

Where L and W are the length and wide of the cell respectively; Z_t is the terrain height and Z_b is the base ground height.

Once thickness is known, as the other two dimensions, it is possible to carry out an earthworks balance.

This new system has been developed, primarily, for linear projects – railways, roads, pipes, ducts, etc. - where huge terrain volumes are involved in one preferential direction and not always its balance control results an easy task.

To achieve this purpose, it is a usual practice to establish longitudinal profiles of terrain excavation-filling regions along the whole layout. This is carried out during the project design phase and not modified during its execution – except in eventual situations for major reasons -.

One of the most important advantages of this new solution is that with geometrical obtained data from UAV, it is possible to reproduce another longitudinal earthwork balance profile and directly compare it with the original plan.

As volume results data are stored with its relative location, after every UAV flight, the database becomes updated and deviations from project foresight could be detected and corrected in real time.
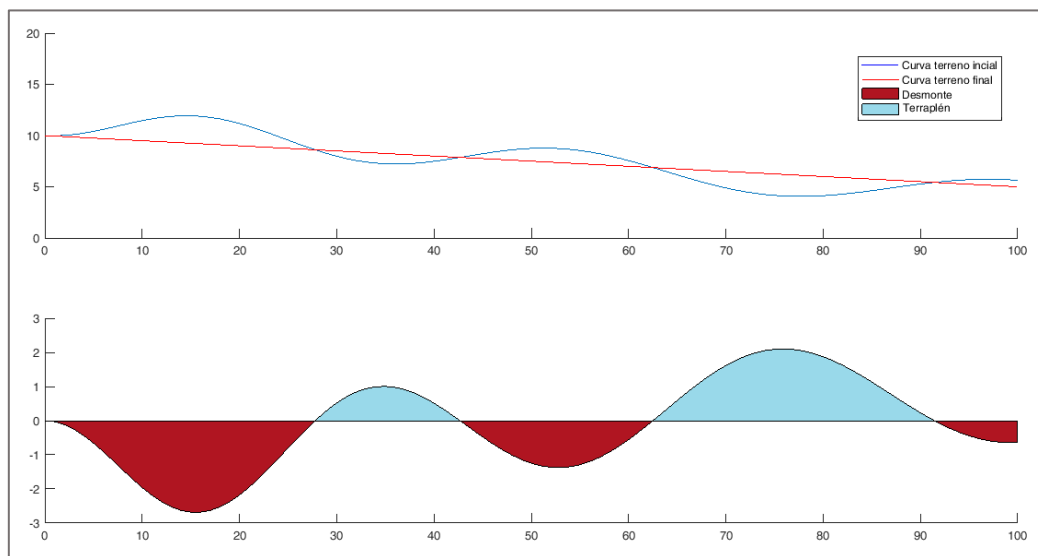
Fig. 5: Earthworks balance longitudinal profile (upper) and excavation / filling absolute volumes vs. longitudinal distance (lower).

References

[1]   Guerrero, J.M, Pajares, G., Guijarro, M. (2011) Técnicas de procesamiento de imágenes estereoscópicas. Departamento de Ingeniería del Software e Inteligencia ArtificialUniversidad Complutense de Madrid.

[2]   Liu, Q., Li, R., Hu, H., Gu, D. (2015) Extracting semantic information from visual data: a survey. School of Computer Science and Electronic Engineering, University of Essex.

[3]. Shi, J., Tomasi, C. (June 1994) Good Features to Track. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 593–600.

A New Forest Measurement and Monitoring System Based on Unmanned Aerial Vehicles Imaging

*Fran Ribes-Llario[*1], Victor Ramos[2], Victor Espert[3], Julia Real Herráiz[4]*

[1,2,3,4] Institute of Multidisciplinary Mathematics, Polytechnic University of Valencia, Camino de Vera,

46022 Valencia, Spain

*Corresponding author. E-mail: frarilla@upv.es. Telephone: +34 650528210

1. Introduction

According to UN data (www.un.org), , global population by June 2017 will exceed in up to 7.6 billion and predictions are that it will increase up to more than 8.6 billion in 2030. As a consequence, natural resources are becoming increasingly scarce. This problem constitutes in many countries supply-demand disequilibrium between forests and wood industries. This results in notorious consequences for economy and environment. In addition, existent measurement and monitoring solutions for forest balance status – growth, decline, illegal practices detection, etc – are shown to be inadequate for this purpose.

For this reason, a new forest status monitoring system based on UAV images taken from has been developed. This solution makes possible: i) Tree detection and account, ii) Individual dendrometric variable determination (tree height and treetop diameter). As a result, from these data, forest mass growing models, biomass and wood quantification and deforestation/degradation models could be developed and achieve an optimum management of forest resources.

2. Tree detection

To achieve this goal, a new image processing algorithm has been developed. First of all, an orthophoto of the whole interest area must be taken. This is carried out by overlapping subsequent images. A Genetic Algorithm, whose role is to maximize the correlation between two sections of both pictures – to compare 20% of the image -, attains this purpose.

Then, this orthophoto is converted from initial RGB to Lab space. After that, three different range filters will be defined for each channel (L, a & b), according to each tree typology studied (pine, eucalyptus and carob) to distinguish between trees and ground pixels.

To avoid false interpretations, while the image obtained is processed we correct possible 'holes' in normal and inverted colour image. All of this is also implemented by using MATLAB image treatment tools.

Fig. 1: Left to right: Original, LAB filtered, Corrected image

Once the image has been corrected, the tree detection algorithm is programmed. This algorithm identifies each object whose size is among 3500 and 13000 pixels as a tree. After each detection process, each tree is removed from the image and saved in another image, and the remaining objects are eroded. Objects of less than 3500 pixels are also deleted.

The erosion and identification process continues until there are no objects left in the image.
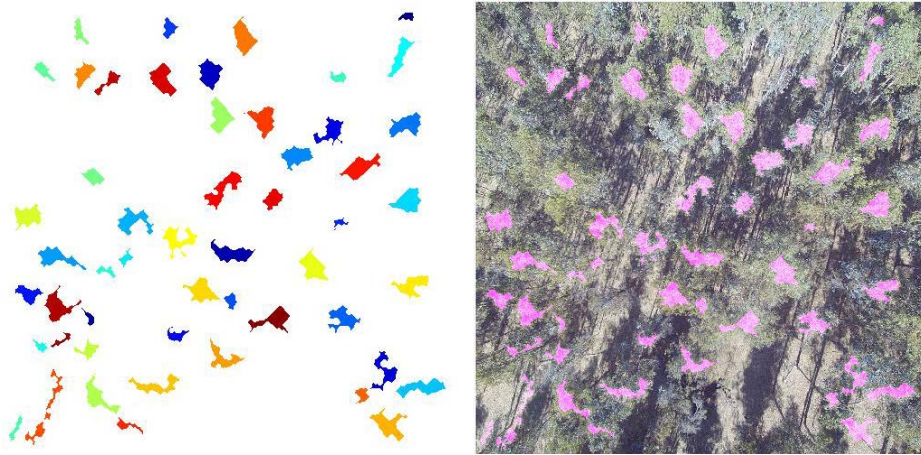
Fig. 2: *Left to right: identified trees, and overlapped image.*

Finally, these objects – trees - are counted, obtaining the exact number of specimens in the studied area.

3. Estimation of characteristics

Once every tree has been identified, it is necessary to estimate the tree characteristics. Those characteristics are the height, and the treetop and trunk diameter.

The height of the trees is estimated by spectroscopy. From two different 2D pictures of the same object taken with a separation of 'L', it is possible to obtain the third dimension, as animals and human do with their eyes, in this case, the depth.

$$L = \frac{t_x \cdot d}{2d_p \cdot \tan\frac{\varphi}{2}} \tag{1}$$

Where 'L' is drone-to-point distance, 't' is axis resolution where the lag is produced, 'd' is distance covered by the drone between both pictures, 'd_p' is this same distance in pixels and Φ is the vision angle (rad.).

This procedure can only be applied to isolated and perimeter trees, because as they are the only ones whose lateral projection is observable. The height of the other trees is assumed to equal the average height of the perimeter trees.

The *Surface Digital Model* (SDM) [1] was obtained after that, which consists of a 3D representation of the area of interest. The difference between this SDM and *Terrain Digital Model* (TDM) – which represents the 3D referent relief, without plants, trees, etc. and must be known previously – forms the *Canopy Height Model* (CHM) [2].
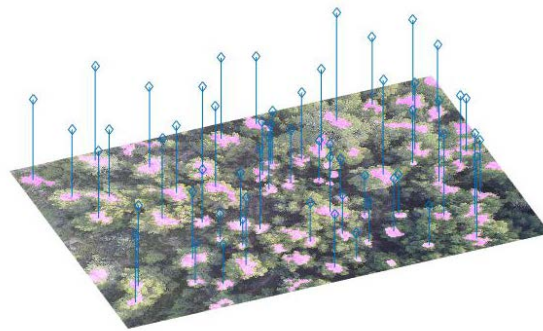
Fig. 3: *CHM image of the trees.*

The height of each tree is implicit in CHM model. Obtained precision in controlled tests has been reached until 95-96 % in height determination.
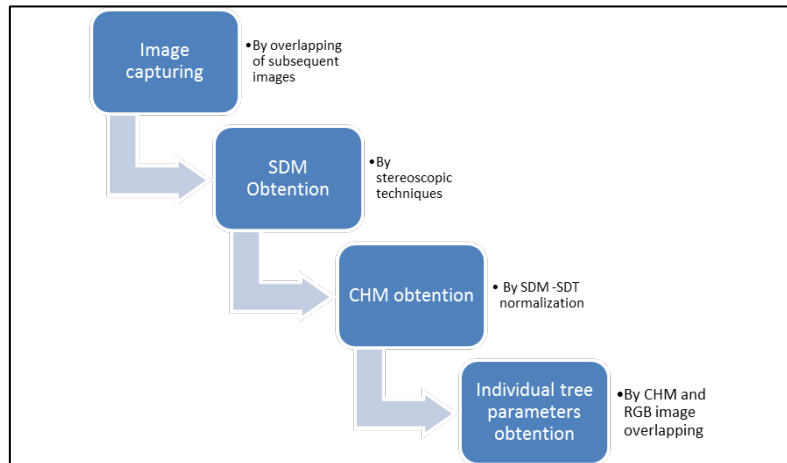
Fig. 4: *Algorithm procedure scheme.*

The treetop diameter will be obtained based on the results of the detection algorithm. The treetop diameter is estimated as the diameter of the disc equivalent to the area of the identified objects. The formula used would be the following:

$$\emptyset = 2 \cdot \sqrt{\frac{A}{\pi}} \tag{2}$$

The difficulty of obtaining the trunk diameter is the same as the height estimation: the treetop is hiding the trunk. So, it is assumed that the trunk diameter of the whole area is equal to the average measured on the perimeter.

For wood volume estimation, two empirical models are possible [3]: i) Wooded area charts as a group – depending on average trunk perimeter, covered area and tree species -, ii) Individual calculated volume – total volume is the summation of each tree -. In this case, empirical expressions depending on the specie, with/without bark volume and calculation model – determine the total volume.

4. Forest control

The data obtained from the previous procedure in different moments allow the analysis of the forest evolution. The growth and deforestation estimation are based on the comparison of different time images. Therefore, the methodology proposed for this control is composed of two stages: i) Comparison of areas with and without vegetation. This stage consists of selecting the layer that exclusively contains zones for the monitoring of change processes, ii) Comparison of images from different dates - to identify degradation/recovery-.

In this context, it is necessary to refer to types of vegetation density depending on the species determined at the beginning of the campaign, since depending on whether it is a native forest or a plantation it will be more productive to study the occupied surface or the volume of wood.

For native species, the growing is calculated as:

$$\Delta BA = BA_{t2} - BA_{t1} \tag{3}$$

Where 'BA' is the basal area of each moment. A negative growing means deforestation of the native forest.

In the case of a plantation, it is interesting to know the increase of the wood volume and the growth rate:

$$\Delta V = V_{t2} - V_{t1} \tag{4}$$

$$Ratio = \Delta V/(t_2 - t_1) \tag{5}$$

In addition, biomass could be also obtained, which is a recognized indicator of the environment status evolution. The method of E. Canafa for biomass estimation requires a subdivision of the area in circular plots. Each one of those plots requires a correction factor. This correction factor is a coefficient that approximates empirical data to real plots, which is governed by the following equation:

$$Fc = \frac{10000}{\pi * R^2} \tag{6}$$

For the calculation of the biomass, only those plots with an occupation of more than 80% are selected, that is in this case they have a number of trees higher than half of the average of the plots approximately.

The equation to calculate a tree biomass is:

$$W = 0.009892 * (d^2 * h)^{1.023} - 0.00434 * d^2 * h + 61.57 - 6.978 * d + 0.3463 * d^2 \tag{7}$$

Where 'W' is biomass dry weight for each tree, 'd' is the diameter of the trunk and 'h' its calculated height.

For the forest calculation, it is necessary multiply for the number of trees and the corrective factor, obtaining the biomass Kilogram/hectare:

$$BIOMASS = \frac{(W1 * N1) + (W2 * N2) + \cdots + (Wn * Nn)}{A} * Fc \tag{8}$$

References

[1]   Weibel, R. & Heller, M. (1991). Digital Terrain Modelling Geographical Information Systems: Principles and Applications.

[2]   Hien, L. Jee-In, K. (2015). Calculation of tree height and canopy crown from drone images using segmentation. Korea.

[3]   Torre Tojal, L. (2016). Diseño y contraste de nuevos modelos de estimación del potencial energético de biomasa forestal en el Territorio de Bizkaia mediante técnicas de análisis estadístico espacial usando herramientas GIS con datos LIDAR.

A new non-intrusive and real time monitoring technique for pavement execution based on Unmanned Aerial Vehicles flights

Teresa Real Herraiz[1], Piter Moscoso Godoy[2], Victor Espert[3], Ana Sancho[4], Julia Real Herráiz[5]*

[1, 3, 4, 5] Institute of Multidisciplinary Mathematics, Polytechnic University of Valencia, Camino de Vera, 46022 Valencia, Spain

[2] DGOP, Ministerio de Obras Públicas, Morandé 59, Santiago de Chile, Chile

*Corresponding author. E-mail: tereaher@upv.es. Telephone: +34 650528210

1. Introduction

In pavement construction, independently of its composition – asphalt or concrete -, its early stages become crucial for its adequate durability, resistance, capacity and final refinements in general. An inadequate care during its installation will incur, in most cases, in third part discords, penalties and, naturally, in a lower quality product – which could involve high economical losses -.

Current execution quality control tools and techniques in this matter have some inconveniences: are not always applicable, could be expensive and could negatively affect to the normal development of its installation – delays, specific times spent on its control, provide results in delayed time, etc. -.

For this reason, a new non-intrusive real time monitoring is proposed herein. It consist of a commercial Unmanned Aerial Vehicle (UAV or drone) equipped with a HD camera and in which have been added a thermal camera and a LIDAR scanner system.

In this way, pavement main properties can be monitored when work managers desire. All of the collected data is sent via Wifi to a remote host, where data processing software is installed.

2. Control parameters

There are three essential parameters for quality control during pavement execution. These parameters are: i) layer thickness, ii) temperature and iii) degree of compaction.

Moreover, what its controlled trough them is exposed herein:

i) Layer thickness: As in asphalt pavements or in concrete pavements, granular subjacent layers, whose execution is critical for surface integrity, must be extended and compacted in limited thickness ratios which are not often accomplished. From it depends the **final bearing capacity** of the blanket. In addition, especially in rigid pavements, surface layer – usually made of reinforced concrete – require to strictly accomplish with minimum project thickness because of **resistance and cracking**.

ii) Temperature: In asphalt pavements, its viscosity and, as a consequence, its **workability, compaction and density** are closely related with its installation temperature. A uniform temperature distribution implies a homogeneous layer densification and avoids thermal segregation. On the other hand, in concrete pavements, temperature is a reliable indicator of **curing process** quality and, as a consequence, about its **final resistance**.

iii) Degree of compaction: Absence of air occlusions is usually synonym of major **final resistance / capacity** ratios. In open-gap asphalt mixtures – for drainage encouraging -, even a certain pores volume is desired, it must be also controlled. It is closely related with density and it is controlled as in surface layer (pavement itself) as in subjacent structural layers.

3. Method

Once the importance of each control parameter is exposed, how does the new solution perform this monitoring is detailed:

i) Layer thickness: For its monitoring, LIDAR scanner creates several subsequent cylindrical {r,y,z} cloud of points which result, after drone GPS synchronization in a virtual 3D image of the environment. It is necessary to assure that the whole area of interest has been covered during drone flight.

To achieve that 3D virtual image and also be capable to measure distances and layer thicknesses, first it is necessary to correct registered points coordinates with rigid body movements – translations and rotations of the UAV -. For this purpose a coordinate transformation matrix (T) should be carried out – more details of this procedure on [1] -.

$$T = \begin{bmatrix} \cos(P)\cos(Y) & -\cos(P)\sin(Y) & \sin(P) & t_x \\ \cos(R)\sin(Y) + \sin(R)\sin(P)\cos(Y) & \cos(R)\cos(Y) - \sin(R)\sin(P)\sin(Y) & -\sin(R)\cos(P) & t_y \\ \sin(R)\sin(Y) - \cos(R)\sin(P)\cos(Y) & \sin(R)\cos(Y) + \cos(R)\sin(P)\sin(Y) & \cos(R)\cos(P) & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (1)$$

Where 'R', 'P' and 'Y' are the three rotation angles: roll angle, pitch and heading. And t_x, t_y and t_z are the translation along the three axes.

After that, layers thicknesses could be obtained as:

$$h_i = c \cdot \Delta t / 2 \quad (2)$$

where 'h_i' is the thickness of layer 'i'; 'c' is the known propagation velocity of the wave and 'Δt' is the wave-travelling time lapse.

ii) Temperature: a thermal camera allows to know in real time surface temperature of the pavement. For asphalt pavement it is possible to observe 'cold points' or other defects in layer being able to act immediately during its construction [2,3].

In case of concrete pavements, observed temperature could be correlated with its maturity level and determine if curing process is being carried out successfully or needs any corrective action.
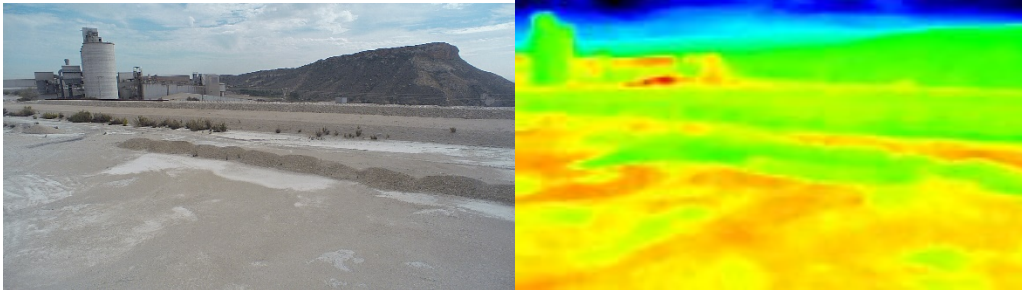
Fig. 1: Thermal image obtained on a test flight during drone devices calibration.

iii) <u>Degree of compaction:</u> It is not possible to direct measure the degree of compaction by non-intrusive measures. For this reason, a numerical FEM model combined with LIDAR thickness measurement results could determine, indirectly, the number of compaction cycles needed to achieve an optimal density value. According with [4], non-linear elasto-visco-plastic response of asphalt during compaction is defined in terms of both constitutive functions: shear modulus function $G(n)$ and viscosity function $\eta(n)$, which depend on the 'n' rolling compaction cycles supported . It changes its status from fluid-like to high viscosity fluid-like status during compaction. For this reason, from a 3D FEM numerical model implemented in ANSYS LS DYNA v14, by a steady-state analysis, the optimal compaction cycles is estimated.

4. Results and conclusions

The solution proposed herein is able to obtain several indicators of the final quality of the pavement, being able to carry out a continuous, global and remote quality control. In addition, these tasks are executed simultaneously to the process construction of each layer that conforms the pavement.

Through the temperature image is possible to control the viscosity of the asphalt mixture or the maturity of the concrete.

By means of combining the thickness variation for each section of the work, it is also possible to control the base or pavement layer compactness and subsequently, is it possible to predict its final capacity.

In this way, it is expected that this new application of thermal scanner, LIDAR and GPS technologies' combination greatly ease the complex task of quality control during the execution of pavement works.

In addition, the potential of this solution does not end only in this kind of works. Other linear works' execution, as railway layouts, channels or ducts that also require the quality control in real time along all of its route could be benefited of this technology as in other constructions that require massive volumes of concrete – building's foundations, platforms, power plants, airports, etc -.

New commercial trending in terms of quality and safety control in civil works are searching for solutions that do not affect to the normal development of the tasks, manage the information in real time and whose results could not be influenced by the human factor – perception deviations of the controller -. For this reason, future lines of investigation will be focused on all of the aforementioned possibilities.

Additionally, the use of commercial devices globally extended enhance the accessibility of the solution in terms of availability and economy.

References

[1]   Puente, I., Solla, M., González-Jorge, H., Arias. P. (2013) Validation of mobile LIDAR surveying for measuring pavement layer thicknesses and volumes. NDT&E International 60. Pp. 70-76.

[2]   Pascucci, S., Bassani, C., Palombo, A., Poscolieri, M., Cavalli, R. (2008) Road asphalt pavement analyzed by airbone termal remote sensing: preliminary results of the Venice Highway.

[3]   Mettas, C., Themistocleous, K., Neocleous, K., Christofe, A., Pilakoutas, K., Hadjimitsis, D. (2016). Monitoring Asphalt pavement damages using remote sensig techniques.

[4]   Masad, E., Scarpas, T., Rajagopal, K.R., Kassem, E., Saradhi, K., Kasbergen, C. (2014). Finite element modelling of field compaction of hot mix asphalt. Part II: Applications.

A New Road Type Response Roughness Measurement

System for existent defects localization and quantification

Franciscoco Jose Vea Folch [1], Claudio Mazanet[2], Mireia Ballester Ramos[3], Raul Redón[4], Julia Real Herráiz[5]*

[1, 3]BECSA - Proyectos de Construcción e I+D+i, Ciudad del Transporte II. C/Grecia, 31, 12006 Castellón, Spain

[2, 4 ,5] Institute of Multidisciplinary Mathematics, Polytechnic University of Valencia, Camino de Vera, 46022 Valencia, Spain

Corresponding author. E-mail: fjvea@becsa.es. Telephone: +34 650528210*

1. Introduction

Road maintenance implies a great investment to avoid pavement deterioration of the road network, in order to ensure the safety and welfare of users. One of the most used parameters as criteria in this maintenance is the surface roughness of the pavement, which is usually quantified by IRI – International Roughness Index – ratio. There exist five different ways to measure the roughness of the pavement: i) manual devices, ii) profilometers, iii) light profilometers, iv) inertial profilometers and v) Road Type Response Roughness Measurement Systems (RTRRMS). However, all of mentioned systems require expensive specific equipment which is not always suitable or include defects effects in roughness calculation, which is inaccurate.

Hence, a new alternative is proposed for the analysis of the pavement condition for road maintenance. It is based on RTRRMS techniques but including accelerometers in a conventional vehicle and a GPS locator. That allows to integrate the IRI obtaining models from vertical accelerometer data [1] with the identification of localized defects [2].

2. IRI estimate

The procedure for obtaining the IRI is as follows: The first step of the procedure consists of measuring the elevations of the terrain that allow representing the profile of the road. Next, a mobile average filtering is performed to obtain a corrected profile. Finally, the Quarter Car model is applied to obtain the roughness profile.
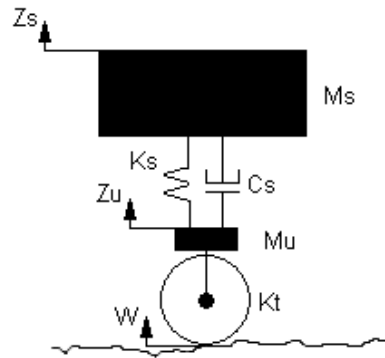
Figure 1: Quarter Car model Outline.

Applying the General Dynamic Equilibrium Equation, the elevation profile of the terrain is obtained:

$$W = Z_u + \frac{M_s \cdot A_s + M_u \cdot A_u}{K_t} \tag{1}$$

Where M_s and M_u are the sprung and unsprung mass; A_s and A_u are the vertical measured accelerations; Z_u is the double integral of vertical acceleration of the vehicle and W is the calculated profile.

A moving average filter is applied to this profile in order to obtain a corrected profile:

$$h_{ps}(i) = \frac{1}{k} \sum_{j=1}^{i+k-1} h_p(j)$$

$$k = \max[1, nint\left(\frac{L_B}{\Delta}\right)] \tag{2}$$

Where h_p is the calculated profile and h_{ps} is the same smoothed profile. L_B being the length of the moving medium, and Δ the longitudinal sampling interval, which in any case should be a maximum of 25mm for class I systems.

A second filter must be applied to the new profile obtained from applying the first filtering, which is based on the quarter car model. The *Quarter Car Model* is defined as a set of masses (sprung mass and unsprung mass) linked to each other through a spring and a linear damper.

This model uses a series of standardized reference parameters, called Golden Quarter Car. These parameters are the following:

$$k_2 = \frac{K_s}{M_s} = 63{,}3 \; ; \quad k_1 = \frac{K_r}{M_s} = 653 \; ; \quad c = \frac{C_s}{M_s} = 6 \; ; \quad \mu = \frac{M_r}{M_s} = 0{,}15 \tag{3}$$

Where Ms is the sprung mass, Mr is the unsprung mass, Ks is the spring constant of the suspension, Kr is the spring constant of the wheel, and Cs is the shock absorber.

The model of the quarter car is described by 4 ordinary differential first-order equations, which can be represented as a matrix as it follows:

$$\dot{x} = Ax + Bh_{ps} \tag{4}$$

Where:

$$x = [z_s, \dot{z_s}, z_r, \dot{z_r}]^T \; ; \quad A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -k_2 & -c & k_2 & c \\ 0 & 0 & 0 & 1 \\ \dfrac{k_2}{\mu} & \dfrac{c}{\mu} & -\dfrac{k_1 + k_2}{\mu} & -\dfrac{c}{\mu} \end{bmatrix} \; ; \quad B = [0,0\,,0,\dfrac{k_1}{\mu}]^T$$

Where hps corresponds to the elevation of the smoothed profile, and zs and zsr correspond, respectively, to the elevations of the sprung and unsprung masses. And x is the matrix of the state variables.

The IRI is an accumulation of the simulation of the movement between the sprung and unsprung mass:

$$IRI = \frac{1}{L} \int_0^{L/V} |\dot{z}_s - \dot{z}_r| \, dt \tag{5}$$

Where L is the normalized length and V is the vehicle speed, which is defined as 80km/h.

3. Fault location

The IRI is one of the most widespread criteria for the identification of the state of the road, because surface damage increases the roughness; however it is not an infallible tool. There could be serious localized damages on the road, and overall roughness criteria would still be met. For this reason, a method to detect localized defects is proposed, based on the analysis by means of filters of the roughness profile.

In order to road damage identification and quantification, the Wavelet Transform is applied to the vertical profile signal:

$$W_f(s, \tau) = \int f(t)\, \psi_{s,\tau}^*(t)\mathrm{dt} \tag{6}$$

By stretching and moving a wavelet, it can be adjusted with the study event, thus allowing finding out its frequency and location in time.

To apply the Wavelet multi-resolution analysis, the roughness profile can be decomposed into components of low (λ) and high (μ) frequency. By the following expressions:

$$\hat{c}_m^0 = \sum_l \tilde{\lambda}_{l-2m} c_l^1 \tag{7}$$

$$\hat{d}_m^0 = \sum_l \tilde{\mu}_{l-2m} c_l^1 \tag{8}$$

According to the 'Theory of the elevation scheme', biorthogonal wavelet type was chosen to decompose the signal in low (λ) and high (μ) frequency components, adding free parameters to the basic functions.

	Low frequency	High frequency
Reconstruction	$h_{k,l} = h_{k,l}^{old} + \sum_m \tilde{s}_{k,m}\, g_{m,l}^{old}$	$g_{m,l} = g_{m,l}^{old}$
Decomposition	$\tilde{h}_{k,l} = \tilde{h}_{k,l}^{old}$	$\tilde{g}_{m,l} = \tilde{g}_{m,l}^{old} + \sum_k \tilde{s}_{k,m}\, \tilde{h}_{k,l}^{old}$

Table 1. Biorthogonal wavelet filters with free parameters in HF component decomposition.

Where $h_{k,l}^{old}$ and $g_{m,l}^{old}$ are the initial filters, and $\tilde{s}_{k,m}$ are the free parameters.

Due to the free parameters, high pass filters will be increased compared to low pass filters in the decomposition process. Applying the filter of high frequency decomposition to the inputs of the roughness profile:

$$d_m^0 = \sum_l (\tilde{g}_{m,l}^{old} + \sum_k \tilde{s}_{k,m} \tilde{h}_{k,l}^{old}) c_l^1 = r_m - \sum_k a_k \tilde{s}_{k,m} \qquad (9)$$

Where r_m corresponds to the high components and a_k to the low frequency components resulting from the decomposition.

Road defects are identified into the HF component. But it is required a training period to calibrate the free parameters. The condition to extract them will be:

$$d_m^0 = r_m - \sum_k a_k \tilde{s}_{k,m} = 0 \qquad (10)$$

For the learning of the free parameters, the algorithm will use 2n signals for its training. Being 2n the number of equations that will be taken from the previous expression. However, the number of unknown variables is 2n+12. This is the reason why the condition that the sum of high-pass decomposition filters must be zero should also be added.

$$\sum_{k=m-n}^{m+n} \tilde{s}_{k,m} = 0 \qquad (10)$$

Once Wavelet filters have been constructed, the wear locations of the surface in the roughness profile can be detected automatically.

Since the free parameters are prepared to cancel the components of the new filters when detecting the characteristics of the training defects, a possible detection strategy would be to find the location that causes d_m^0=0 to be met.

However, this could result in the detection of points where the high-frequency components are close to zero, both in the initial filters and in the new ones. In this way,

to avoid false positives, it has been decided to maximize a new function I_m, which is the difference between the components of both filters.
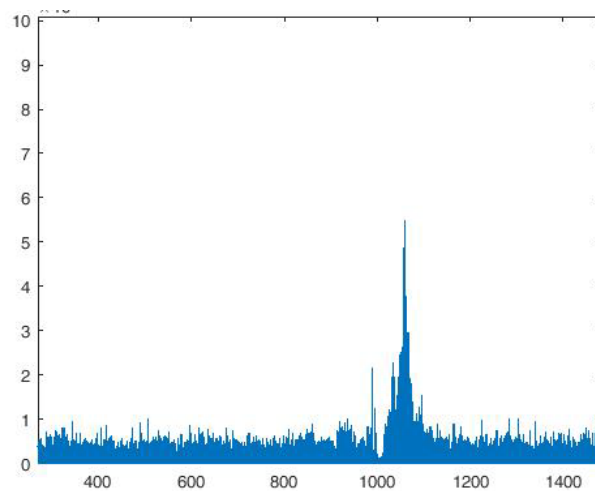
Figure 2: Fault detection function I_m.

When I_m becomes greater than a certain threshold, it will be determined that there are defects in the pavement. Thus, a graphic representation of the value of I_m, can be obtained, where the location of the defects and their severity can be clearly identified.

References

[1]   Du, Y., Liu, C., Wu, D. (2014) Measurement of International Roughness Index by using Z-Axis Accelerometers and GPS.

[2]   Abulizi, N., Kawamura, A., Tomiyama, K., Fujita, S. (2016) Measuring and evaluating of road roughness conditions with a compact road profiler and ArcGIS.

[3]   Sweldens, W. (1998). The Lifting Scheme: a construction of Second Generation Wavelets.

[4]   Shokouhi, P., Gucunski, N., Maher, A. and Zaghloul, S.M. (2005) Wavelet-Based Multiresolution Analysis of Pavement Profiles as a Diagnostic Tool.

Application of an analytical solution based on beams on elastic foundation model for precast railway transition wedge design automatization

Jose Luis Pérez Garnes[1], Miriam Labrado[2], Teresa Real Herraiz[3], Adrián Zornoza[4], Julia Real Herráiz[5]*

[1] Torrescamara, Avenida del puerto, 332, 46024, Valencia Spain

[2, 3, ,4, 5]Institute of MultidisciplinaryMathematics, PolytechnicUniversity of Valencia, Camino de Vera, 46022 Valencia, Spain

*Corresponding author. E-mail: jureaher@upv.es. Telephone: +34 650528210

1. Introduction

Every time a train crosses from an embankment area (soil) to a bridge, tunnel, viaduct or box culvert – and vice versa – the vertical stiffness of support under the track superstructure changes too abruptly. This areas are called transition zones and require much more maintenance than regular track [1] due to the great difference between materials stiffness – concrete/steel vs. embankment ground -. For this reason, a new transition wedge formed by steps made of precast concrete slabs has been conceived.

2. Problem approach

In common transition zones, the progressive compaction of the embankment material with several load cycles during its lifespan produces a progressive differential settlement between the terrain supported section and the part supported by the pre-existent structure.

This is caused by an abrupt change in the vertical stiffness of the support and originates that vertical dynamic forces become increased – up to twice according with [2] -. All of

these repetitive effects result in a potential source of damage on the rolling stock and on the same railway track.

On the other hand, for the execution of traditional granular transition zones a huge amount of resources and time are required – up to 72 days per wedge in order to carry out an adequate compaction and humidity rates – and its final quality is challenging to be checked during the works.

For this reason, a new precast transition wedge has been designed. It consists on stepped solution formed by precast reinforced concrete slabs that is placed under the ballast layer.
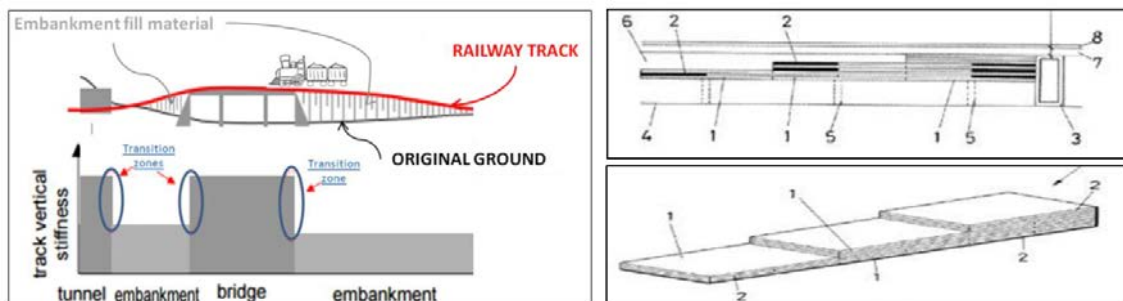
Fig 1: Stiffness abrupt changes in traditional solutions (left) and precast concrete transition wedge purposed (right).

Usually, in Civil Engineering, these problems are solved by means of FEM (Finite Element Models) software tools. In this case, due to PTW application is desired to be possible in any potential scenario – type of railway and infrastructure -, its particular optimized design is not compatible with iterative FEM calculations because of the huge amount of computational consumption time required.

For this reason, the design of this structural element involves the achievement of three different goals: i) the consecution of an accurate – comparable in results to FEM - analytical model which vertical stiffness / longitudinal distance function could be implemented in mentioned optimization problem, ii) automatic optimization algorithm design and iii) result in a PTW that efficiently works for each particular scenario.

In this research, first of those objectives is of concern. The selected analytical model must be capable of supporting stepped stiffness variations along the rail and correlate aleatory configurations of moving loads with vertical deformation in the rail top.

3. Method

For new precast transition wedge (PTW) design, different variables are identified: number of steps, length of steps, depth of the wedge, height of the steps, etc. It means that its optimal design could be approached as a mathematic optimization problem where the minimization functions are: i) stiffness / longitudinal distance gradient and ii) resources used for its construction – volume of the wedge –.

A numerical model based on analytical expressions is chosen to develop the future optimization problem.

Several simplifying hypothesis are assumed to minimize computation times and to soften up the model requirements: i) Load and shape longitudinal symmetry – respect to the rail track axis – ; ii) Linear elasticity of the materials – which fits properly with one single load cycle (vehicle) pass -; iii) Material homogeneity – a 2D vertical middle section trough the rail web is representative of the vertical deformations of interest -.

4. Numerical modelling

Winkler idealization is the first approach to accomplish with model requirements [3]. It consists of the soil as a linear elastic medium where the load-deformation constitutive relation is defined by infinite individual springs of 'k' stiffness constant. In 1946, M. Hetenyi published a new Winkler-based model. In this case, interaction between adjacent springs was modelled by incorporating an elastic flexural deformation only beam/plate.

Flexural rigidity of the beam/plate -Young modulus and bending axis inertia- must be known and incorporated as new parameters to Winkler approach. In addition, this model allows to abruptly change the stiffness among adjacent sections by juggling deformations in their contact edges. Its expression and scheme are:

$$p = kw + D\nabla^4 w \qquad (1)$$
$$D = (E_b h_p^3)/(12(1 - \mu_p)^2) \qquad (2)$$

*where 'D' is the flexural stiffness of the elastic beam/plate, 'p' is the pressure, 'k' is Winkler spring constant and 'w' is vertical deflection, 'E_b' is the Young modulus of the beam, 'h_p' is the beam height and 'μ_p' its respective Poisson ratio.

Beam or Plate with flexural rigidity D
Springs
Rigid Layer
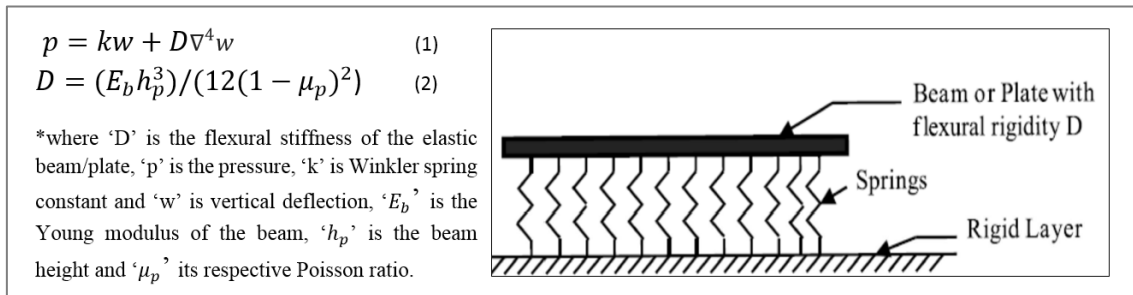
Fig 2: Hetenyi's model constitutive equation and 2D scheme. Source: Self-made from [4]

5. Model implementation and validation

Once the model is chosen, it was implemented in MATLAB by dividing the transition zone in different section lengths, according to its vertical stiffness variations due to the wedge steps. A PWT was reproduced in a 3D numerical FEM model in ANSYS APDL v16 commercial software and its equivalent 2D longitudinal mid-section was also implemented in mentioned MATLAB code in order to compare results and analyze the new model behavior.
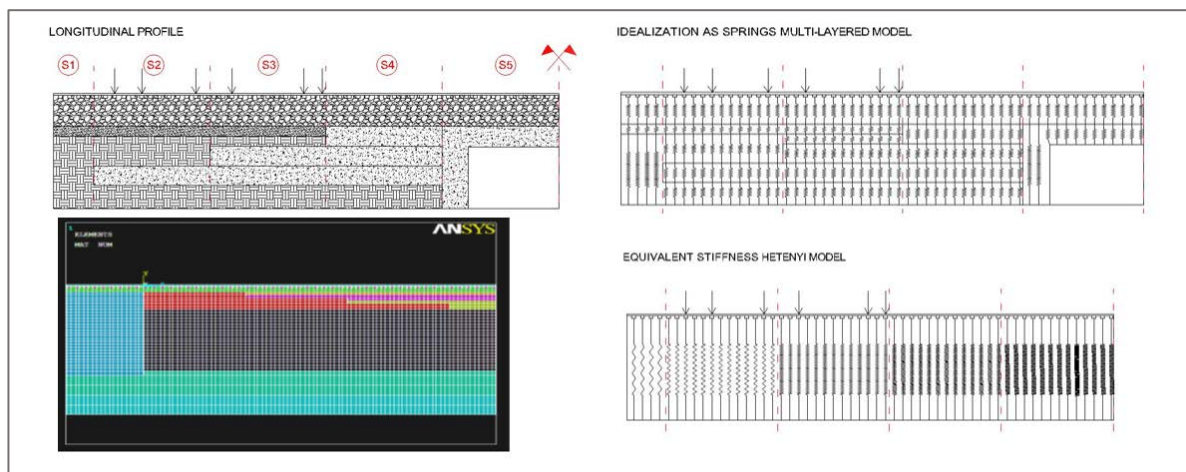
Fig 3: Initial longitudinal transition profile (left, up), ANSYS FEM model (left, down), 2D sprung idealized section (right, up) and Hetenyi 2D model implemented in MATLAB.

Equivalent stiffness of each layer was calculated attending to Hooke's Elasticity relationship with Winkler constitutive equation (Eq. 3) and in series spring adding (Eq. 4):

$$k_{eq,i} = \frac{E_i \cdot A_l}{h_i} \qquad (3) \qquad ; \qquad K_{eq} = \frac{1}{(1/k_{eq,i}) + (1/k_{eq,j}) + \cdots + (1/k_{eq,n})} \qquad (4)$$

Properties of C30 concrete (according to EN 1992:1-1), UIC-54 for rail and ADIF transition zones normative specifications for embankment (ADIF PGP-2008) were used in order to simulate as accurately as possible a standard transition zone. Results profile for vertical displacement on the rail top under a 9 tn load – reproducing a typical axle load – are:
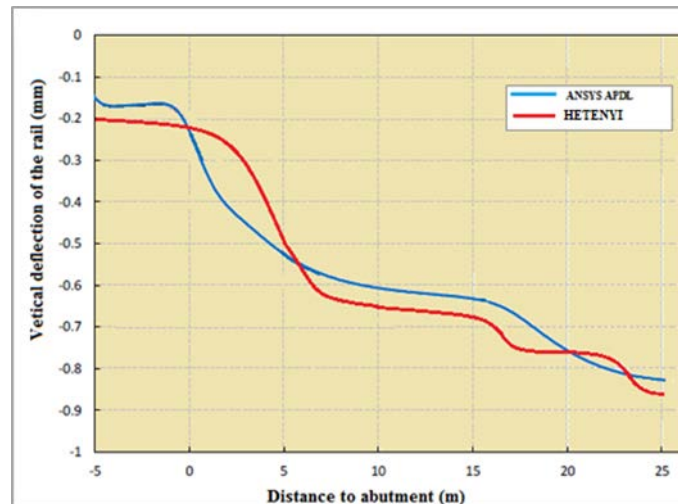
Fig 4: Vertical deflection in longitudinal rail mid-section profile. ANSYS results vs HETENYI multi-layered model.

Results obtained required a total time of computation of 24 seconds in MATLAB application for Hetenyi's model and greater than 3 hours for ANSYS steady state simulation with only an average deviation of 6.8% in vertical deflection, which makes approachable the problem of design optimization in future phases.

6. Conclusions

In this research, a new method based on analytical expressions has been defined to calculate the railway transition element.

Accurate results, with an average deviation lower than 7% from 3D FEM models has been achieved.

Only 24 seconds per calculation are required carry out a simulation with the proposed method while more than 3 hours were required for the equivalent 3D FEM steady-state analysis – by using a conventional PC [Intel Core Duo 2.93 GHz and 4.00 GB RAM specifications], which means a substantial reduction in computation time.

The analytical expressions set chosen together with volume of material needed estimation allow developing the desired future automatic design algorithm.

References

[1]   Paixao, A., Fortunato, E., & Calçada, R. (2013). Design and construction of backfills for railway track transition zones. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit.

[2]   Sadeghi, J. and Shoja, S. (2015) Influences of track and rolling stock parameters on the railway load amplification factor. Institution of Mechanical Engineers. Journal of Rail and Rapid Transit.

[3]   Chandra, S. (2014) Modelling of Soil behaviour. Indian Technology Institute of Kampur.

[4]   Lee, J.K., Jeong, S., Lee, J. (2014) Natural frequencies for flexural and torisional vibrations of beams on Pasternak foundation. Soils and Foundations, Vol. 54. Pp: 1202-1211.

Development of an innovative wheel damage detection system based on track vibration response on frequency domain

Ramón Auñon[1], Beatriz Baydal Giner[2], Sheila Nuñez[3], Julia Real Herráiz[4]*

[1] AMINSA, Conde Altea, 1, 46005, Valencia Spain

[2, 3, 4]Institute of MultidisciplinaryMathematics, PolytechnicUniversity of Valencia, Camino de Vera, 46022 Valencia, Spain

*Corresponding author. E-mail: jureaher@upv.es. Telephone: +34 650528210

1. Introduction

More than 4000 M€ per year are spent in European rail infrastructure maintenance [1] whose 20% is only invested in repairing railway track damages caused by defects on train wheels [2]. The reason is that current Wheel Impact Detectors (WID) cannot accurately predict the occurrence of the failure. In addition, those systems do not indicate which and when maintenance operations must take place.

The solution proposed herein is an automatable algorithm that allows, through strategical disposition of inertial sensors (accelerometers) in the railway track, to detect subtle changes in wheel tread and become a useful tool for infrastructure managers for future maintenance operations prediction.

2. Wheel pathologies

First of all, it is important to know the different pathologies that a wheel can present. In general, depending on their morphology, and although they have different origins, can be classified as spalling, wheel flats, corrugation or generalized wear.

If a sudden braking occurs while a train moves along the track owing to an incidence, an instant blockage of the wheels could happen, this causes the wheel slide over the track. During the blockage, the rolling stock is modified producing a chamfer on its running surface.

A spalling defect is also known as a cavity and refers to the loss of material. This type of defect can appear in only one point defect localization, although it is common to appear as a group of points. This is what is known as a wheel flat.

Corrugation is characterized by wavy geometry; this pathology can cause a loss of circularity or wavy wear. The last pathology type is generalized wear defect which refers to the equal loss of wheel radius.

Of these, the most common defect and which can have the most influence is the wheel flat defect. Additionally, the defects type spalling have a similar behavior and the other kinds of defects have a much lower importance in the deterioration of the track. For these reasons this research project will focus on the detection of pathologies type wheel flats.

3. Theoretical model

All these pathologies create situations where the tracks suffer high intensity impacts or forces that result in permanent damage to the rails. Since these forces are applied vertically to the rails, they create accelerations in the same axis that can be measured by the means of sensors, such as accelerometers. These accelerations can be later analysed mathematically to detect the presence or the starting of these pathologies with high accuracy.

The objective of the algorithm is to analyze the vibrations signals caused by the train passing to the early detection of wheel pathologies. For the study of vibratory patterns, a

study of the signal in the frequency domain can be carried out using the Fourier transform or a wavelet filtering.

For this purpose, a research about vibration patterns using Wavelet Transform has been carried out. The Wavelet Transform (WT) allows to analyse sudden transient signals in frequency domain without losing time domain information – as happens with Fourier Transform -. In addition, it is possible to divide the whole frequency broadband registered in ranges and localize in which ones the defects are located.

In order to know the vibratory frequencies that characterize a wheel flat pathology and therefore the decomposition level, a model of the theoretical vibratory patterns has been made. To this end a model for predicting vibrations in response to different stresses (numerical model of finite elements - FEM) and a model for calculating the temporal history of the contact force for both a smooth wheel without defects and for a wheel with defects (analytical models) have been developed.

In the next figure can be seen the wavelet transform obtained for a vibratory signal of a health wheel.
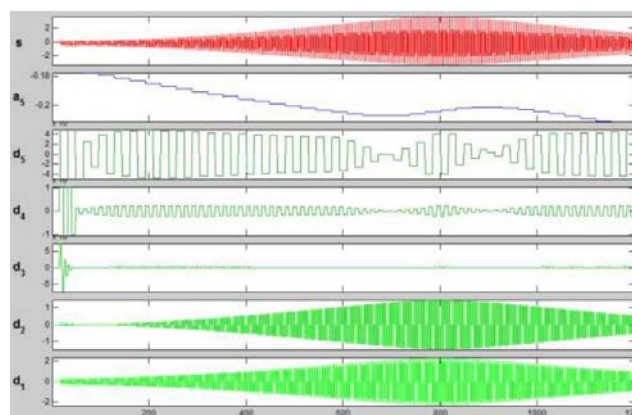
Fig. 1: *Wavelet transform for a vibratory signal of a health wheel*

The wavelet on the signal without defects we observe how in the details of level 3 and 4 does not see any kind of disturbance. Looking these levels details for the signal caused by damage wheel:
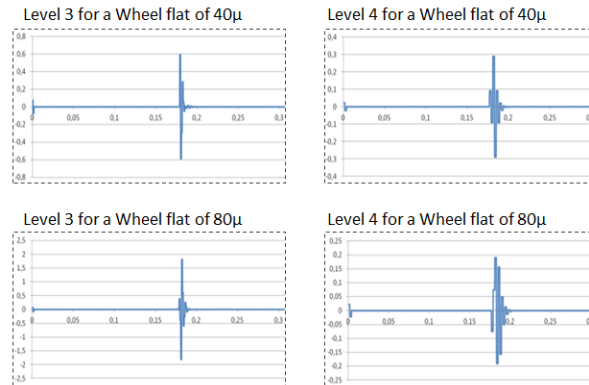
Fig. 2: *Details of level 3 (left) and level 4 (right) for a Wheel flat of 40μ (up) and 80μ (down)*

Looking the last figures it can be concluded that the existence of wheel damage is defined by the existence of peaks in levels 3 (frequencies between 250 and 500 Hz) and 4 (frequencies between 125 and 250 Hz).

4. Signal analysis procedure

The aim of this research is to identify and locate wheel flats by analysing a Wavelet Transformed acceleration signal in order to determine if any corrective action is necessary and where it should be applied – in axis 'n' of bogie 'm' -. This type of defect is very usual and could produce stresses up to 4 times greater than in healthy wheels.

Haar wavelet type was chosen (step function), because other typologies could produce signal distortion after its treatment.

First of all, it is necessary to determine time lapse between consecutive bogies. It is possible by peak to peak time lapse analysing in original signal (t_b). If axle separation is known (d_b), it is possible to determine vehicle velocity as: $v = d_b/t_b$.

For time lapse between the axles detection, a type of wavelet filtering has been used to detect abrupt changes in the frequency of the signal and after them a simple peak detection algorithm.
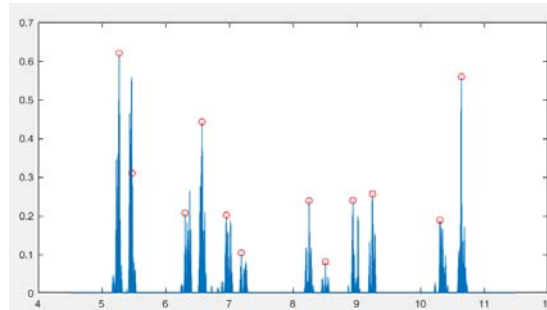
Fig. 3: Axle detection with a vibratory signal of 12 axles train.

In order to detect wheel pathologies, a damage detection criteria based on acceleration amplitude was needed. For this purpose and, due to vertical acceleration registers are closely related with vehicle velocity, a statistical analysis based on box-whisker and normal probability distribution (NPD) was performed:
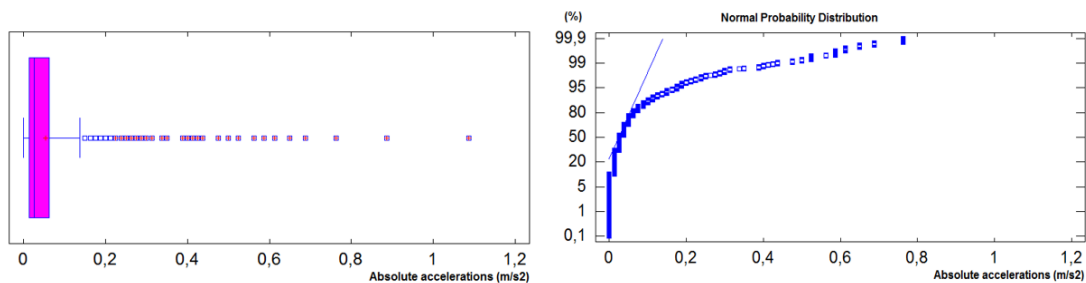
Fig. 4: Statistical analysis for wavelet decomposed signal of one accelerometer register. Box-whisker plot (left) and NPD (right).
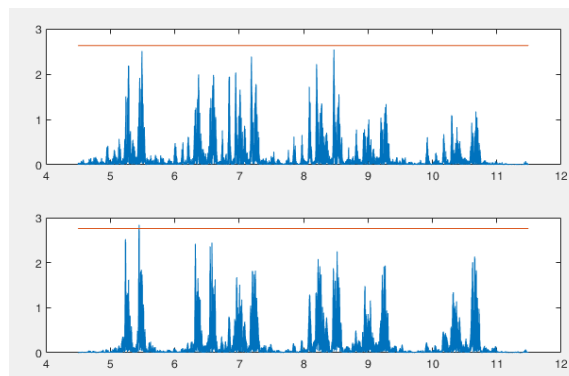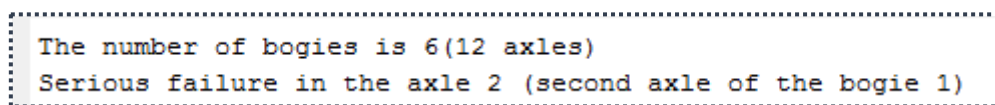
Fig. 5: Analysis of levels 3 (up) and 4 (down) of vibratory signal of a train with a wheel flat on the second axle.

As all of registers were analysed, it was determined that wheel flat identification criteria will be for those values which exceed in, at least, a 50% the obtained for 99 percentile of each sample.

5. Results and conclusions

This made possible to correlate WT obtained signal in decomposition levels of interest (3rd and 4th) with potential defects appearance in time domain and, consequently, to deduce which bogies or axles need maintenance actions.

```
The number of bogies is 6(12 axles)
Serious failure in the axle 2 (second axle of the bogie 1)
```

Fig. 6: Conclusion of the algorithm with the signal of the figure 5.

Applying the method with vibratory patterns caused by the train passing with validated data on the parked vehicle, it can be concluded that detail of level 4 provides better results, not showing false positives in which it detects a possible failure when it is not. Then it can be used to detect current failures.

The detail of level 3 is more precise and would detect as pathology some irregularities of little significance, such as marks that are hardly noticeable to the touch. In this way it can be used to detect possible future failures.

References

[1]   Real, T., Marzal, S., Morales, S., Real, J. I. (2016) Wheel-rail irregularities dection System using analytical and numerical methods. Modelling for Engineering & Human Behaviour 2016.

[2]   Wavelet signal processing applied to railway wheelflat detection. Belotti, V., Crenna, F., Michelini, R.C., Rossi, G.B. (2003) University of Genova. XVII IMEKO World Congress, Metrology in the 3rd Millennium.

Mathematical characterization of liquefaction phaenomena for structure foundation monitoring

Piter Moscoso Godoy[1], Rubén Sancho[2], Ernesto Colomer[3], Julia Real Herráiz[4]*

[1] DGOP, MOP, Morandé 59, Santiago de Chile, Chile

[2, 3, 4]Institute of MultidisciplinaryMathematics, PolytechnicUniversity of Valencia, Camino de Vera, 46022 Valencia, Spain

*Corresponding author. E-mail: pitermoscoso@yahoo.es.Telephone: +34 650 52 82 10

1. Introduction

Every year, important structures whose foundation is supported by granular or low-cohesive soils in wet environments have to deal with one of the most unpredictable and harmful events, the soil liquefaction. This phaenomenon is continuously causing severe material, human and economic damages, especially in seismic areas.

Herein, a mathematical characterization of liquefaction is exposed. The objective of this research is to determine distinctive patterns which could be mathematically expressed. In this way, this phaenomena could be predicted, simulated or detected automatically by real-time monitoring systems. For this purpose, liquefaction is analysed herein from structural and terrain effects.

2. Mathemathical characterization

Attending to Terzaghi's Principle (1925), in a static equilibrium situation, granular soils behaviour is determined by: $\quad \sigma = \sigma' + U$ (1)

where "σ" is the total pressure; " σ' " is the solid particle contact pressure and "U" is the

pore water pressure. Under sudden and high intensity shakings, as produced during an earthquake, the liquid phase of the soil – pore water – is ejected into a particle rearrangement process, causing a rapid increase of water pressure ($\Delta U >> 0$) and, as a consequence, the loss of foundation stability ($\Delta\sigma' << 0$).

Additionally, a foundation stability loss implies a decrease on stiffness on structural support systems. Mathematically it means a substantial diminution in stiffness matrix in dynamic equilibrium equation – based on Newton's Second Law - (Ec. 2).

$$[M]\{\ddot{x}\} + [C]\{\dot{x}\} + [K]\{x\} = \{F(t)\} \tag{2}$$

$$[M]\{\ddot{x}\} + [K]\{x\} = \{0\} \tag{3}$$

where [M] , [C] and [K] are mass, damping and stiffness matrixes respectively; $\ddot{x}, \dot{x}, x$ are acceleration, velocity and displacement vectors and $\{F(t)\}$ resultant forces vector.

A particular case of this equation, where no damping is considered ([C]=0) and resultant forces are zeroed ($\{F(t)\} = 0$) – known as 'free vibration' conditions – allows to obtain natural vibration frequencies and modes (Ec. 3) as an eigenvalues/eigenvectors problem – which is known as modal analysis of a structure -:

From both equations, it is deduced that if stiffness matrix decreases, natural frequencies will be affected, concretely, a diminution up to 50% of initial value could be achieved [1]. In addition, damping ratio could increase, in conditions of full liquefaction of soil, up to 20% - from usual values of 2-5% - [1]. This variation additionally reduces natural

frequencies as: $$f_{undamped} = \frac{f_{damped}}{\sqrt{1-\zeta}} \tag{4}$$

Regarding stiffness loss, it is a complex interaction among air, water and granular phases of soil. Deep non-linear processes are involved in this event and it hinders its mathematical representation. For this reason, it is possible to distinguish three different phases during liquefaction [1]:

i) No liquefaction – for low range excitations ($\Delta U < 0.1\ U_{max}$). Soil's bearing capacity variations do not endanger structural integrity. Expectable natural frequency decrease up to 15%.

ii) Partial liquefaction – for mid-range excitations – ($0.1\ U_{max} < \Delta U < 0.4\ U_{max}$). Soil's bearing capacity loss is significant. Structural damages appear. Instability could produce displacements from original structure position. Expectable natural frequency decrease up to 50%.

iii) Transient and full liquefaction – for high intensity excitations – ($0.4\ U_{max} < \Delta U$; $\sigma' \approx 0$). Risk of collapse. Serious or irreparable damages. Expectable natural frequency decrease up to 60%.

3. Method: numerical analysis

As it was aforementioned, there are two recognizable aspects that characterize liquefaction: increase of pore pressure (ΔU) and soil stiffness decrement.

Regarding ΔU, it is only relevant for liquefaction detection. Field disposition of piezometers or other adequate pressure sensors may constitute an effective alert system of liquefaction is taking place.

Stiffness decrement could be analysed by different ways. For stability loss scenarios simulation, 3D FEM models where soil is modelled as a multi spring-dashpot system is one of the most efficient options for computational analysing.

According with [2], during liquefaction, the ideal value of the spring stiffness diminish by a 'α' or 'p-multiplier' empirical factor. For an accurate simulation, this factor is only

applied to those springs which are under liquefaction conditions. This technique is used in huge offshore structures design, among others [3].
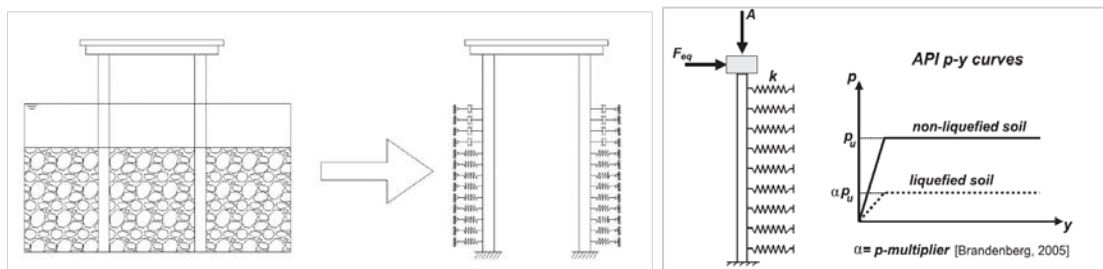
Fig. 1: Example of a pier foundation interaction by spring-dashpots system (left). Spring discretization and 'p-multiplier' effect on spring stiffness. Source: own elaboration & [2]

4. Evaluation

Extracting eigenvalues from Frequency Domain Decomposed (FDD) vibration signal shows the natural frequency values and its possible time variation. Thus, it is possible to detect if potential damage caused by liquefaction is being produced and estimate severity of liquefaction attending to mentioned criteria.

In addition, disposing inertial and pressure sensors allow to know not only if damage is producing (natural frequency decrement) but which part of the structure could be more degraded (where maximum pore pressure values have been reached).

Moreover, all of this procedure could be automated by means of an algorithm definition where aforementioned behaviour patterns would be identified.

In this algorithm, Power Spectrum Density (PSD) decomposition is carried out in order to identify the eigenvalues by peak picking techniques. Previously, the noise of the signal should be reduced / removed by means of a Moving Average Filter.

A Savitzky-Golay filter is also applied in order to enhance the peack picking identification procedure.

Once the peaks have been identified, natural frequencies are known in current conditions, without the necessity of artificial means of excitation (Modal Operational Analysis).

If only noise is detected during this treatment, the curve is discarded and not included into the study of natural frequencies' variation.

The scheme of this procedure is shown below:
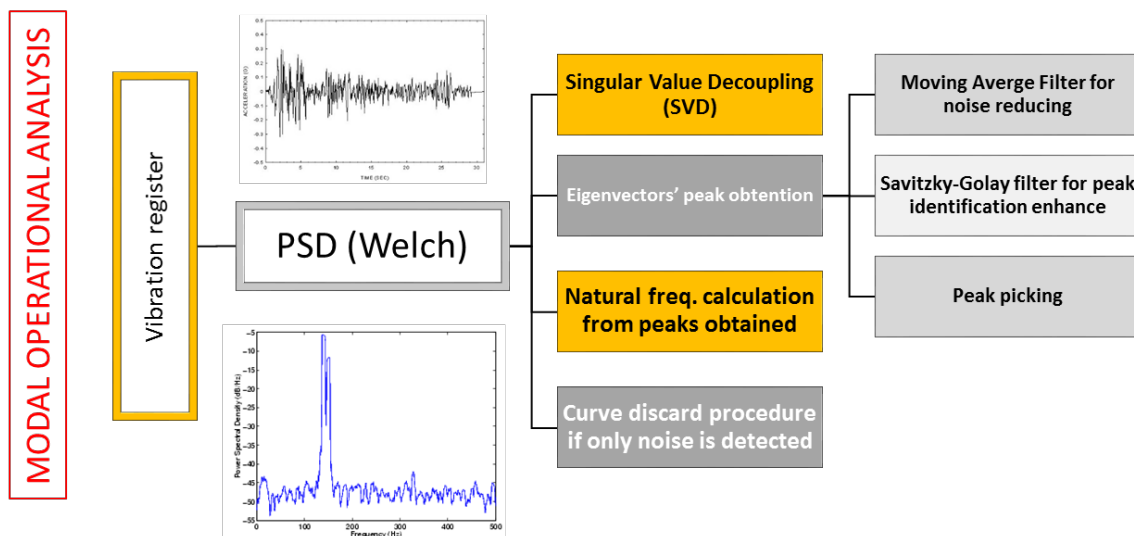
Fig. 2: Scheme of the automatic data treatment algorithm proposed. Source: own elaboration

All of the signal analysis tools mentioned are implemented in commercial software MATLAB, what makes easier its constitution and use.

5. Conclusions

In this research, a new procedure for liquefaction monitoring based on structural and support response parameters control has been defined.

This procedure is based on FDD and eigenvalue extraction of the vibration response of the structure and also, by means of the pore pressure analysis in its support. This double approach allows to:

- Detect the liquefaction in real time. Pore pressure variations in different points of the structure support make possible to identify an estimate the appearance of this event and its severity.

- Predict damage severity. By means of pore pressure register & previous FEM simulations, it is possible to correlate the loss of support stiffness with the variation of the stability in the structure.

- Use commercial software and measurement devices to implement and carry out this process.

In addition, the applicability of this methodology does not depend on the type of structure and is a non-intrusive technique.

References

[1]   Lombardi, D. & Bhattacharya, S. (2013). Modal analysis of pile-supported structures during seismic liquefaction.

[2]   Brandenberg, S.J. (2002) Behavior of Pile Foundations in Liquefied and Laterally Spreading Ground.

[3]   American Petroleum Institute (API). (2000) Recommended Practice for Planning, Designing and Constructing Fixed Offshore Platforms. Recommended Practice 2A-WSD.

Neural Network application for concrete compression strength evolution prediction

Teresa Real Herraiz[1], Miriam Labrado Palomo[2], Beatriz Baydal Giner[3], Julia Real Herráiz[4]*

[1, 2, 3, 4]*Institute of MultidisciplinaryMathematics, PolytechnicUniversity of Valencia, Camino de Vera, 46022 Valencia, Spain*

*Corresponding author. E-mail: tereaher@upv.es. Telephone: +34 650528210

1. Introduction

One of the most challenging aspects in 'in-situ' elaborated concrete is the inability to accurately obtain the maximum compression strength during its beginnings. This induces several uncertainty when shuttering removal, when is necessary to apply special curing cares and other common situations in concrete element manufacturing.

The solution proposed herein aims to determine and control in a rapid and non-destructive way the evolution of concrete compression strength by means of electrical resistivity monitoring registered at work – real time data monitoring -. It also allows to predict which will be the maximum reached resistance after its hardening.

2. Neural Network

First of all, a study was made of the factors that influence the resistivity-compression resistance of the concrete. By means of an ANOVA analysis with real concrete specimens, the most influential factors were selected. It can be differentiated between the factors related to the dosage of the concrete and the factors related to the curing conditions.

Regarding to the dosage factors: i) cement type, ii) cement portion, iii) water/cement ratio, iv) percentage of fines in the sand, v) maximum diameter of the aggregate, vi) consistency. These properties a priori are known and controllable, then it can be use as inputs for the prediction algorithm. Due to the amount of factors, an Artificial Neural Network (ANN) was chosen to develop the prediction algorithm.

On the other hand, the curing conditions are not in any sense controllable: vii) curing technique, viii) curing time, ix) temperature, x) humidity. For this reason, to take them into account they will be used as corrections of the real time resistivity measurement.

The ANN will be used to predict the resistivity and compression resistance curves of a concrete with known dosing data. The data of the necessary curves are 1, 3, 7, 14 and 28 days of maturation. Therefore the neural network will have 6 input data (mentioned factors 'i' to 'iv') and 10 output data (5 for resistivity and 5 for resistance). For this purpose, the number of hidden nodes (9) has been determined according to next equation:

$$hidden\ nodes = \left(\frac{inputs\ number\ +\ outputs\ number}{2}\right) + 1 = 9 \qquad (1)$$

ANN training phase has been divided into three different stages – training, validation, test – according to the percentage of aleatory data used in each one - [70% - 20% -10%] – in order to avoid any processing bias.

Based on this, the performance of different neural networks has been studied, testing with different number of hidden neurons, with only neural network for resistivity and resistance, with two separate neural networks. Finally, attending to the validation error, the better results have been obtained using a single neural network with nine hidden neurons.

For ANN working, a -1 to 1 normalization process of input data is carried out:

$$Xn = \frac{X - X\,min}{X\,max - X\,min} \cdot (1 - (-1)) + 1 \tag{2}$$

where 'Xmax' and 'Xmin' are the maximum and minimum potential values of each variable.

The ANN allows predicting the concrete curves from the concrete dosing data, both in terms of compression strength and resistivity over time, as well as the relationship between them.

In this way, for its use it is necessary to normalize the input data first. With the neural network predict the normalized output data and proceed to the denormalization of them. Finally, to have intermediate data in the curves and carry out a continuous study of the concrete resistance to compression evolution during its maturation, it was decided to perform a linear interpolation between the predicted days.
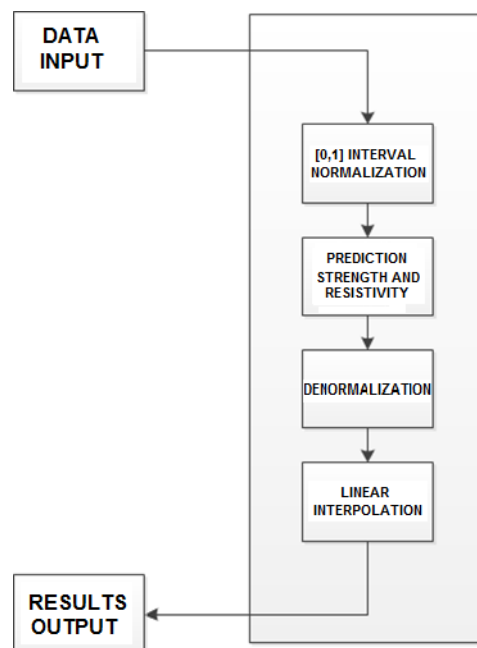
Fig. 1: Algorithm working scheme.

3. Prediction algorithm

In addition, a correction algorithm was also implemented to transfer field data to a standardised potential situation of resistance evolution according to normative - nCh 1037 in this case -.Then with the algorithm, it will be possible know the state at the time of the measurement from the data measure in situ of the concrete.

From [1] and [2], it is possible to obtain the relationship between resistivity and temperature; this relationship can be used to solve the problem that the resistivity measure does not correspond to a state under standardized conditions:

$$\rho = \frac{e^{-E/RT}}{e^{-E/RT_{ref}}} \tag{3}$$

where E is the activation energy (33258 J/mol); R is the gas constant (8.314 J/mol·K) and T the absolute temperature in K.
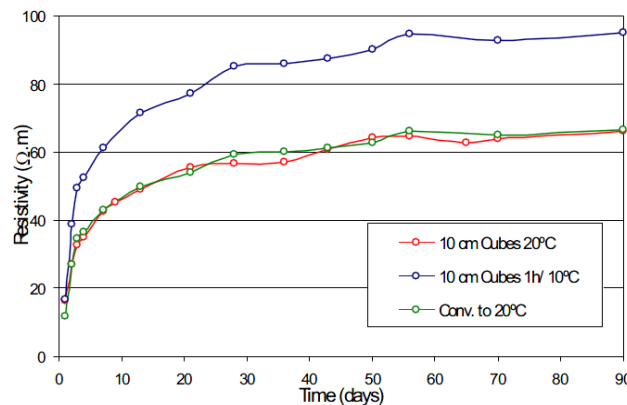
Fig. 2: *Resistivity data corrected by temperature factor.* [2]

Comparing the curves of the resistivity measured at 20 degrees Celsius with the measurement at 10 degrees Celsius the difference between the two is evident. Multiplying by the correction factor both curves overlap.

Once this corrected value is known, it is possible to obtain, from the resistence (MPa) – resistivity (Ohm·m) graphic the real status of concrete strength.
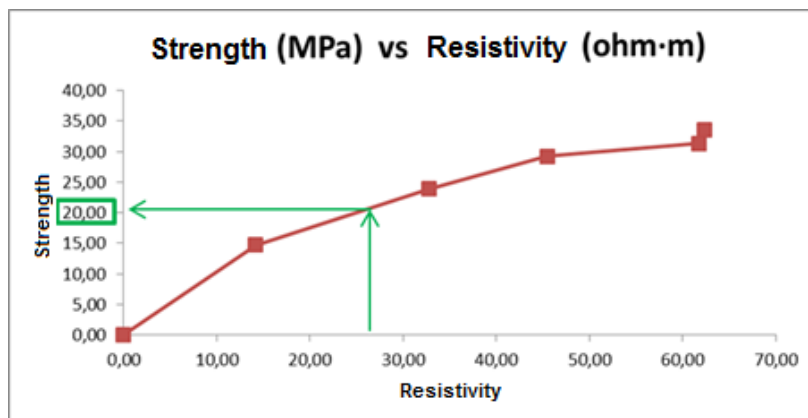
Fig. 3: Strength – Resistivity curve.

After that, it is possible to determine the time lapse between both situations (real and theoretical) under the standardised conditions.
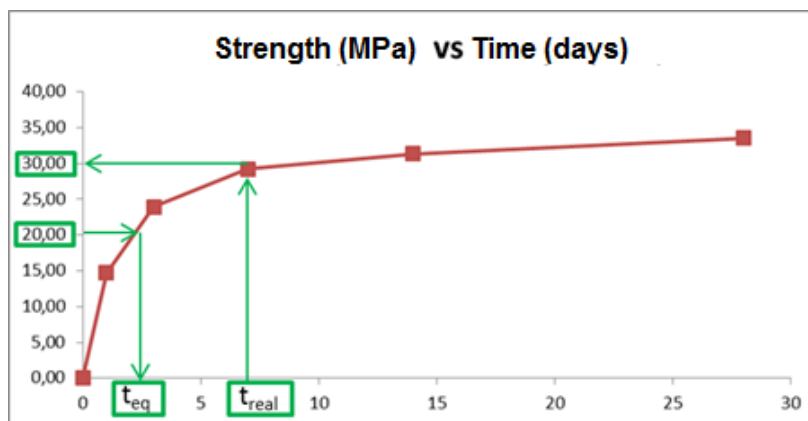
Fig. 4: Equivalent standardised time vs real measured time. Delay from ideal conditions.

As it could be appreciated in last figure, processed example shows a maturity delay of 5 days if its fabrication and curing conditions were as normative predictions indicate. This evidences that possibly, additional cares are needed. Opposite situation, where $t_{real} < t_{eq}$ is also possible and it would indicate that curing is developing in more favourable conditions than predicted by normative.

4. Results and conclusions

The correct work of the developed algorithm has been verified. At first the concrete curves have been obtained from the input data relative to the dosage.
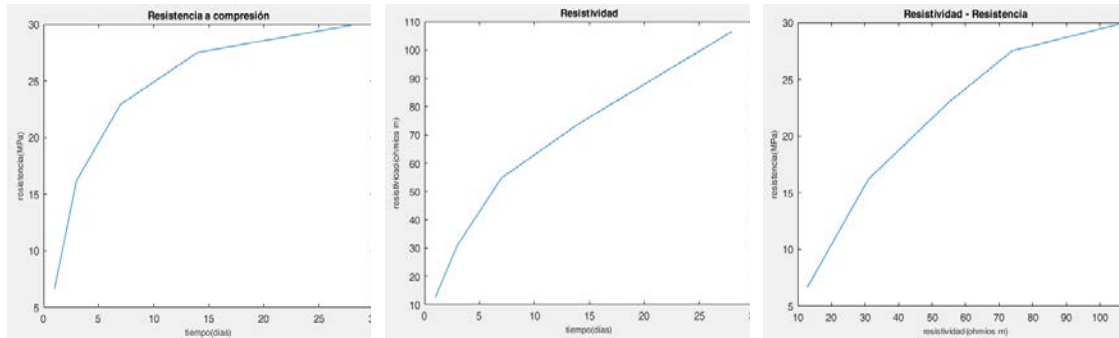
Fig. 5. Predicted curves for a known concrete. Concrete strength over time, resistivity over time and relationship between resistivity and resistance.

The curves obtained with the ANN were compared with known real curves, obtaining an error of 12%. This error is assumable because the resistivity measurement error is of the same order of magnitude.

Once the concrete curves have been obtained, the correct performance of the prediction algorithm has been verified. This algorithm works with the concrete curves, the data taken in situ and the temperature. And it is able to calculate the resistivity, predict the compression strength and calculate the theoretical time if the concrete had been in standard conditions.

References

[1] Ferreira, R.M., Jalali, S. (2010). NDT measurements for the prediction of 28-day compressive strength. NDT&E International, Vol. 43.

[2] Ferreira, R.M., Jalali, S. (2010).Quality control based on electrical resistivity measurements. ESCS-2006. European Symposium on Service Life and Serviceability of Concrete Structures. Helsinki (Finland)

Numerical simulation of lateral railway dynamic effects for a new stabilizer sleeper design

Francisco J. Fernández Martínez [1], Teresa Real Herraiz[2], Adrian Zornoza[3], Julia Real Herráiz[4]*

[1] ACCISA, C/ Alfonso Álvarez Miranda 11 C, 39408 Los Corrales de Buelna, Spain

[2, 3, 4]Institute of Multidisciplinary Mathematics, Polytechnic University of Valencia, Camino de Vera, 46022 Valencia, Spain

Corresponding author. E-mail: jureaher@upv.es. Telephone: +34 650528210

1. Introduction

According to future trends predictions in 2011 Transport White Paper, freight and passenger in Europe will have significant increases in demand, velocity and tonnage in near future. This means that robust and capable railway infrastructures will be needed. One of the main challenges that these infrastructures have to deal with is the misalignment defects due to lateral dynamic effects – forces and deformations -, which suppose more than 26% of track maintenance costs [1]. To deal with this, a new concrete sleeper equipped with additional stabilizer appendixes has been conceived.
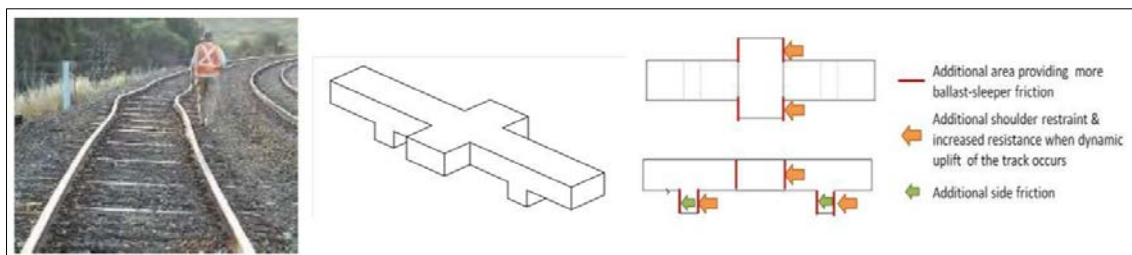
Fig. 1: *Misalignment defect or buckle (left) and new stabilizer sleeper solution proposed.*

2. Method

To achieve an adequate design, a 3D numerical FEM (Finite Element Model) has been implemented in ANSYS LS DYNA v14 commercial software.

One of the most relevant causes of lateral misalignment is rail dilation due to high increase of temperatures – especially in continuous welded rails -. Mathematically, modelling this phaenomena implies serious challenges that are not obvious to save.

In a conventional basic static model, displaced geometry is obtained by a linear relation between loads – thermal in this case – and structural stiffness – determined by material and shape properties. However, in this case, interaction between rail and ballast, non-linear geometrical effects – buckling -, and sliding of some elements – rail / support pads and sleepers / ballast – presumable have serious importance on lateral response.

For this reason, in mentioned 3D FEM model, a straight-curve-straight section of conventional metric wide ballasted track has been reproduced.
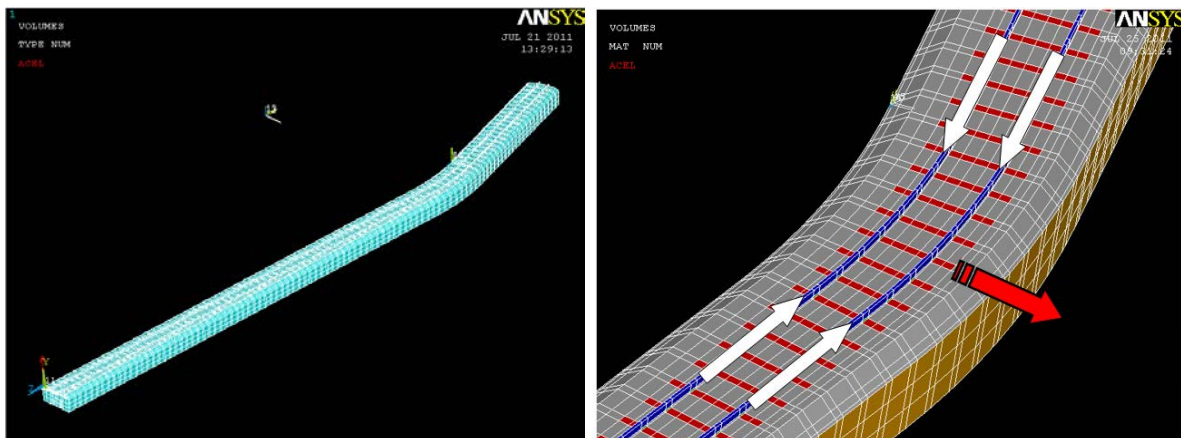
Fig. 1: Numerical modelization of a curved section of ballasted railway. Thermal effects on buckling.

Rail has been modelled as linear elastic – determined only by Young modulus (E) and Poisson coefficient (ν) - material with rectangular shape and equivalent inertia in vertical and cross axis inertia than a 45 kg/m Vignole type has. Oak wood sleepers of 1.9x0.22x0.13 meters have been also modelled as linear elastic. Ballast and foundation layers have been modelled as elastoplastic Drucker-Prager materials – which additionally require internal friction angle (Φ) and cohesion definition (c) -.
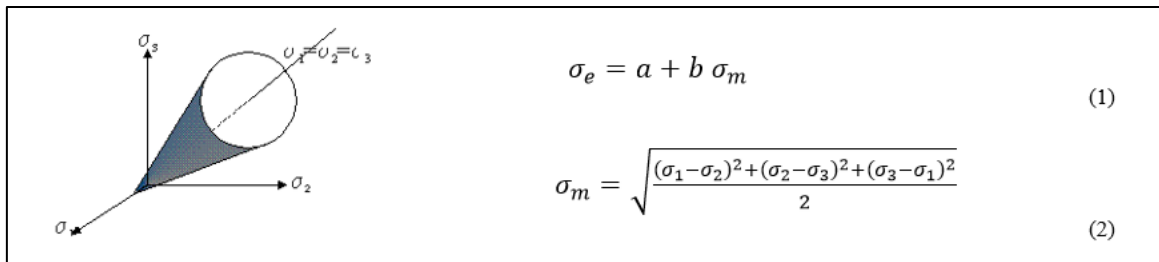
Fig. 2: Drucker-Prager elastoplastic criteria related with average stress.

where 'a' and 'b' are characteristic parameters of the material; 'σ_e' is the Drucker-Prager yielding surface; '$\sigma_{1,2,3}$' are 3D principal stresses and 'σ_m' the average hydrostatic stress.

Thermal supporting elements have been used to rail modelling in order to accurately reproduce dilation process.

Regarding boundary conditions, displacement has been constrained at limit surfaces in its perpendicular direction. Rail longitudinal displacement has been released from support pads but coincident with them in vertical and cross direction. 'Contact Elements', which are non-linear 2D contour elements that allow displacement among different deformable volumes have been disposed in sleeper surfaces in order to allow its relative displacement. These elements follow the Hertzian Contact Theory. The displaced geometry between the 'contact elements' requires iterative calculation for its equilibrium solution. A maximum penetration depth of 0.1 mm was allowed as a compromise solution between accuracy and convergence of the solution.
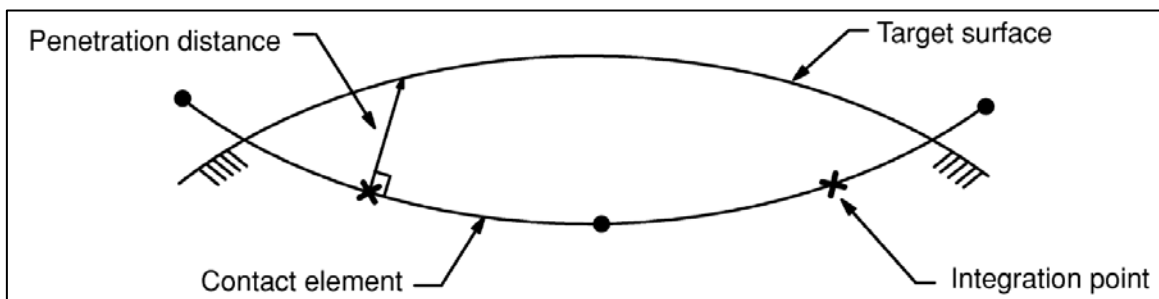
Fig. 3: Contact elements working scheme in ANSYS. Source: ANSYS APDL Element Reference Manual.

Ballast Young modulus and friction angle; foundation Young modulus and sleeper-ballast friction coefficient for contact elements are defined through carrying out a model calibration with real field test data.

Because the original geometry of the model should be distorted because of the combination of different non-linear processes, a large displacement steady state analysis has been performed. The Newton-Rhapson algorithm was used to determinate the final shape of the railway section.

Additionally, due to the high level of non-linear assumptions required, the Arc-Length method was included during solution calculation.

3. Results and model validation

Validation of the model was carried out by measuring lateral displacement on rail web in the mid cross-section of the curve on a real test bank – in Solares, Spain - with similar geometry and properties as the model exposed, by increasing progressively the temperature of rails from 8 to 38 ºC.

Fig. 4: 3D FEM lateral displacement results (left), real test (mid) and results comparison for ΔT and validation (right).

Once the model was validated and calibrated, new stabilized design concrete sleepers were introduced and lateral displacement for identical temperature increases were compared with different modular shapes: i) H-shaped, ii) Cross-shaped, iii) with vertical heels and iv) Cross-shaped combined with heels.
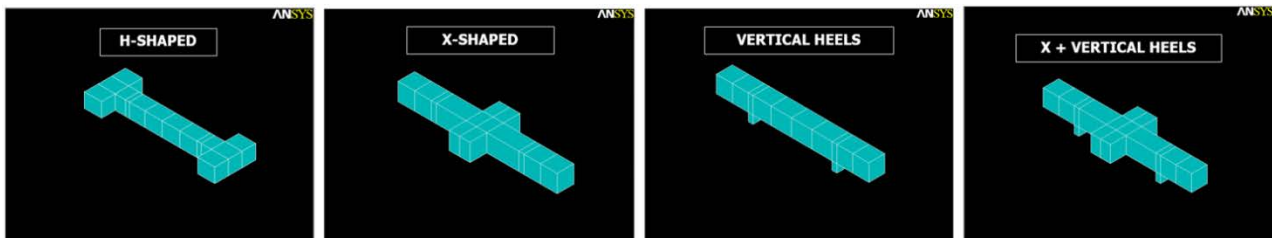
***Fig. 5:** Different shape alternatives tested by numerical simulation.*

Results of numerical modelling revealed that 'X+Vertical Heels' alternative had a better response than the others and increases lateral resistance respect to wood sleepers up to 91% more.

According to [2 – 4], other commercial solutions – winged sleepers, frictional sleepers, framed sleepers, etc. - only achieve to increase up to 59% of lateral resistance but being, at least, 21% more expensive in terms of track maintenance and renewal cost than the proposed one .

4. Conclusions

In this research, the numerical 3D FEM modelling technique has been applied in order to design a new sleeper which improves the lateral stability of the railway. Lateral displacement of the track elements over the support layers involves a challenging task from a mathematical convergence perspective. For its convergence, Newton-Rhapson algorithm combined with the Arc-Length criteria have been used during calculation process. Soil layers' plasticity has been also included in the model by following Drucker-Prager non-linear model. Thus, a steady-state analysis with non-linear geometry - large displacements – and temperature effects was performed.

A real field data test measurement in Solares (Spain) provided the necessary information for the calibration of the model and the order of magnitude of expected results, which was necessary to set the parameters of mentioned resolution algorithms in order to convergence assurance.

Once the model had been calibrated, four different design alternatives of the sleepers were simulated. Those ones equipped with vertical heels shown the best response in terms of lateral stability – minimum displacement -.

Up to a 91% gain in lateral resistance respect to conventional sleepers was estimated in the model results. On the other hand, pre-existent commercial solutions are only able to achieve a 59% more lateral resistance than conventional ones. All of this by needing an average cost increment of 21% more than the solution proposed herein.

References

[1]   Gong (2016). The interaction between railway vehicle dynamics and track lateral alignment. Journal of Rail and Rapid Transit.

[2]   Koike (2014). Numerical method for evaluating the lateral resistance of sleepers in ballasted tracks. Soils and Foundations, 54(3), 502-514.

[3]   Fakhari, M. (2012) Experimental Investigation of Frictional Sleeper Effect on the Lateral Resistance of Railway Track.

[4]   Ciotlăuș (2012). Increasing railway stability with support elements. Special sleepers. Acta Tech. Napocensis: Civ. Eng. Archit, 55(2), 165-172

Operational costs optimization method in transport systems for open-pit mines

Felipe Halles[1], Fran Ribes-Llario[2], Claudio Mansanet[3], Raul Redón[4], Julia Real Herráiz[5]*

[1] Altavia, Av. Presidente Kennedy, Santiago de Chile, Chile

[2, 3, ,4,5]Institute of MultidisciplinaryMathematics, PolytechnicUniversity of Valencia, Camino de Vera, 46022 Valencia, Spain

*Corresponding author. E-mail: jureaher@upv.es. Telephone: +34 650528210

1. **Introduction** In open-pit mining, it is crucial to maximize productivity while the costs of transporting the ores, which is vital to guarantee the future and profitability of the operation. It represents approximately 50% of the total operating cost. In this way, not only saving costs by the adequate maintenance of mining trucks is of major concern, also, it is possible to achieve a significant resources saving by improving road conditions.

A considerable amount of research has been developed in recent years in order to deal with this major challenge [1 – 10]. The most of these studies are focused on analysing the main characteristics of the network of mining roads – roughness, elevation profile, surface condition, etc. -.

The methodology proposed herein is to stablish an optimized action plan of maintenance operations on an open-pit mining road based on a sequence of different algorithms which will work by real time data input taken by sensors installed on the mining vehicle.

2. **Neural Network**

For this purpose, at first, an Artificial Neural Network (ANN) is defined to calculate the elevation profile and the roughness of the roads – by obtaining the Displacement Spectral Density (DSD) -. The input data are the vertical accelerations of both sides of the front axle of the mining vehicle. An ANN can be defined as a calculation model based on an interconnected set of nodes (neurons) organized in several layers.

Mathematically, each neuron adds the input signals weighted with a series of synaptic weights, and applies a function, called the activation function, which depends on this sum and the activation threshold of the neuron.

$$u_n = \sum_{i=1}^{m} w_{nj} x_j \qquad (1)$$

$$y_n = \varphi(u_n + b_n) \qquad (2)$$

Where x_j represents the input signals; w_{nj} the synaptic weights of the neuron; u_n is the result of the sum of the weighted input signals, b_n is the activation threshold, and $\varphi()$ the sigmoid activation function, which results in the output signal y_n. Sigmoid function is mathematically defined as [10]:

$$\varphi(u) = \frac{1}{1 + e^{au}} \qquad (3)$$

Where "a" is the characteristic parameter of the function.

A training stage must be carried out, which means to provide the neural network the input data in order to adjust the network parameters until obtain accurate output signals. The proposed model is based on a NARX-type neural network (Non Linear with Exogenus Inputs Networks) with 7 input neurons, 5 in the hidden layer and 2 in the output – where the elevation profile and the DSD are obtained -. This implies that the calculation of each result should be based on the values of the input signals as well as the past values obtained from the results.

$$y(t) = f(u(t), u(t-1), \dots, u(t-d), y(t-1), \dots, (y(t-d)) \qquad (4)$$

3. Defect analysis

Once the elevation profile is obtained, an identification and classification of road defects is carried out attending to its geometrical properties – length, width, angle, etc. -. There are several types of defects that could appear on the surface of a mining road, such as undulations, obstacles, gullies or bumps.

The methodology used is based on a search of the protuberances and concavities of the profile with a minimum size, with respect to the average profile. For each of those points, the main characteristics of the defect are extracted.

The type of defect found is determined from the comparison with the requirements marked for identification.

Once the defects are classified by type, a new classification is applied, separating the faults according to its gravity. This way it is possible find the faults that must be repaired.

Therefore, the repair cost of each defect is calculated according to its necessary corrective operations, searching in the database developed.

4. Rolling Resistance

A rolling resistance (RR) estimation based on the characteristics of the terrain – obtained from the DSD – is used to predict the operating cost for the mining road operation.

In this calculation, amortization costs, vibration and impacts effects and fuel consumption are considered to finally establish a daily operation cost. For example, fuel consumption can be calculated as:

$$C_t = \frac{Pv \cdot D \cdot FC}{1000 \cdot V} \tag{5}$$

Where D is the transport distance in meters, FC is the consumption factor (l/h•t), V is the vehicle speed (km/h), and Pv is the weight of the loaded or empty vehicle (t).

In order to calculate the operating costs of each road, a simulation of the operation of the vehicle will be practiced from the rolling resistance of the different sections.

5. Action Plan

The minimization of the total transport costs (daily operation and daily road maintenance costs) becomes into the Optimized Action Plan – where, when and how many resources are necessary to apply -.
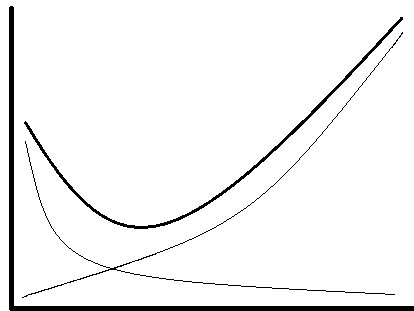
Figure 1: Daily costs evolution.

Whenever the overall expenses of the maintenance of the vehicle and the losses caused by the road condition are higher than the fixing costs, the corresponding maintenance operation will be carried out in each section. Otherwise, the losses and the cost of the operation and maintenance of the vehicle will be assumed.

References

[1] Thompson, R.J. & Visser, A. T (2003). Mine haul road maintenance management systems.

[2]   Thompson, R.J (2014). Mine Haul Road Design, Construction & Maintenance Management.

[3]   Marais, W.J., Thompson, R.J., Visser, A.T. (2008). Managing Mine Road Maintenance Interventions Using Mine Truck On-Board Data. University of Pretoria.

 [4]   Ngwangwa, H.M. & Heyns, P.S. (2014). Application of an ANN-based methodology for road surface condition identification on mining vehicles and roads. Journal of Terramechanics. 2014.

[5]   Rojas, R. (1996). Neural Networks: A Systematic Introduction. Springer-Verlag New York, Inc.

[6]   Thompson, R.J., Visser, A.T., Miller, R.E., Love, N.T. (2003). The Development of a real-time mine road maintenance management system using haul truck and road vibration signature analysis.

[7]   Hugo, D. (2005). Haul road defect identification and condition assessment using measured truck response. University of Pretoria.

[8]   Ngwangwa, H.M., Heyns, P.S. , Labuschangne, F.J.J., Kululanga, G.K. (2009) . Reconstruction of road defects and road roughness classification using vehicle responses with artificial neural networks simulation. Journal of Terramechanics.

[9]   Ngwangwa, H.M., Heyns, P.S., Breytenbach, H.G.A., Els., P.S. (2014) Reconstruction of road defects and road roughness classification using Artificial Neural Networks simulation and vehicle dynamic responses: Application to experimental data. Journal of Terramechanics.

[10] Hugo, D., Heyns, P.S.,Thompson, R.J.,Visser, A.T. (2008) Haul road defect identification using measured truck response.

Structural Railway Bridge Health monitoring by means of data analysis

Fran Ribes-Llario[1], Clara Zamorano[2], Piter Moscoso godoy[3], Julia Real Herráiz[4]*

[1, 4] *Institute of Multidisciplinary Mathematics, Polytechnic University of Valencia, Camino de Vera, 46022 Valencia, Spain*

[2] *UPM, Calle del Prof. Aranguren, 3, 28040 Madrid, Spain*

[3] *DGOP, Ministerio de Obras Públicas, Morandé 59, Santiago de Chile, Chile*

*Corresponding author. E-mail: frarilla@cam.upv,es. Telephone: +34 965 145 205

1. Introduction

The conservation of the railroad tracks is essential in order to maintain the geometric quality standards that the circulation demands at a certain speed. Railway maintenance has evolved over time in parallel with the changes introduced in the track grid design, given that the new track designs needed, on the one hand, fewer human resources for maintenance and, on the other hand, higher-performance track machinery.

Regarding the inspection methods for bridges, if we focus only on the structure itself we find two types of inspections, the basic inspection and the main inspection. These main inspections are not carried out systematically or on a regular basis, but as a consequence of the damages detected in a main inspection or as a consequence of an extraordinary situation

Once the issue has been framed, this paper intends to synthesise different mathematical methodologies that allow the diagnosis of the structural health of railway bridges to be carried out automatically from the excitations of the road operation itself.

2. Pathologies

Metallic truss bridges were widely used during XIX and XX century in railway infrastructures. These structures are light and allow to cross large spans and resist the exigent railway loads by using a low quantity of material. But their worst enemy is the fragile cracking, it is produced mainly by two causes: fatigue and corrosion.

Fatigue is caused by repetitive load-unload cycles where the strength limits of the material are not exceeded. During several cycles of deformation, the material accumulates deformation that could finally break.

Fatigue propagation velocity was described by Paris-Erdogan LAW. If loads doesn´t exceed a percent of material strength (usually 50%-60% for steel), unlimited life is assumed.

$$\frac{da}{dN} = C\Delta K^m \tag{1}$$

Interrogating this curve, the maximum number of cycles supported by a constant load level is obtained. This is known as the S-N curves (stress-number of cycles).

$$N = \frac{A}{\Delta\sigma^n} \tag{2}$$

For elements that support variable stress levels, it is necessary to normalize its contribution to final cracking in order to account them as a whole. This is achieved by the Palmgren-Miner expression and its normalized damage coefficient 'D'.

$$D_L = \sum \frac{n_i}{N_i} = \frac{n_1}{N_1} + \frac{n_2}{N_2} + \frac{n_3}{N_3} \tag{3}$$

3. Maintenance strategies

There are different maintenance strategies and techniques for structure conservation. Predictive maintenance with non intrusive means are the most treding techniques because they guarantee a good health of the structure without huge investments.

These structures have demonstrated being proper designed and many of them have successfully overcome their lifespan. It is also possible that their initial planed loading conditions have changed -weightier and faster vehicles, more traffic, etc-. For this reason, it is necessary to know its current status by non-intrusive on-site techniques in real time. There are two different ways for structural defects detection: critical points and global structure.

Fig 1. *Types of maintenance. Source: self made*.

In critical elements, which are supporting continuously load-unload cycles and where potential fatigue could appear, it is possible to know the stress level by its deformation measure using strain gauges.

On the other hand, in global structures is used natural frequencies variation to detect damages in unexpected structural element.

4. Method

For fatigue analysis, first critical element must be identified. Its initial lifespan estimation will be evaluated. After installing and obtaining first data set about load cycles and intensity. Damage index is obtained and current element lifespan is updated and the procedure starts again, damage index is obtained by Rainflow Algorithm.

After obtaining equivalent stress level in time domain, this signal is decomposed on cycles of representative intensity.

The number of cycles, average stress supported and the amplitude of stress added by each cycle are identified.

Stress normalization could be done by different ways, such as, Soderberg, Gerber and Goodman. Goodman is the most extended because it uses yielding point as a reference. The other ones use braking point, which is more difficult to know accurately.

Once the stress has been normalized, by using the characteristic S-N curve of our steel, it is possible to determinate which part of the stress is contributing to fatigue of the material per cycle -N_i- and finally obtaining the Damage coefficient D_i.

On the other hand, for general structure monitoring, the analysis of vibrations produced by normal operation and environment give us the natural frequencies of the structure in a determinate status.

It´s Power Spectrum Density decomposed in Singular Values give us the natural frequencies of the structure. A decrease in natural frequencies is indicator of structure stiffness variation that could be originated by a structural cracking.

For both procedures is only necessary to use commercial devices as accelerometers, strain gauges, microcontrollers and a transmission data Gateway.

For theorical validation of the procedure, a numerical FEM was implemented in ANSYS. A metallic truss structure was represented according to it´s building plans.
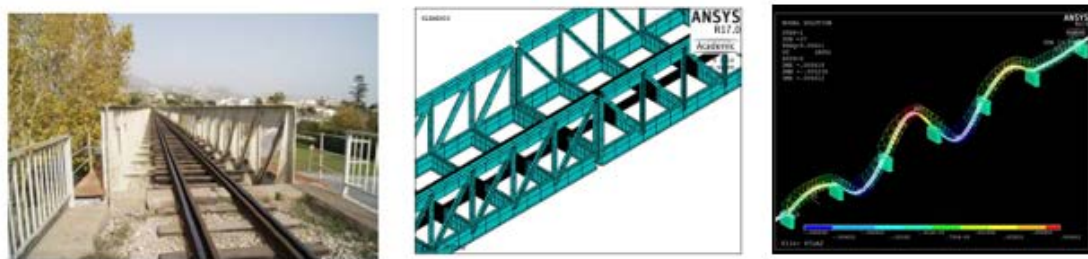
Fig 2. Numerical simulation. Source: self made

Two vehicle loading running through the bridge was simulated by means of a transient

analysis where 8 loads simulate the effect of the wheels.
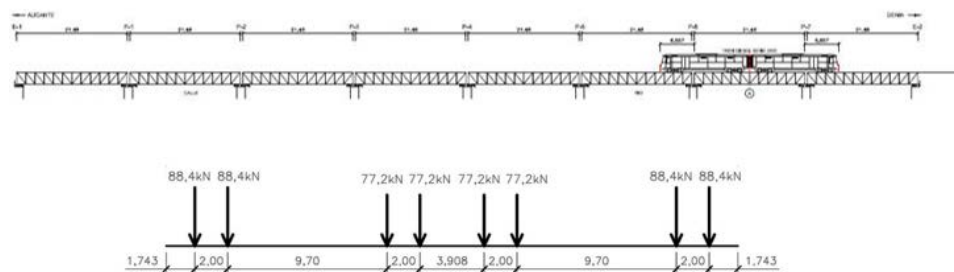
Fig 3. Loads of numerical simulation. Source: self made

5. Theorical result

Fatigue analysis procedure was used for forces and moments, and equivalent Von
Mises stress was obtained in a longitudinal beam.
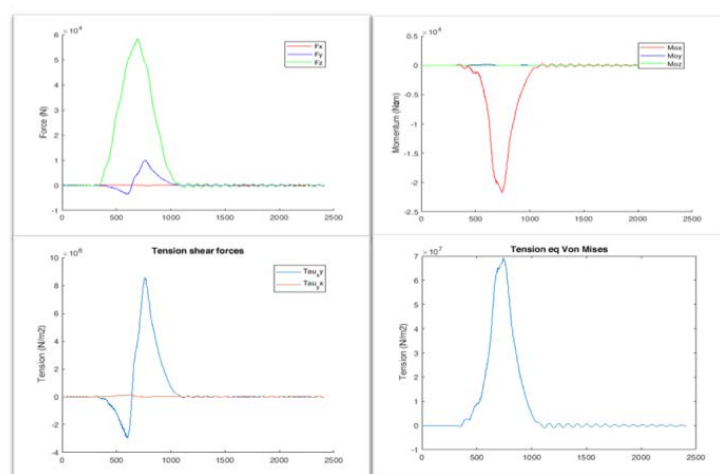
Fig 4. Forces of fatigue (top-left), moments of fatigue (top-right), tension shear forces (down-left) and tension eq. Von Mises (down-right). Source: self made

Rainflow algorithm results are shown on figure 5. S-N curve was also represented for normalized stresses and an unlimited life of this element was obtained. This means that the structure will not suffer fatigue for it´s current railway traffic.
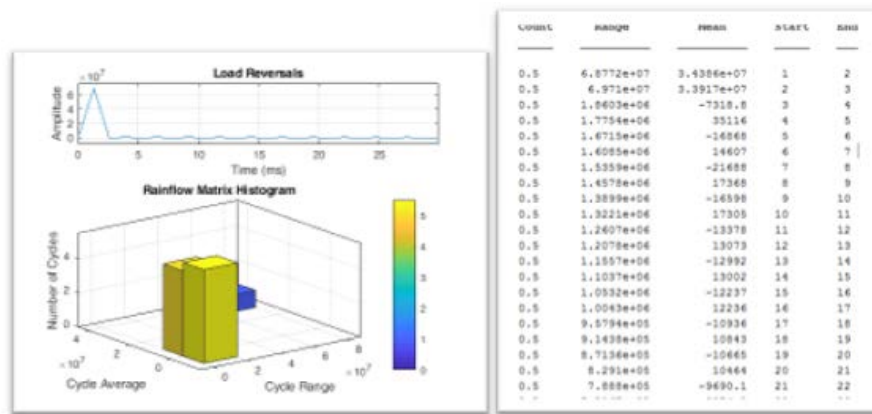
Fig 5. Rainflow algorithm results. Source: self made

Regarding modal analysis, the breaking of this same longitudinal element was simuled, which originated a NF decrease of 3,1% and a spectral amplitude decrease in vertical accelerations of 20%.
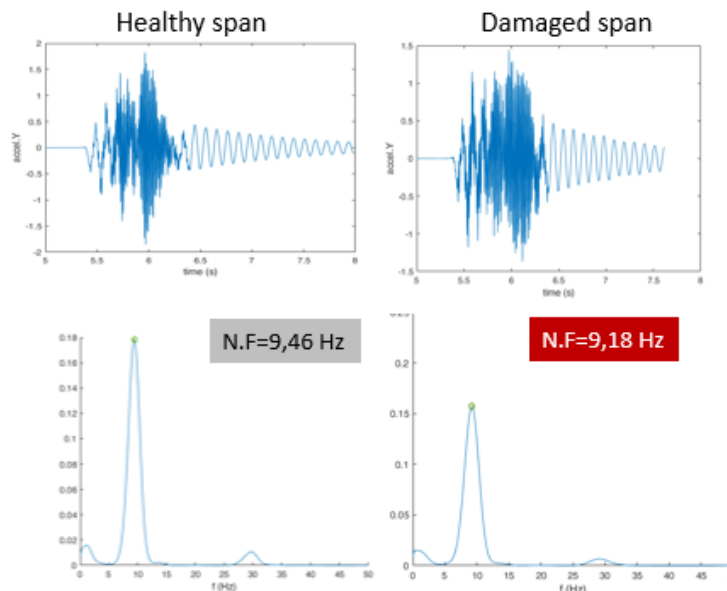
Fig 6. Cracking Healthy span (left) and Damaged span (right). Source: self made

A spatial model for mean house mortgage appraisal value in boroughs of the city of Valencia

M.A.López[♭], N.Guadalajara[♭] [*], A.Iftimi[†] and A. Usai[‡]

(♭) Universitat Politècnica de València, Spain

(†) Universitat de València, Spain

(‡) University od Sassari, Italy

November 30, 2018

1   Introduction

In the aftermath of the recent boom and the real estate burst in developed countries, growing interest in appraisal methods has been shown (Krause and Bitter, 2012), with an extended use of spatial econometrics. The geographical location of homes and the spatial dependence relations among prices are taken into account in these spatial models. The existence of spatial autocorrelation, which requires the use of a weight matrix, is assumed in these models (Dubin, 1998). Autoregressive models have recently progressed to more sophisticated models (Brasington, 2004; Liao and Wang, 2012; Liu, 2013) with different weight matrices (Affuso et al., 2017), which may lead to distinct results (Dubin, 1998).

Nonetheless, home purchase-sale prices have been used in all models (Brasington, 2004; Chernobai et al., 2011; Liao and Wang, 2012; McCluskey et al., 2013), and the few exceptions are: home price offers (Chasco and Le

[*]e-mail:nguadala@omp.upv.es

175

Gallo, 2013) when purchase-sale prices are not available, rent values (Pholo et al., 2014), and cadastral values (Zoopi et al., 2015). However, other housing values exist, such as mortgage appraisal values (McGreal and Taltavull, 2012), which no previous work has modeled. Mortgage appraisal values are used for financial institutions to grant home loans (Cerruti et al., 2017), and for evaluating financial risks (Lacour-Little and Malpezzi, 2003).

Given the importance of such appraisals, it is very important to develop appraisal models that identify the exact characteristics and services that underlie appraisals of home prices. This study aimed to analyse the effects that the mean characteristics of the homes in borough, the demographic and socio-economic aspects, citizen security and co-existence, environmental aspects, and closeness to public transport and education centres, had on the mean housing appraisal value in each borough. The city chosen for this study was Valencia.

2   Data and Methodology

The city of Valencia had 790 201 inhabitants in 2016, and is located in east Spain on the Mediterranean coast. It is divided into 19 districts in administration terms, which are all divided, in turn, into differents boroughs, which amount to 85. Districts 17 to 19 are on the outskirts, far from the city centre, and have a small population and a few buildings. For this reason, the analysis was limited to districts 1 to 16, which contain 70 boroughs. The mortgage appraisal price is available for 5551 residential properties in the province of Valencia, of which 2564 are located in these 70 boroughs. The information available for each residential property is: mortgage appraisal value (in euros), physical variables (surface (m^2), air conditioning, preservation, swimming pool, lift) and temporary variables (age of building). A hierarchy variable is also available, and refers to the cadastral value of the location, which varies between 1 (the highest value) and 67 (the lowest value). It corresponds to the zones with values restricted by the Cadastre to do calculations for cadastral reports. A map of value zones results from hierarchising the land values throughout the territory, obtained with national coordinate-homogeneity criteria. The home mortgage appraisal values were provided by four official appraisal societies according to the method set out in Article 2a) of Order ECO/805/2003. Each residential property was valued only once by a single company. The appraisal values of the residential

properties was a mean of 143 450.29 euros and 1270.86 euros/m2. The mean surface area of each property was 104.74 m2 and its mean age was 41.35 years. The geographical location in UTM coordinates was known for each property. The analysis included two phases: the first, an exploratory one, aimed to detect if spatial autocorrelation existed in the data; the second, a modeling one, built regression models to explain the mean price for a square meter. Presence of the spatial component in these models depends on the result of phase 1. We begin with a brief description of the tools to carry out the exploratory spatial data analysis (ESDA), followed by the spatial autoregressive and error regression models.

The first of the models, known as a spatial lag model, is expressed as:

$$y = \beta_0 + \sum_{n=1}^{\infty} \beta_k x_k + \rho W y + \varepsilon \tag{1}$$

where y is the response variable, $x_{k,k=1,...,p}$ are the independent variables with which we wish to model y, and ϵ the error term. The difference with a multiple linear regression model lies in the inclusion of term W_y, which represents the mean of the y values observed in its neighborhood. This model aims to capture the influence that the neighbors have on an observation. A positive ρ value indicates that y increases because of this influence, provided that ρ significantly differs from 0.

The model is denoted as spatial autoregressive by its analogy with the AR models in time series, in which the temporal autocorrelation is modeled by including temporal lags y_{t-k} of the dependent variable.

The second model is known as the spatial error model, which is expressed as:

$$y = \beta_0 + \sum_{n=1}^{\infty} \beta_k x_k + \rho W \varepsilon + \varepsilon \tag{2}$$

This model assumes that the errors in a multiple linear regression model are spatially autocorrelated. The spatial autoregressive model (1) does not allow such a direct interpretation as it happens in a multiple linear regression model (each one represents the variation that the response variable undergoes when the independent variable related to the parameter increases by one unit, but the others remain unchanged). If a similar notation to that used for the previous model is adopted, then:

$$y = \beta_0 \iota_n + \rho W y + X \beta + \varepsilon \tag{3}$$

where ρ is the spatial coefficient and W is the neighborhood matrix. A simple operation allows (3) to be rewritten as follows:

$$y = (I_n - \rho W)^{-1}(\beta_0 \iota_n + X\beta + \varepsilon) \qquad (4)$$

which evidences that now the partial derivatives matrix is not diagonal, which is what happened before. Now the variation of a covariable in one of the boroughs directly affects the value of the dependent variable for this borough, but also its neighbors and the neighbors of the other boroughs. Thus, by means of a kind of boomerang effect, the original borough is also affected.

3   Results

Moran's Index value for the spatial distribution of $\ln(\text{mvm}^2)$ is 0.4078, with a p-value of $2.44E^{-09}$ or 0.001, depending on whether it was obtained by the asymptotic test or by the random permutations test. This value implies the acceptance that spatial autocorrelation exists.

We observe coefficients ρ and λ of the spatial terms, the former is significant but not the latter. Another justification for using spatial models is provided by Moran's Indices for the residuals of the three models, with values of 0.123, 0.015 and -0.011 for the OLS, autoregressive and error model, respectively. The p-value of the OLS model residuals, 0.029, indicates that spatial autocorrelation still exists, while those that correspond to the spatial autoregressive and spatial error models indicate the exact opposite.

The three models show good performance for goodness of fit, which is deduced from its determination coefficients and from the other measures. With respect to the spatial models, the autoregressive model offers the best value for these indicators and its spatial term is significant, which does not happen with the error model. The autoregressive model also has the advantage of allowing a simpler and natural interpretation.

4   Conclusions

Like home purchase-sell prices, mortgage appraisal prices can also be modeled by hedonic and spatial models.

The OLS and the spatial error or autoregressive models provided very good fit results, which somewhat improved when spatial aspects were added,

spatial aspects which in turn eliminated the spatial autocorrelations observed in the OLS model.

The characteristics of the considered residential properties (state of preservation, lift, swimming pool, and the hierarchy variable of the cadastre value) clearly influenced the mean prices in the borough.

Regarding the borough-related variables, the variables associated with vehicles'age or cylinder capacity positively corresponded to the mean price of residential properties. The mean size of a family property, however, negatively affected prices, and the distance from a metro station or from infant or primary education centers also had a negative influence.

However, the models showed some unexpected results, such as the positive influence on the price of residential properties caused by the ratio of incidences due to drug addiction, unauthorized public activities or distances to secondary education centers. Nor does it seem logical that the mean distance to a tram or bus stop did not influence the models.

References

[1] Affuso E, Cummings JR and Le H. Wireless towers and home values: an alternative valuation approach using a spatial econometric analysis. *Journal of Real Estate Finance and Economics* , 2017. doi:10.1007/s11146-017-9600-9.

[2] Brasington DM. House prices and the structure of local government: an application of spatial statistics. *Journal of Real State and Economics*, 29(2): 211?231, 2004. doi:10.1023/B:REAL.0000035311.59920.74

[3] Cerruti E, Dagher J and Dell?Ariccia G. Housing finance and real-estate booms: A cross-country perspective. *Journal of Housing Economics*, 38:1-13, 2017. doi.org/10.1016/j.jhe.2017.02.001

[4] Chasco C and Le Gallo J. The Impact of Objective and Subjective Measures of Air Quality and Noise on House Prices: A Multilevel Approach for Downtown Madrid. *Economic Geography*, 89 (2): 127-148, 2013. doi: 10.1111/j.1944-8287.2012.01172.x

[5] Chernobai E, Reibel M and Carney M. Nonlinear spatial and temporal effects of highway construction on house prices. *Journal of Real Estate*

Finance Economics, 42(3): 348-370, 2011. doi:10.1007/s11146-009-9208-9

[6] Dubin RA (1998) Spatial autocorrelation: a primer. *Journal of Housing Economics*, 7: 304?327, 1998.

[7] Krause AL and Bitter C. Spatial econometrics, land values and sustainability: Trends in real estate valuation research. *Cities*, 29: S19?S25, 2012. doi.org/10.1016/j.cities.2012.06.006

[8] Lacour-Little M and Malpezzi S. Appraisal Quality and Residential Mortgage Default: Evidence from Alaska. *Journal of Real Estate Finance and Economics*, 27(2): 211?233, 2003. doi.org/10.1023/A:1024728420837

[9] Liao WCh and Wang X. Hedonic house prices and spatial quantile regression. *Journal of Housing Economics*, 21(1): 16?27, 2012. doi.org/10.1016/j.jhe.2011.11.001

[10] Liu X. Spatial and Temporal Dependence in House Price Prediction. *Journal of Real Estate Finance Economics*, 47: 341?369, 2013. doi:10.1007/s11146-011-9359-3

[11] McCluskey WJ, McCord M, Davis PT, et al. Prediction accuracy in mass appraisal: a comparison of modern approaches. Journal of Property Research 30 (4): 239?265, 2013. dx.doi.org/10.1080/09599916.2013.781204

[12] McGreal S and Taltavull P. An analysis of factors influencing accuracy in the valuation of residential properties in Spain. *Journal of Property Research*, 29(1): 1?24, 2012. dx.doi.org/10.1080/09599916.2011.589531

[13] Pholo A, Peeters D and Thomas I. Spatial issues on a hedonic estimation of rents in Brussels. *Journal of Housing Economics*, 25: 104?123, 2014. doi.org/10.1016/j.jhe.2014.05.002

[14] Zoopi C, Argiolas M and Lai S. Factors influencing the value of houses: Estimates for the city of Cagliari, Italy. *Land Use Policy*, 42: 367?380, 2015. doi.org/10.1016/j.landusepol.2014.08.012

Third order root-finding methods based on a generalization of Gander's result

S. Busquier[♭], J. M. Gutiérrez[†,*], and H. Ramos[‡]

(♭) Dept. Applied Mathematics and Statistics,

Polytechnic University of Cartagena, Spain,

(†) Dept. Mathematics and Computer Sciences,

University of La Rioja, Logroño, Spain,

(‡) Dept. Applied Mathematics,

Higher Polytechnic School, University of Salamanca, Zamora, Spain.

November 30, 2018

1 Derivation of a new family of third order iterative methods

In a classical result given by Gander in 1985 [3] it is shown that third order methods for solving nonlinear equations $f(x) = 0$ can be written in the form $x_{n+1} = G(x_n)$ where

$$G(x) = x - H(L_f(x))\frac{f(x)}{f'(x)}, \tag{1}$$

H is a twice differentiable function around $x = 0$ that satisfies the conditions $H(0) = 1$, $H'(0) = 1/2$ and $|H''(0)| < \infty$ and $L_f(x)$ is defined as the quotient

$$L_f(x) = \frac{f(x)f''(x)}{f'(x)^2}. \tag{2}$$

---

*e-mail: jmguti@unirioja.es

There exists a geometrical way to obtain these methods, as it can be seen in [2]. It is based on the idea of approximating the function $f(x)$ by means of adequate interpolating functions of conic type, in particular parabolic or hyperbolic curves. In this paper we present two new iterative methods that are constructed from interpolating functions of exponential and logarithmic type. In the first case, we approximate $f(x)$ by a function $g(x)$ of exponential type given by

$$g(x) = a + \exp(c + bx),$$

where a, b, c are chosen to satisfy

$$g(x_n) = f(x_n), \quad g'(x_n) = f'(x_n), \quad g''(x_n) = f''(x_n)$$

at a given approximation x_n. In this way, we get the numerical method given by

$$x_{n+1} = x_n + \log\left(1 - L_f(x_n)\right) \frac{f'(x_n)}{f''(x_n)}, \tag{3}$$

where $L_f(x)$ is the quotient defined in (2).

In the second case, we approximate $f(x)$ by a function $h(x)$ of logarithmic type:

$$h(x) = a + c\log(x + b),$$

where a, b, c are chosen to satisfy the tangency conditions

$$h(x_n) = f(x_n), \quad h'(x_n) = f'(x_n), \quad h''(x_n) = f''(x_n)$$

at a given approximation x_n. In this way, we deduce the numerical method

$$x_{n+1} = x_n + (1 - \exp(L_f(x_n))) \frac{f'(x_n)}{f''(x_n)}. \tag{4}$$

The idea of considering interpolating functions of exponential type has been considered, for instance by S. Amat and S. Busquier [1]. In view of the expressions (3) and (4), and taking into account Gander's result, we can formulate a convergence result for methods of the form $x_{n+1} = S(x_n)$ for

$$S(x) = x - T(L_f(x))\frac{f'(x)}{f''(x)}. \tag{5}$$

Theorem 1 *Let α be a simple zero of the equation $f(x) = 0$ that satisfies $f''(\alpha) \neq 0$. Let us assume that $T : \mathbb{R} \to \mathbb{R}$ is a differentiable enough function around zero that satisfies $T(0) = 0$, $T'(0) = -1$ and $T''(0) = -1$. Then the iterative method given by $x_{n+1} = S(x_n)$, where S is defined in (5) is convergent to α with at least order three.*

Remark 1 *As an application of this result, we see that iterative methods of the form*

$$x_{n+1} = x_n - \left(L_f(x_n) + \frac{1}{2}L_f(x_n)^2 + \sum_{j \geq 3} a_j L_f(x_n)^j \right) \frac{f'(x_n)}{f''(x_n)} \qquad (6)$$

are cubically convergent to simple roots of $f(x) = 0$. The previous series development must be seen in a formal way, assuming that the corresponding convergence conditions are fulfilled.

Remark 2 *In particular, if $a_j = 0$ for $j \geq 3$ in (6), we obtain a new construction of the well-known Chebyshev's iterative method*

$$x_{n+1} = x_n - \left(1 + \frac{1}{2}L_f(x_n) \right) \frac{f(x_n)}{f'(x_n)}. \qquad (7)$$

This method and some of its properties have been studied by many authors (see [4] for more details).

Remark 3 *We would like to highlight that the family of methods given in (6) is essentially the same as the one given previously by Gander in (1). In fact,*

$$
\begin{aligned}
x_{n+1} \quad &= x_n - \left(L_f(x_n) + \frac{1}{2}L_f(x_n)^2 + \sum_{j \geq 3} a_j L_f(x_n)^j \right) \frac{f'(x_n)}{f''(x_n)} \\
&= x_n - \left(1 + \frac{1}{2}L_f(x_n) + \sum_{j \geq 2} a_{j+1} L_f(x_n)^j \right) L_f(x_n) \frac{f'(x_n)}{f''(x_n)} \\
&= x_n - \left(1 + \frac{1}{2}L_f(x_n) + \sum_{j \geq 2} a_{j+1} L_f(x_n)^j \right) \frac{f(x_n)}{f'(x_n)}.
\end{aligned}
$$

2 Two local convergence results with asymptotic error constants

We present new local convergence theorems for the iterative methods (3) and (4), that in addition prove their cubic order of convergence. In the statement of these results we consider the usual notations

$$e_n = x_n - \alpha, \quad c_j = \frac{f^{(j)}(\alpha)}{j!f'(\alpha)}, j \geq 2. \qquad (8)$$

Theorem 2 *Assume that $f(x) : D \to \mathbb{R}$ is a sufficiently many times differentiable function with a simple zero $\alpha \in D$, with D an open interval, and let x_0 be an initial guess close enough to α. Then, the exponentially-fitted method defined in (3) has third-order of convergence and the error equation is*

$$e_{n+1} = \frac{2c_2^2 - 3c_3}{3} \, e_n^3 + \mathbb{O}(e_n^4) \, .$$

Theorem 3 *Assume that $f(x) : D \to \mathbb{R}$ is a sufficiently many times differentiable function with a simple zero $\alpha \in D$, with D an open interval, and let x_0 be an initial guess close enough to α. Then, the logarithmically-fitted method defined in (4) has third-order of convergence and the error equation is*

$$e_{n+1} = \frac{4c_2^2 - 3c_3}{3} \, e_n^3 + \mathbb{O}(e_n^4) \, .$$

We finish our work with some numerical examples where we have compared the introduced methods with other well-known methods of same order. From the numerical examples, we conclude that the proposed methods outperform the other methods for each of the types of functions for which they have been designed, and can be comparable for other kind of functions. Thus, the proposed methods may be considered as alternative methods for solving nonlinear equations, particularly if the functions involved present exponential or logarithmic behaviors.

References

[1] S. Amat and S. Busquier. Geometry and convergence of some third-order methods *Southwest J. Pure Appl. Math.*, 2:61–72, 2001.

[2] S. Amat, S. Busquier and J. M. Gutiérrez. Geometric constructions of iterative functions to solve nonlinear equations. *J. Comput. Appl. Math.*, 157:197–205, 2003.

[3] W. Gander. On Halley's iteration method. *Amer. Math. Monthly*, 92:131–134, 1985.

[4] M. García-Olivo and J. M. Gutiérrez. Notas históricas sobre el método de Chebyshev para resolver ecuaciones no lineales. *Miscelánea Matemática*, 57:63–83, 2013.

ASSESSMENT OF A GRAPHIC MODEL FOR SOLVING DELAY TIME MODEL INSPECTION CASES OF REPAIRABLE MACHINERY. PREDICTION OF RISK WHEN SELECTING INSPECTION PERIODS.

Fernando Pascual[1], Emilio Larrodé[1,2], Victoria Muerza[2,3]

[1]UNIVERSITY OF ZARAGOZA. Dpto. de Ingeniería Mecánica. C/ María de Luna, 3 – 50018. Zaragoza (Spain). Tel: +34 976 762319. fpascual@unizar.es, elarrode@unizar.es

[2]ARAGON INSTITUTE OF ENGINEERING RESEARCH (i3A). Edificio de I+D+i, C/ Mariano Esquilllor s/n – 50018, Zaragoza (Spain). Tel: +34 976 761888. vmuerza@unizar.es

[3]MIT INTERNATIONAL LOGISTICS PROGRAM. ZARAGOZA LOGISTICS CENTER. C/ Bari 55, Edificio Náyade 5 – 50197, Zaragoza (Spain). Tel: +34 976 077604. vmuerza@zlc.edu.es

1. INTRODUCTION

This article focuses on the study of the maintenance cost optimization of an annuity under the Life Cycle Cost (LCC) approach based on reliability engineering. For doing this, the Delay Time Model approach is adapted to the railway case and a previous proposal of graphic resolution method is used (Pascual et al., 2017). In this paper, we present the discussion of the model where a study is made of the boundaries of the domain in which the graphic method is valid. Once the limits of the domain are established, a graphical analysis of the sharpness of the curve is performed, which we interpret as the risk in the selection of inspection periods, depending on the different variables of the problem. Likewise, a graphic study of the asymmetry of the model is carried out, depending on the different variables of the problem. On the other hand, an assessment of the influence of the use of exponentials in the graphic model used and the possibility of simplification that implies is performed. Finally, the drawing of graphs developed for the prediction of risk in the selection of inspection periods is proposed, and after its evaluation, the construction of a simplified method of interpretation of results is proposed.

2. ASSESSMENT

Initially the limits of the work domain will be discussed, that is, the ranges of the physical variables (failure rate, inspection period in days, delay time, etc...) for which the developed graphic method is valid, concluding that the abacuses are valid for most applications with physical meaning.

Next, the shape of the "curve of the bathtub" is studied, explaining how the asymmetry of the same and its sharpening varies depending on each of the variables involved, emphasizing the relevance of the sharpening of the function C(T) - total cost depending on the inspection period (see Eq. (1)). This sharpening represents the risk that is assumed (in increment of incurred cost) when deviating from the optimal inspection period. Flat curves imply that the cost increase will be low if it is not inspected with the optimum frequency. However, very sharp curves will involve a large increase in cost (great risk) if it is not inspected at the optimum frequency.

To study the influence of the use of exponential distributions on risk, we start with the graphic study of each of the terms that make up the cost function as a function of period. The expression of the cost function is the following:

$$C(T) = \frac{\lambda \cdot \left[\frac{e^{-\beta T}-1}{\beta}+T\right] \cdot C_b}{T} - \frac{\lambda \cdot \left[\frac{e^{-\beta T}-1}{\beta}\right] \cdot C_{CM}}{T} + \frac{C_i}{T} \tag{1}$$

As a result, new abacuses are presented, representing the level of risk or sharpening for the entire field of solutions, proposing a possible categorization of equipment to be maintained according to their risk and cost (see Figure1 and Figure 2).
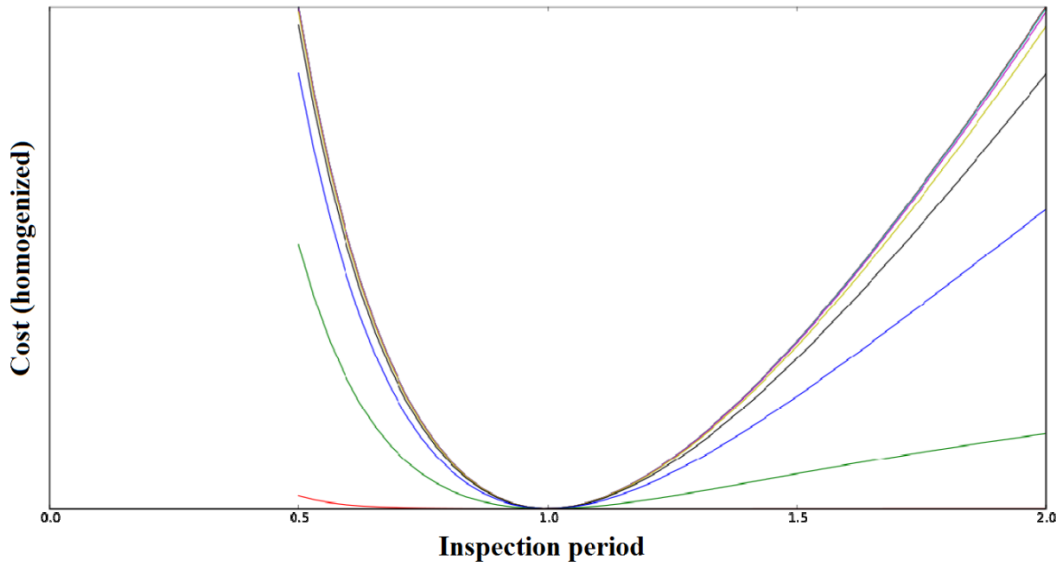
Figure 1. Cost functions as a function of the period for the same period of T = 1 day, identical failure rate and variable delay time rate. Curves taken to the same abscissa point to study the shape.

Figure 1 shows potential achievable solutions for an assortment of delay time values. All cost values have been homogenized to the same cost value to ease visual comparison. Two relevant shape parameters become then noticeable: both the sharpness and the asymmetry of the curves.

Regarding the sharpness, the longer the delay time used on the model, the flatter the curve gets. Sharper curves imply higher risk whereas a flat curve means a lower risk related if a suboptimal inspection period is chosen. It has been observed that the curve sharpness achieves an asymptotical limit that can also be observed on the General Graphic Model (Pascual et al., 2017)
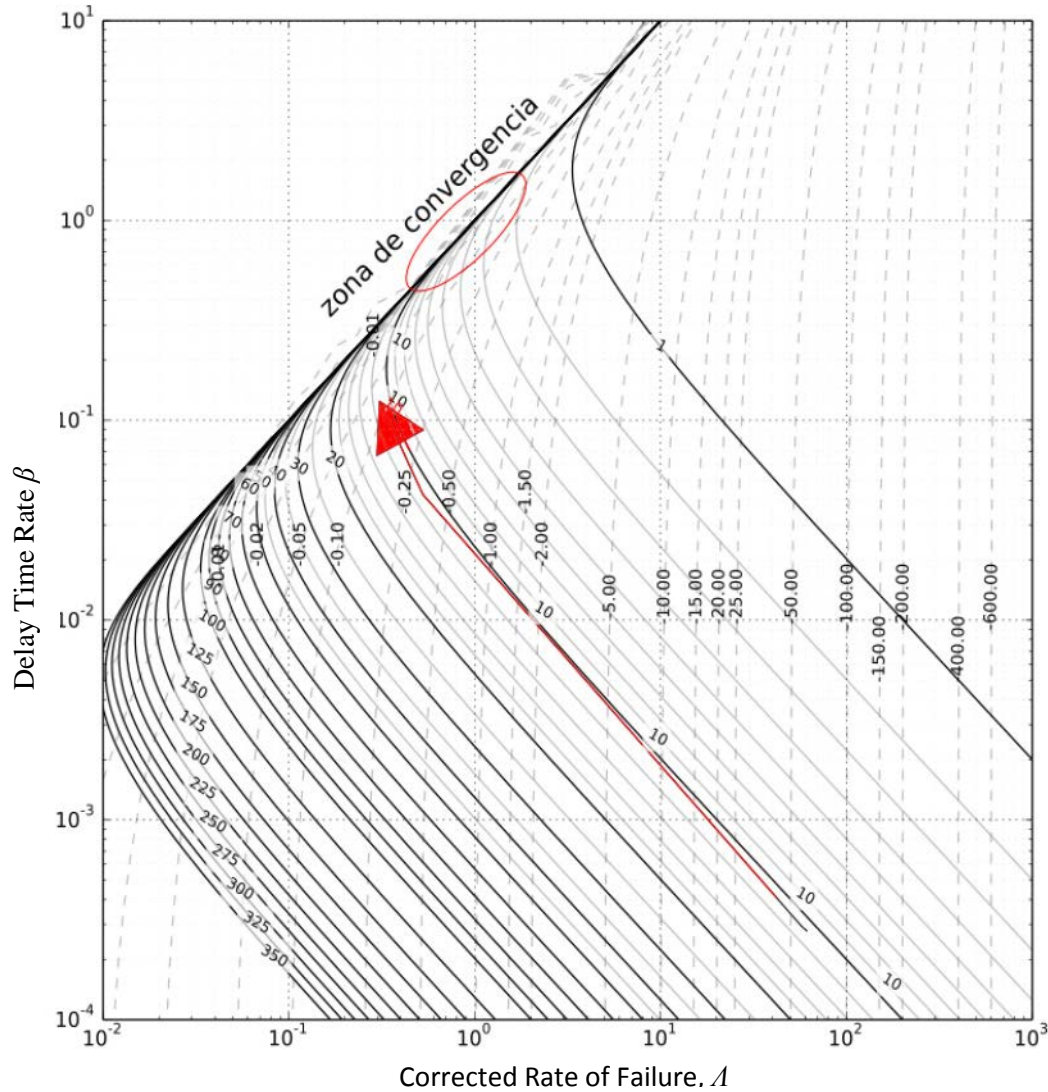
Figure 2. Optimum inspection period to obtain maximum savings per cycle based on the Corrected Rate of Failure Λ and Delay Time Rate β

The isovalues line for inspection times T=10 shows how an increase on the Delay Time Rate β implies a decrease on the Corrected Rate of Failure Λ until it arrives to a convergence zone.

To study the sharpness variation depending upon the inspection period, Figure 1 was calculated and plotted for several inspection periods. The resulting potential cost functions for an assortment of time delay values are shown on Figure 3 for inspection periods equal to T=10, T=20 and T=150
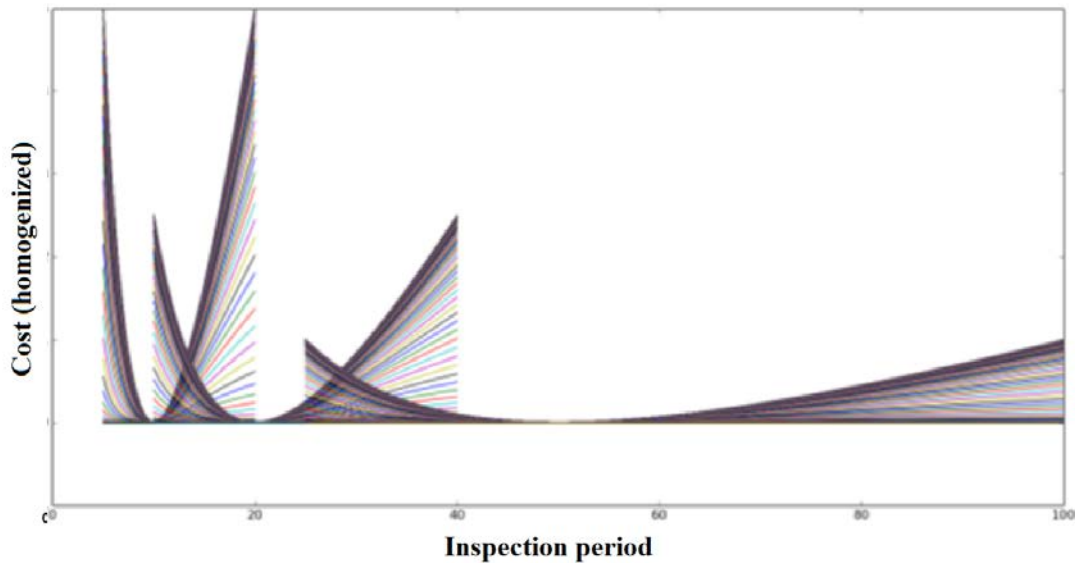
Figure 3. Families of cost curves for periods T = 10, T = 20 and T = 50; Cost scale (abscissa) homogenized and variable delay time

A visual inspection of the potential solutions reveals two observable facts:

First, certain asymmetry is shown for any given delay time. A stiffer negative slope is observed on inspection periods shorter than the optimal inspection period. Beyond the optimal inspection period, the positive slope becomes softer.

On the other hand, maximum sharpness occurs for lower inspection periods. This behavior arises from the use of exponential distribution to model both, the failure rate and the delay time. Thus, the Delay Time Model (Christer 1984) when applied with exponential distributions will yield risk-decreasing solution curves as the optimal inspection period increases. This outcome entails that physical cases with intrinsic long inspection periods will get low-risk, flat cost curves.

3. CONCLUSION

This paper shows the assessment of the graphical model for solving the Delay Time Model with exponential distributions for its application in inspection of repairable machinery. The objective is to establish the limits of the domain for its practical application in this type of operations. The discussion of its limits of application allows us to make a correct use and obtain an enriched knowledge of the problem to study in an agile and precise way. The possibility of obtaining effective information that avoids the risk of expensive decision making when setting the delay times in maintenance operations, is another of the main contributions of the graphic model.
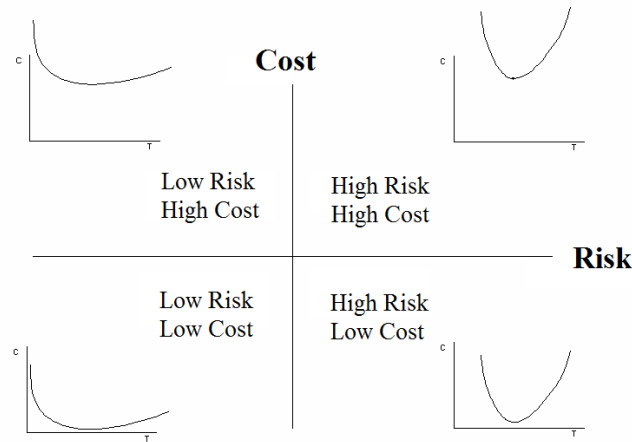
Figure 3. Qualitative graphic classification according to risk and cost

Assigning to each element studied a cost value (the optimal cost) and a value for the risk (the average of the slopes around the optimum), the elements can be categorized, visually ordering them in a perceptual map in which each value is represented according to its cost and risk (Figure 3).

REFERENCES

Baker, R. D., & Scarf, P. A. (1995). Can models fitted to small data samples lead to maintenance policies with near-optimum cost?. IMA Journal of Mathematics applied in Business and Industry, 6(1), 3-12.Christer, A. H., and Waller, W. M. (1984). Delay time models of industrial inspection maintenance problems. *Journal of the Operational Research Society*, 35(5), 401-406.

Dunk, A. S. (2004). Product life cycle cost analysis: the impact of customer profiling, competitive advantage, and quality of IS information. *Management Accounting Research*, 15(4), 401-414.

González-Fernández, J. (2012). Teoría y Práctica del mantenimiento industrial. FC editorial, Madrid.

Jardine, A. K., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical systems and signal processing,* 20(7), 1483-1510.

Pascual, F., Larrodé, E., and Muerza, V. (2017). Proposal of a graphic model for solving delay time model inspection cases of repairable machinery. In: Modelling for Engineering and Human Behaviour 2017. Eds: Jódar, L., Cortés, J.C., and Acedo, L., Instituto Universitario de Matemática Multidisciplinar, pp. 234-237.

Tang, Y., Jing, J. J., Yang, Y., & Xie, C. (2014). Parameter Estimation of a Delay Time Model of Wearing Parts Based on Objective Data. Mathematical Problems in Engineering.

Wang, H. (2002). A survey of maintenance policies of deteriorating systems. *European Journal of Operational Research*, 139(3), 469-489.

Christer, A.H., "Operational Research applied to industrial maintenance and replacement" in: Eglese and Rands (eds), Developments in Operational Research, Pergamon Press, Oxford 1984, 31-58

A high order iterative scheme of fixed point for solving nonlinear Fredholm integral equations

M.A. Hernández-Verón$^{\flat *}$, María Ibáñez†,

Eulalia Martínez‡, and Sukhjit Singh$^{\diamond}$

($\flat$) Department of Mathematics and Computation,

University of La Rioja, Spain,

($\dagger$) Facultat de Ciències Matemàtiques,

Universitat de València, Valencia, Spain,

($\ddagger$) Instituto Universitario de Matemática Multidisciplinar,

Universitat Politècnica de València, Spain,

($\diamond$) Department of Mathematical Sciences,

I.K.G. Punjab Technical University, Jalandhar, India.

November 30, 2018

1 Introduction

In this paper, we consider the integral equations given by

$$x(s) = f(s) - \lambda \int_a^b K(s,t)H(x(t))dt, \qquad (1.1)$$

with $H : \mathbb{R} \longrightarrow \mathbb{R}$ a derivable scalar function, $f : [a,b] \longrightarrow \mathbb{R}$ a continuous function and $K : [a,b] \times [a,b] \longrightarrow R$ continuous function in both arguments.

Nonlinear integral equation (1.1) is a particular case of Fredholm integral equations [1, 2]. These equations have strong physical background and arise from the electromagnetic fluid dynamics and play a very significant role in several applications, as for example the dynamic models of chemical reactors.

---

*e-mail: mahernan@unirioja.es

As the Fredholm integral equations of form (1.1) cannot be solved exactly, we can use numerical methods to solve them, such as the different techniques that can be found in the references of this work. If we pay attention to the iterative methods that can be applied for approximating a solution $x^* \in \mathcal{C}[a, b]$ of (1.1), the method of successive approximations plays an important role (see, [3]-[6]).

This method consists of applying the fixed point theorem to the equation

$$x(s) = F(x)(s), \tag{1.2}$$

with $F : \Omega \subseteq \mathcal{C}[a, b] \longrightarrow \mathcal{C}[a, b]$, where Ω is a nonempty convex domain in $\mathcal{C}[a, b]$, with $F(x)(s) = f(s) - \lambda \int_a^b K(s, t) H(x(t)) dt$ and obtaining a sequence $\{x_{n+1} = F(x_n)\}_{n \in \mathbb{N}}$ that converges to a solution $x^* \in \mathcal{C}[a, b]$ of (1.1), i. e., a fixed point of F.

So in this paper, we consider an iterative process for approximating a fixed point of F, whose iterative algorithm is

$$
\begin{cases}
y_n = x_n - [I - F'(x_n)]^{-1}(x_n - F(x_n)) \\
x_{n+1} = y_n - [I - F'(x_n)]^{-1}(y_n - F(y_n)), n \geq 0,
\end{cases} \tag{1.3}
$$

where $x_0 \in \mathcal{C}[a, b]$ is given.

We obtain a semilocal convergence result for the iterative process (1.3) from which we will carry out the qualitative study for the equation (1.1).

In what follows, we consider the Nemytskii operator $\mathcal{H} : \Omega \subseteq \mathcal{C}[a, b] \longrightarrow \mathcal{C}[a, b]$ such that $\mathcal{H}(x)(s) = H(x(t))$. Obviously, it is a Frechet differentiable operator and then, the operator

$$F(x)(s) = f(s) - \lambda \int_a^b K(s, t) H(x(t)) dt$$

verifies

$$[F'(x)y](s) = -\lambda \int_a^b K(s, t)[\mathcal{H}'(x)y](t) dt = -\lambda \int_a^b K(s, t) H'(x(t)) y(t) dt.$$

1.1 Existence and location of a solution for (1.1)

Now, to obtain a semilocal convergence result for (1.3), let us assume that the following conditions are satisfied:

(I) $\Gamma_0 = [I - F'(x_0)]^{-1}$ exists for some $x_0 \in \Omega \subseteq \mathcal{C}[a, b]$, with $\|\Gamma_0\| \leq \beta$, $\|\Gamma_0(x_0 - F(x_0))\| \leq \eta$.

(II) $\mathcal{H}'$ is a ω-Lipschitz continuous operator such that

$$\|\mathcal{H}'(u) - \mathcal{H}'(v)\| \leq \omega(\|u-v\|) \text{ for } u,v \in \Omega, \tag{1.4}$$

where $\omega : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ is a continuous and nondecreasing function satisfying $\omega(\alpha z) \leq \phi(\alpha)\omega(z)$ for $\alpha, z \in [0,+\infty)$ with $\phi : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ a continuous and nondecreasing function.

As first step, from the previous conditions, we easily obtain the following result for the operator F'.

Lemma 1.1. F' *is a ω-Lipschitz continuous operator in Ω such that*

$$\|F'(u) - F'(v)\| \leq |\lambda| M \omega(\|u-v\|) \text{ for } u,v \in \Omega,$$

with $M = \max\limits_{s \in [a,b]} \left| \int_a^b K(s,t)dt \right|.$

As second step, denoting $\Gamma_n = [I - F'(x_n)]^{-1}$, we prove the existence of these operators for each $n \in \mathbb{N}$ from Banach Lemma.

Lemma 1.2. *Given $R \in \mathbb{R}_+$, if $x_n \in B(x_0, R) \subseteq \Omega$, and $\beta|\lambda|M\omega(R) < 1$ then $[I - F'(x_n)]^{-1}$ exists and $\|[I - F'(x_n)]^{-1}\| \leq \beta_R$, where*

$$\beta_R = \frac{\beta}{1 - \beta|\lambda|M\omega(R)}.$$

From now, we denote $\theta(t) = \dfrac{\beta}{1 - \beta|\lambda|M\omega(t)}$, then $\beta_R = \theta(R)$.

In what follows, we tested two more technical lemmas by analyzing how the iterative method (1.3) works for our problem (1.2) under conditions (I) and (II) established in subsection 2.1, we obtain the bounding conditions to define the recurrence relations need for the sequences $\{x_n\}$ and $\{y_n\}$. First of all, we define the following parameters

$$\begin{aligned} r_0 &= \eta \\ s_0 &= \psi_R(r_0)r_0 \\ S &= 1 + \psi_R(r_0) \\ T &= \chi_R(r_0, s_0)\psi_R(r_0) \end{aligned}$$

where $\psi_R(u) = \beta_R\,|\lambda|\,M\,Q\,\omega(u)$, $\chi_R(u,v) = \beta_R\,|\lambda|\,M\,(\omega(u)+Q\,\omega(v))$ and $Q = \int_0^1 \phi(t)dt$.

So, by considering the following scalar sequences $r_n = Tr_{n-1}$ and $s_n = \psi_R(r_n)r_n$, we have that, if $T < 1$, r_n and s_n are decreasing scalar sequences and after proving some results we establish the existence of the fixed point x^*.

Theorem 1.1. *With the previous notations, let $F : \mathcal{C}[a,b] \to \mathcal{C}[a,b]$ be the nonlinear Fréchet differentiable operator given by $[F(x)](s) = f(s) - \lambda \int_a^b K(s,t)\mathcal{H}(x)(t)dt$. If the equation:*

$$t = \frac{1 + \theta(t) \ |\lambda| \ M \ Q \ \omega(\eta)}{1 - \theta(t) \ |\lambda| \ M \ \big(\omega(\eta) + Q \ \omega(\theta(t) \ |\lambda| \ M \ Q \ \omega(\eta)\eta)\big)\theta(t) \ |\lambda| \ M \ Q \ \omega(\eta)}\eta \qquad (1.5)$$

*has at least one positive real root and the smallest positive real root, denoted by R, satisfies $\beta \ |\lambda| \ M \ \omega(R) < 1$, $B(x_0,R) \subseteq \Omega$ and assumptions (**I**) and (**II**) hold, then, for the starting point x_0, the method (1.3) converges to a fixed point x^* of (1.2). Moreover, $x_n, y_n, x^* \in \overline{B(x_0,R)}$.*

Finally, we apply all the above-mentioned to some nonlinear Fredholm integral equations for obtaining different results on the existence and uniqueness of the solution of these applied problems.

References

[1] A. D. Polyanin, A. V. Manzhirov, Handbook of integral equations. CRC Press, Boca Raton, 1998.

[2] J. Rashidinia, A. Parsa, Analytical-numerical solution for nonlinear integral equations of Hammerstein type. International Journal of Mathematical Modelling and Computations, 1, 61–69, (2012).

[3] Traub, J.F.: Iterative Methods for the Solution of Equations. Prentice-Hall, Englewood Cliffs, New Jersey (1964).

[4] Argyros, I.K., Hilout, S.: Numerical methods in Nonlinear Analysis. World Scientific Publ. Comp. New Jersey (2013).

[5] Cordero, A., Ezquerro, J.A., Hernández, M.A., Torregrosa, J.R.: On the local convergence of a fifth-order iterative method in Banach spaces. Appl. Math. Comput. 251, 396-403 (2015).

[6] Singh, S., Gupta, D.K. Martínez, E., Hueso, J.L.: Enlarging the convergence domain in local convergence studies for iterative methods in Banach spaces. Mediterr. J. Math. 13, 4219-4235 (2016).

Some parametric families improving Newton's method *

Alicia Cordero, Sergio Mallasén[†], and Juan R. Torregrosa

Institute for Multidisciplinary Mathematics, Universitat Politècnica de València,

Camino de Vera, s/n, 46022-Valencia, Spain

November 30, 2018

1   Introduction

The problem of finding a simple zero $\bar{x}$ of a nonlinear function $f(x)$ yields frequently to the use of an approximating method. Newton's scheme is the best known one to find $\bar{x}$, being a one-step method. Based on it, extensive research has been developed over the years to improve its behavior, in terms of order of convergence or computational efficiency. Some good texts about both kind of procedures can be found in [1–3].

In this manuscript, we present several parametric classes that can hold the quadratic convergence of Newton's method and whose sets of converging initial estimations can be wider than those of Newton's scheme, for some problematic functions.

The proposed iterative family is

$$x_{k+1} \;\; = \;\; x_k - \frac{f(x_k)}{f'(x_k) + \alpha f(x_k)}, \;\; k = 0, 1, 2, \dots, \tag{1}$$

where α is a parameter. Note that for $\alpha = 0$, we obtain Newton's scheme. The following result establishes the convergence of iterative methods (1).

[†]e-mail: sermalqu@ade.upv.es

Theorem 1 *Let $f : D \subseteq \mathbb{R} \longrightarrow \mathbb{R}$ be sufficiently differentiable at each point of an open interval D such that $\bar{x} \in D$ is a simple solution of equation $f(x) = 0$ and the initial estimation x_0 is close enough to $\bar{x}$. Then, sequence $\{x_k\}_{k \geq 0}$ obtained from expression (1) converges to $\bar{x}$ with order 2 with independence of parameter α, being in this case the error equation*

$$e_{k+1} = (\alpha + C_2)e_k^2 + O(e_k^3),$$

where $c_j = \dfrac{1}{j!} \dfrac{f^{(j)}(\bar{x})}{f'(\bar{x})}$, $j = 2, 3, \ldots$ and $e_k = x_k - \alpha$.

Other proposed families are

$$x_{k+1} = x_k + \frac{1}{\alpha} \ln\left[1 - \alpha \frac{f(x_k)}{f'(x_k)}\right], \tag{2}$$

or even

$$x_{k+1} = x_k - \alpha \frac{f(x_k)}{f'(x_k)} - \beta \left(\frac{f(x_k)}{f'(x_k)}\right)^2. \tag{3}$$

Let us observe that family (3) includes Newton's method for $\alpha = 1$ and $\beta = 0$. We can establish a similar result to Theorem 1 for these both classes of iterative methods. The methods of families (2) and (3) have order of convergence two, for any values of the parameters.

2 General second-order weight function structure

These families can be generalized with the following weight function structure

$$x_{k+1} = x_k - H(t(x_k)), \tag{4}$$

where $t(x_k) = \dfrac{f(x_k)}{f'(x_k)}$.

The order of convergence of (4) is established in the following result.

Theorem 2 *Let $f : D \subseteq \mathbb{R} \longrightarrow \mathbb{R}$ be a real function with the second derivative in D. Let $\bar{x} \in D$ be a simple root of $f(x) = 0$. If we choose an initial guess close enough to $\bar{x}$ and a sufficiently differentiable function $H(t)$ such*

that $H(0) = 1$, then methods described by (4) converge to $\bar{x}$ with quadratic order of convergence, being their error equation

$$e_{k+1} = (-H'(0) + c_2)e_k^2 + O(e_k^3),$$

where $c_j = \dfrac{1}{j!}\dfrac{f^{(j)}(\bar{x})}{f'(\bar{x})}$, $j = 2, 3, \ldots$ and $e_k = x_k - \bar{x}$.

3 Numerical performance

In this section, the numerical results obtained by applying Newton's method and the families denoted by M1 (1), M2 (2) and M3 (3) are compared using the following test functions.

- $f_1(x) = \arctan(x), \quad \bar{x} = 0,$

- $f_2(x) = x^3 - 2x + 2, \quad \bar{x} \approx -1.769292,$

- $f_3(x) = \sin(x) + x\cos(x), \quad \bar{x} = 0,$

- $f_4(x) = x^2 e^{x^2} - \sin^2(x) + x, \quad \bar{x} = 0.$

The numerical computations have been carried out using MATLAB R2017 with variable precision arithmetics and 1000 digits of mantissa. The stopping criterion used is $|x_{k+1} - x_k| + |f(x_{k+1})| < 10^{-100}$. For each class and test function, we calculate the number of iterations, the value of residual $|f(x_{k+1})|$ at the last iteration and the computational order of convergence $ACOC$, approximated by (see [4])

$$p \approx ACOC = \frac{\ln(|x_{k+1} - x_k| \,/\, |x_k - x_{k-1}|)}{\ln(|x_k - x_{k-1}| \,/\, |x_{k-1} - x_{k-2}|)}. \tag{5}$$

As it can be seen in Table 1, M1 and M2 (for specific value of $\alpha = 0.319$ and $\alpha = -0.107$, respectively) converge to the root of $f_1(x)$ with $x_0 = 2$ when Newton's method does not. The same can be said for $f_2(x)$ when $x_0 = 1$, Newton's scheme diverges while M1, M2 and M3 converge to the root. For M3 we observe that its approximated computational order of convergence (ACOC) is only 1 due to it does not satisfy $H(0) = 1$, which is a necessary condition in Theorem 4 for having quadratic order of convergence. In Table 1, the no convergence is represented by $-$.

	Function		$f_1(x)$	$f_2(x)$	$f_3(x)$	$f_4(x)$
	x_0		2	1	0.5	3
iter	Newton		31	-	7	18
	M1	$\alpha = 0.319$	7	64	8	22
	M1	$\alpha = -0.107$	21	11	8	20
	M2	$\alpha = -0.574$	10	32	8	22
	M3	$\alpha = 0.834, \beta = 0.289$	-	142	130	143
incr	Newton		-	-	2e-774	9e-504
	M1	$\alpha = 0.319$	4e-214	4e-297	5e-333	1e-312
	M1	$\alpha = -0.107$	4e-313	2e-221	4e-302	3e-263
	M2	$\alpha = -0.574$	4e-357	3e-325	2e-285	1e-347
	M3	$\alpha = 0.834, \beta = 0.289$	-	2e-101	6e-102	9e-102
ACOC	Newton		-	-	3.0	4.0
	M1	$\alpha = 0.319$	2.0	2.0	2.0	2.0
	M1	$\alpha = -0.107$	2.0	2.0	2.0	2.0
	M2	$\alpha = -0.574$	2.0	2.0	2.0	2.0
	M3	$\alpha = 0.834, \beta = 0.289$	-	1.0	1.0	1.0

Table 1: Numerical results for the different methods and several test functions.

(a) M1, $\alpha = 0.1$ (b) M2, $\alpha = 0.1$ (c) M3, $\alpha = \beta = 0.1$
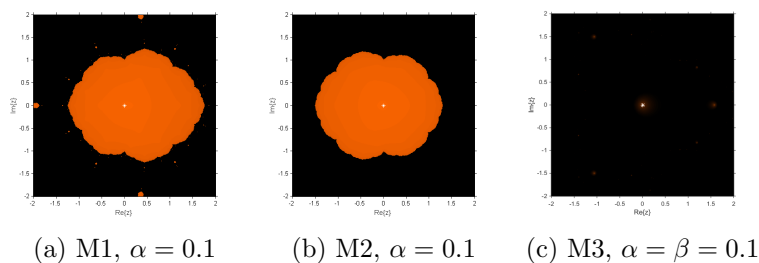
Figure 1: Dynamical planes of all the methods for $f_1(x)$

In Figure 1, different dynamical planes are shown where a mesh of 400×400 initial estimations have been defined and, by using the routines described in [5] and the value of parameter $\alpha = 0.1$ in case of Figures 1a and 1b and $\alpha = \beta = 0.1$ for M3 family in Figure 1c, each initial point has been plotted

in orange color if it has converged to the zero of $f_1(x)$ (up to a tolerance of 10^{-3} and in black color if it has not happened in a maximum of 80 iterations. It can be observed that the set of initial converging points is a bit wider in case of class M1 than that of M2, being the basin of attraction of the zero much smaller in case of M3.

(a) M1   (b) M2   (c) M3
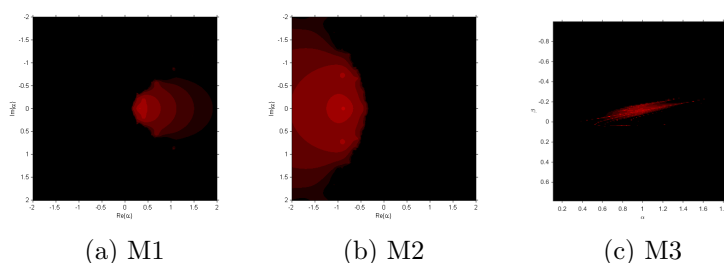
Figure 2: Parameter planes of all the methods for $f_1(x)$ and $x_0 = 2$

However, in Figure 2, the point of view is different: the initial estimation is fixed in x_0 (the value used in Table 1) and the mesh of points defines complex values of parameter α (for families M1 and M2, Figures 2a and 2b) or pairs of real values of (α, β) in case of class M3, in Figure 2c. We observe that the best values of α to converge to the root of $f_1(x) = 0$ are those with real part positive and close to zero, meanwhile in case of M2, α should have the real part negative. In case of M3, the get the best results, real α must be inside the interval $[0.5, 1.5]$ and $\beta \in [-0.2, 0]$. These results agree with the numerical performance of the families presented in Table 1.

References

[1] J.F. Traub, Iterative methods of the solution of equations. Prentice-Hall, New York, 1984.

[2] S. Amat, S. Busquier, Advances in Iterative Methods for Nonlinear Equations, in: SEMA SIMAI Springer Series, vol. 10, Springer International Publishing, Switzerland, 2016.

[3] M.S. Petković, B.Neta, L.D. Petković, J.Džunić. Multipoint methods for solving nonlinear equations. Elsevier, 2013.

[4] A. Cordero, J.R. Torregrosa. Variants of Newton's method using fifth-order quadrature formulas. *Appl. Math. Comput.* 190: 686–698 (2007).

[5] F.I. Chicharro, A. Cordero, J.R. Torregrosa. Drawing dynamical and parameters planes of iterative families and methods. *The Scientific World Journal* ID 780153, pages 1–11 (2013).

Modeling consumer behavior in Spain

P. Merello[♭] *, L. Jodar[†], G. Douklia[†], and E. de la Poza[‡]

(♭) Department of Accounting,

University of Valencia, 46071, Valencia, Spain,

(†) Instituto Universitario de Matematica Multidisciplinar,

Universitat Politecnica de Valencia, 46022 Valencia, Spain,

(‡) 2Centro de Ingenieria Economica,

Universitat Politecnica de Valencia, 46022 Valencia, Spain.

November 30, 2018

1   Introduction

Consumption is one of the main drivers of the economy. It is promoted in our society due to its contribution to the public revenues through taxation as well as its impact on macroeconomics indicators. Indeed, in western counties, shopping is classified as a leisure activity, a way to manage emotions or a means of expressing the self-identity.

In this context, a current culture of consumption has been generated, leading to the development of different types of consumers. The rational consumer who bases his purchase on a logical mental process and that is governed by satisfying primary needs has given way to consumers who obtain pleasure through the purchased product or the shopping experience and, ultimately, consumers who consume with spending patterns that interfere in their interaction with society, carrying serious labour, social and economic problems [1].

This overconsumption is driven by external or internal factors, but have in common that the consumer receives a profit from the product or from the

---

*e-mail:paloma.merello@uv.es

purchase experience itself. Thus, we can find hedonistic consumption and conspicuous consumption. In the first, the consumer is self-satisfied thanks to the act of consumption or the pleasures derived from the product [2], [3]; in the latter, the consumer emulates the upper classes through the purchase of similar products [4], [5].

In XXI century, digital revolution has brought the emergence of e-commerce and individuals are changing their patterns of behaviour. The literature provides evidence that consumers' compulsion differences exist compulsion depending on their frequency of online shopping [6]. The results of [6] illustrate that the type of goods purchased is not a determinant to distinguish compulsive and non-compulsive buyers; but also, Lam and Lam [7] stated that buying on internet more than once a week increases significantly the risk of becoming a problematic shopper.

The age has been identified as a determinant factor in the development of a pathological consumption disorder [1]. Concretely, age is inversely correlated with the disorder indicating that younger people are more prone to manifest the pathology.

In this paper, the transit of individuals among subpopulations is explained through diverse factors such as the hedonistic consumption (Pascal effect) [8], the imitation or conspicuous consumption (Veblen effect), [9], bandwagon effect, economy, psychological, technological and demographic.

In this vein, the interest of the study relies on quantifying the magnitude of the over consumers in Spain and the reasoning behind their behaviour. Also, it is relevant to determine the drivers that explain how an impulsive consumer might transit to pathological buyer [10]. Thus, to design prevention strategies and/or develop a more responsible culture of consumption, reinforcing from childhood the values that prevent the risk of development of pathological consumption behaviour, [11].

2   The model

Consumers can be classified depending on their buying behaviour into three categories: ordinary consumers, impulsive consumers and pathological consumers. Ordinary consumers (N_j) are those buyers with a rational and planned purchase behavior, impulsive consumers (S_j) are characterized by spontaneous, immediate loss of control buying [1],[12] and pathological consumers (A_j) repeat inappropriate spending patterns that interfere with social,

work, or role functioning [1],[13].

In this work we identify and classify the Spanish population by their level of consumption paying special attention to the differences by ages. Thus, four age groups $j = 1, 2, 3, 4$ comprising the age intervals [16-25]; [26-35]; [36-64] and over 65 years old, respectively, are identified.

We propose a compartmental discrete non-linear model of difference equations to explain consumers' behaviour trends during the period of time 2016-2020 [14]. We consider a short term period for the simulations as the hypotheses need to be accurate for a finite domain and that is only guaranteed in short-term for harsh human problems [15].

The compartmental difference equations model for the dynamic of buying behavior in Spain is as follows (t in semesters),

$$
\begin{aligned}
N_{t+1} &= (\sigma - \gamma \times \sigma - \mu \times \sigma)B_t + \sum_{j=1}^{4} i_{j,t}\tau - d_j N_{j,t} + N_t + \alpha\frac{1}{58}S_{3,t} + \\
&\quad \sum_{j=1}^{4} i_{j,t}(1-\tau)\frac{N_{j,t}}{B_{j,t}} - \left(V'_{j,t} + P'_{j,t}\right)N'_{j,t} + rA_t\,, \\
S_{t+1} &= S_t + (\gamma \times \sigma) - d_j S_{j,t} - \alpha\frac{1}{58}S_{3,t} + \sum_{j=1}^{4} i_{j,t}(1-\tau)\frac{S_{j,t}}{B_{j,t}} + \\
&\quad \left(V'_{j,t} + P'_{j,t}\right)N'_{j,t} - (P+V)\left(u_j + E_{j,t}\right)S_{j,t}\,, \\
A_{t+1} &= A_t + (\mu \times \sigma)B_t - d_j A_{j,t} + \sum_{j=1}^{4} i_{j,t}(1-\tau)\frac{A_{j,t}}{B_{j,t}} + \\
&\quad (P+V)\left(u_j + E_{j,t}\right)S_{j,t} - rA_t\,.
\end{aligned}
\tag{1}
$$

The demographic variables considered are the birth rate (σ) and the death rate (d_j) for every j. Besides, new incomers in the model needs to consider γ and μ as the prevalence rate of impulsive (S) and pathological consumption (A) in high school students. The migratory balance is also considered (parameter $i_{j,t}$), and defined as immigrants minus emigrants for every sub-population.

In addition, a retirement effect is considered, defining α as the proportion of impulsive consumers of 64 years old that transit to rational consumers when they become 65 years old and get retired [16].

As regards to the transits from ordinary (N) to impulsive buyers (S), Pascal and Veblen effects are considered. This way, we define P and V as the percentage of consumers affected by the Pascal and Veblen effect, respectively. Furthermore, e-commerce influences the frequency with which consumer goods are purchased [6], hence it is assumed a catalyst for the Veblen and Pascal effects and involved in the formulation of $P_{j,t}$ and $V'_{j,t}$.

Finally, transits from impulsive (S) to pathological buyers (A) include also Pascal and Veblen effects, but they need an external trigger to take

j	N	$\%$	S	$\%$	A	$\%$
$j=1$	$1,600,827$	34.9	$2,138,373$	46.6	$848,213$	18.5
$j=2$	$1,913,703$	31.8	$2,997,857$	49.8	$1,114,155$	18.5
$j=3$	$6,688,474$	33.7	$11,581,381$	58.3	$1,588,683$	8.0
$j=4$	$5,988,292$	69.2	$2,375,051$	27.4	$294,362$	3.4
Total		41.4		48.8		9.8

Table 1: Initial sub-populations in Spain at January 2016, $t = 0$.

place. That kind of transit is complex and we consider that only an external distress, economic $(E_{j,t})$ and/or emotional (u_j), can affect those consumers suffering from anxiety or depression increasing their levels of compulsion.

As the recognition of the addiction has been maintained in our society, we consider the same therapy recovery rate (r) for pathological buyers as in [17] and adapt its value to semi-annual transits.

The initial sub-populations are computed and take the values as Table 1 shows.

The initial sub-populations of pathological buyers (A_j) have been calculated according to the literature values for different samples and ages [17],[18],[19]. Rational consumers (N_j) perform planned purchases; this category also comprises those consumers that due to their beliefs (religion/culture) or forced by their economic situation (long-term unemployed workers and people at risk of poverty) do not spend their economic surplus in purchases. Finally, the impulsive consumers (S_j) in $t = 0$ are calculated by difference for every j.

3   Results

As the economic situation is considered in the model, we simulate the short-term future sub-populations in July 2020 under different possible future evolutions of the poverty risk rate.

The following three scenarios are considered. The base scenario does not consider any substantial changes neither in the political nor in the economic situation of Spain in the following years but also from January 2017 the poverty risk rates keep constant. The recovery scenario considers a possible labor reform with reinforcement of the salaries, temporality and labor condi-
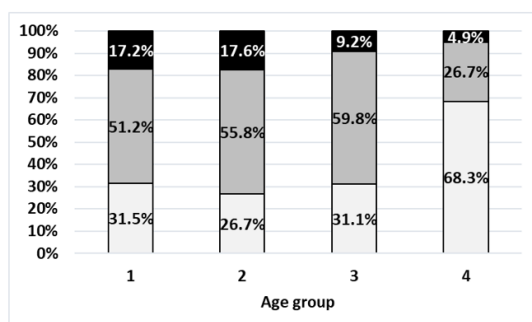
Figure 1: Sub-populations of consumers by age in Spain in July 2020; base scenario. White block represents the ordinary consumers, gray block represents impulsive consumers and black block represents pathological consumers.

tions as a result of the change in the Spanish government. Under the recovery scenario, the poverty risk rates are decreasing 0.5 points every semester since January 2017 for $j = 1, 2, 3$; however, it remains constant for $j = 4$ as retirement pensions will rise with the CPI and their purchasing power will not change. The worsening scenario assumes an opposite situation. As far as we know, the current macroeconomic indicators in Spain are improving but precarious workers also increase [20]. If we consider that the government does not develop any labor reform, precariat will increase. Under the worsening scenario we consider an increase in the poverty risk rates about 0.5 points every semester since January 2017.

The results of the simulations are shown in Figure 1 and 2. Figure 1 illustrates the evolution of the different sub-populations if the economic situation remains constant.

Under base scenario, ordinary buyers decrease for all age groups, as well as pathological consumers increase for $j = 3$ and 4 (being 8% and 3.4% the percentage estimated for 2016 and 9.2% and 4.9% in 2020, for $j = 3$ and 4, respectively) and there is a lightly decrease for $j = 1$ and 2 (being 18.5% the percentage estimated for 2016 and 17.2% and 17.6% in 2020, for $j = 1$ and 2, respectively).

It is remarkable that impulsive buyers increase for all sub-populations minus for $j = 4$. A sensitivity analysis is performed for the retirement effect coefficient (α) considering α between 0.1 and 0.9 under the base scenario. The results predict an increase of the impulsive consumers in the total Spanish consumers' population for all possible values of α.
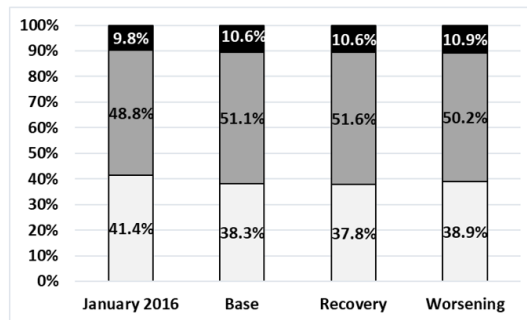
Figure 2: Total sub-populations of consumers in Spain in January 2016 compared to sub-populations in July 2020 under the three scenarios considered. White block represents the ordinary consumers, gray block represents impulsive consumers and black block represents pathological consumers.

As regards to the differences between scenarios, Figure 2 shows that ordinary consumers decrease for the three scenarios considered, meanwhile impulsive and pathological buyers increase. Pathological consumers increase the most under worsening scenario, due to the economic effect of the increase of precariat. The recovery scenario presents the largest increase in the impulsive buyers' subpopulation, from 48.8 in January 2016 to 51.6% in July 2020.

4 Conclusions and discussions

The compulsive buying has been a topic of interest in the recent decades as the consumption culture has led to the development of different types of shopping behavior.

Impulsive consumption is governed by two internal behavioral mechanisms that respond fundamentally to the hedonism and the emulation or Veblen effect. This paper presents a compartmental mathematical model that allows estimating in the short term the ordinary, impulsive and pathological buyers in Spain under three different economic scenarios.

The results show that impulsive and pathological buyers will increase under all economic scenarios. Notable differences in the number of ordinary buyers are found for the group over 65 years old.

The limitations of this work are the impossibility of finding real data

for Spain for some of the transit parameters. As is common in this type of models, the estimates are calculated under some assumptions and hypotheses of simplification of human behavior. In the future, it would be interesting to validate the estimates of our model by comparing with data from a real sample of Spanish consumers.

References

[1] Muller, A., Mitchell, J.E., De Zwaan, M. Compulsive buying. *American Journal on Addictions*, 24(2):132–137, 2015.

[2] Ozen, H., Engizek, N. Shopping online without thinking: being emotional or rational? *Asia Pacific Journal of Marketing and Logistics*, 26 (1):78–93, 2014.

[3] Lu, J., Liu, Z., Fang, Z. Hedonic products for you, utilitarian products for me. *Judgment and Decision Making*, 11(4):332–341, 2016.

[4] Freedman, Alix M. Little Wishes Form the Big Dream. *The American Way of Buying (Wall Street Journal)*:4–10, 1991.

[5] I., Ryabov. Conspicuous consumption among Hispanics: Evidence from the Consumer Expenditure Survey. *Research in Social Stratification and Mobility*, 44:68–76, 2016.

[6] Harnish, R.J., Bridges, K.R., Karelitz, J.L. Compulsive Buying: Prevalence, Irrational Beliefs and Purchasing. *International Journal of Mental Health and Addiction*, 15 (5):993–1007, 2017.

[7] Lam, L.T., Lam, M.K. The association between financial literacy and Problematic Internet Shopping in a multinational sample. *Addictive Behaviors Reports*, 6:123–127, 2017.

[8] B. Pascal, Pensees, edition de Leon Brunschwig, Garnier-Flammarion, Paris, France, 1976.

[9] O. Bomsel, L'economie immaterielle, Industries et marches d'experiences, Gallimard, Paris, France, 2010.

[10] C. Nakken. The Addictive Personality. Understanding the Addictive Process and Compulsive Behavior. 2th Edition, Hazelden, Minessota, United States, 1998.

[11] J. Rifkin, La era del acceso, Paidos, Barcelona, Spain, 2013.

[12] Yi S. Heterogeneity of compulsive buyers based of impulsivity and compulsivity dimensions. A latent profile analytic approach. *Psychiatry Research*, 208:174–182, 2013.

[13] McElroy, S.L., Keck, P.E., Pope, H.G, et al. Compulsive buying: A report of 20 cases. *Journal of Clinic Psychiatry*, 55:242–248, 1994.

[14] J.H. Goldthorpe, Sociology as a Population Science, Cambridge University Press Cambridge, Cambridge, U.K, 2016.

[15] De la Poza, E., Jódar, L. . A Short-Term Population Model of the Suicide Risk: The Case of Spain. *Culture, Medicine and Psychiatry*:1–21, 2018, Article in Press.

[16] Vida Caixa. (2017). Barometer Vida Caixa. Portrait of a Spanish retiree. Available at: https://www.vidacaixa.es/documents/51066/151087/conclusiones-barometro-vidacaixa-retrato-del-jubilado.pdf/641c7683-4ae7-0d01-bdf6-797947d88ef7 (Accessed 21.09.2018)

[17] De la Poza, E., Garcia, I., Jódar, L., Merello, P. Does VAT growth impact compulsive shopping in Spain? In Cortes, J.C; Jódar, L.; Villanueva R.J. Eds. *Mathematical Modeling in Social Sciences and Engineering*, Nova Science Publishers, Inc., New York, USA, 2014.

[18] Harnish, R. J., Bridges, K. R. Compulsive buying: the role of irrational beliefs, materialism, and narcissism. *Journal of Rational-Emotive Cognitive-Behavior Therapy*, 33:1–16, 2014.

[19] Maraz A, Griffiths MD, Demetrovics Z. The prevalence of compulsive buying: a meta-analysis. *Addiction*, 111(3):408–19, 2016.

[20] De la Poza, E., Fernández, A.E., Jórdar, L., Merello, P. A theoretical model to explain the rise of the European Precariat. *Proceedings of the 4th International Conference on European Integration*, Ostrava, Czech Republic, May 17-18, 2018.

Hamiltonian approach to human personality dynamics: an experiment with methylphenidate

Joan C. Micó[1]*, Salvador Amigó**, Antonio Caselles****

(*) Institut Universitari de Matemàtica Multidisciplinar.
Universitat Politècnica de València.
Camí de Vera s/n., 46022, ciutat de Valencia, Spain.
(**) Departament de Personalitat, Avaluació i Tractaments Psicològics.
Universitat de València,
Av. Blasco Ibáñez 21, 46010, ciutat de Valencia, Spain.
(***) IASCYS member, Departament de Matemàtica Aplicada.
Universitat de València.
Dr. Moliner 50, 46100 Burjassot, Spain.

1. **Introduction**

The present work is an attempt to define a minimum action principle to describe the short term dynamics of personality as a consequence of a stimulus, including the Lagrangian and the Hamiltonian functions of this formalism [1]. In physics, the current problem consists in getting the dynamics (by a set of coupled second order differential equations) from a known Lagrangian. However, the inverse Lagrange problem [2] consists in finding the Lagrangian from the known dynamics. In the context of this paper, the inverse Lagrange problem is applied to the short term dynamics of personality as a consequence of a stimulus. Once the Lagrangian is found, the Hamiltonian is derived by its definition from the Lagrangian.

Personality is here measured by the *Five-Adjective Scale of the General Factor of Personality* (GFP-FAS) [3], which measures dynamically the *General Factor of Personality* (GFP), i.e, a way to measure the overall human personality [4]. The so-called response model is the mathematical tool used to model the personality dynamics [5]. However, the response model here presented has a slight different mathematical structure, which produces a more realistic dynamics [6]. The response model presented is an integro-differential equation where the stimulus is an arbitrary time function. It is transformed in a second order differential equation for which a Lagrangian and a Hamiltonian are found, solving like this the corresponding inverse Lagrange problem.

An application case is presented: an individual consumes 20 mg of methylphenidate, and the GFP-FAS are observed every 7.5 minutes during 3 hours. Methylphenidate is a stimulant drug that can be modelled by a known time function [7], which produces significant changes in the biological bases of personality [7, 8]. This time function is considered in the second order differential equation. The response model is then calibrated with the experimental outcomes of the individual GFP-FAS. The corresponding Hamiltonian dynamics is also reproduced.

2. **The response model**

The response model is given by the integro-differential equation:

$$\left.\begin{array}{c} \frac{dy(t)}{dt} = a(b - y(t)) + p \cdot s(t) \cdot y(t) - q \cdot \int_0^t e^{\frac{x-t}{\tau}} \cdot s(x) \cdot y(x)\, dx \\ y(0) = y_0 \end{array}\right\} \qquad (1)$$

In Eq. 1, $y(t)$ represents the GFP dynamics; and b and y_0 are respectively its tonic level and its initial value. Its dynamics is a balance of three terms, which provide the time derivative of the GFP: the homeostatic control $(a(b - y(t)))$, i.e., the cause of the fast recovering of the tonic level b, the excitation effect $(p \cdot s(t) \cdot y(t))$, which tends to increase the GFP, and the inhibitor effect $(q \cdot \int_0^t e^{\frac{x-t}{\tau}} \cdot s(x) \cdot y(x)\, dx)$, which tends to decrease the GFP and is the cause of a continuously delayed recovering. Parameters a, p, q and τ are named respectively the homeostatic control power, the excitation effect power, the inhibitor effect power and the inhibitor effect delay. In addition, the $s(t)$ time function represents the dynamics of an arbitrary stimulus.

Taking the time derivative in Eq. 1 and subsequently substituting the inhibitor effect (the integral term) in this equation, the second order differential equation, in addition of the initial conditions, arises:

---

[1] E-mail: jmico@mat.upv.es

$$\left.\begin{array}{c} \ddot{y}(t) = \left(-a - \frac{1}{\tau} + p \cdot s(t)\right)\dot{y}(t) + \left(-\frac{a}{\tau} - q \cdot s(t) + \frac{p}{\tau} \cdot s(t) + p \cdot s'(t)\right)y(t) + \frac{a \cdot b}{\tau} \\ y(0) = y_0 \\ \dot{y}(0) = a(b - y_0) + p \cdot s_0 \end{array}\right\} \quad (2)$$

Eq. 2 is an equivalent version of Eq. 1. In it, s_0 is the amount of in the initial time $t=0$. From now onwards Eq. 2 is the version of the response model to be used.

3. Hamiltonian for the response model

The minimum action principle applied to Eq. 2 asserts that the Action between two arbitrary times t_1 and t_2, defined as $A = \int_{t_1}^{t_2} L(t, y, \dot{y})\, dt$, being $L(t, y, \dot{y})$ the Lagrangian, must be minimum under the parameter variation, i.e., $\delta A = 0$. The last equation provides the so-called Euler-Lagrange equation:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{y}}\right) = \frac{\partial L}{\partial y} \quad (3)$$

The Lagrange inverse problem consists in finding the Lagrangian that provides Eq. 2. To solve it, if $u(t)$, $v(t)$ and $w(t)$ are unknown time functions by the moment, the following Lagrangian is essayed:

$$L(t, y, \dot{y}) = \frac{1}{2}u(t) \cdot \dot{y}^2 + \frac{1}{2}v(t) \cdot y^2 + w(t) \cdot y \quad (4)$$

Applying Eq. 3 to Eq. 4:

$$\ddot{y}(t) = -\frac{u'(t)}{u(t)}\dot{y} + \frac{v(t)}{u(t)}y + \frac{w(t)}{u(t)} \quad (5)$$

By comparing Eq. 5 and Eq. 3:

$$-\frac{u'(t)}{u(t)} = -a - \frac{1}{\tau} + p \cdot s(t) \; ; \frac{v(t)}{u(t)} = -\frac{a}{\tau} - q \cdot s(t) + \frac{p}{\tau} \cdot s(t) + p \cdot s'(t) \; ; \frac{w(t)}{u(t)} = \frac{a \cdot b}{\tau} \quad (6)$$

Eq. 6 provides:

$$\left.\begin{array}{c} u(t) = u_0 e^{\left(a + \frac{1}{\tau}\right)t - p \int_0^t s(x)\, dx} \\ v(t) = u(t)\left(-\frac{a}{\tau} - q \cdot s(t) + \frac{p}{\tau} \cdot s(t) + p \cdot s'(t)\right) \; ; \; w(t) = u(t)\frac{a \cdot b}{\tau} \end{array}\right\} \quad (7)$$

Thus, the Lagrangian is given by Eqs. 4 and 7. Note that it is undetermined by the constant u_0. The canonical momentum must be defined to find the Hamiltonian:

$$\gamma = \frac{\partial L}{\partial \dot{y}} = u(t) \cdot \dot{y} \quad (8)$$

And the Hamiltonian, through its known formula, becomes:

$$H(t, y, \gamma) = \frac{\partial L}{\partial \dot{y}}\dot{y} - L(t, y, \dot{y}) = \frac{1}{2}\frac{\gamma^2}{u(t)} - \frac{1}{2}u(t) \cdot \bar{v}(t) \cdot y^2 - u(t)\frac{a \cdot b}{\tau}y + H_0 \quad (9)$$

after having used Eq. 8 to substitute $\dot{y}$ by γ, where $\bar{v}(t) = -\frac{a}{\tau} - q \cdot s(t) + \frac{p}{\tau} \cdot s(t) + p \cdot s'(t)$.

Note that the Hamiltonian found is not a conserved amount, due to, as it is known in the general case, $\frac{dH}{dt} = \frac{\partial H}{\partial t}$, but by Eq. 9: $\frac{\partial H}{\partial t} \neq 0$. Thus, $\frac{dH}{dt} \neq 0$. However, as a hypothesis, the Hamiltonian of Eq. 9 can be considered a non-conserved energy of personality as a consequence of a stimulus $s(t)$. Note again that it is undetermined by the constant u_0, and also by an additive constant added H_0, which help us to define the convenient energy value in the initial time t=0.

4. The application case

The application case consists in providing 20 mg of methylphenidate to an individual (male) of 54 years old. Methylphenidate is a stimulant drug whose dynamics can be modelled by a set of two coupled differential equations [7] as:

$$\left.\begin{array}{c} \frac{dm(t)}{dt} = -\alpha \cdot m(t) \\ m(0) = M \end{array}\right\} \qquad \left.\begin{array}{c} \frac{ds(t)}{dt} = \alpha \cdot m(t) - \beta \cdot s(t) \\ s(0) = s_0 \end{array}\right\} \tag{10}$$

In Eq. 10 $m(t)$ is the non-assimilated methylphenidate amount, M is the initial amount of methylphenidate of a single dose and α is the methylphenidate assimilation rate. In addition $s(t)$ represents the stimulus, i.e., the amount in organism of the methylphenidate non-consumed by cells, s_0 is the amount of methylphenidate present in organism before the dose intake, and β is the methylphenidate metabolizing rate. The time function $s(t)$ corresponding to the stimulus dynamics is obtained by integrating the system of Eq. 10:

$$s(t) = s_0 \mathrm{e}^{-\beta \cdot t} + \begin{cases} \frac{\alpha \cdot M}{\beta - \alpha}\left(\mathrm{e}^{-\alpha \cdot t} - \mathrm{e}^{-\beta \cdot t}\right): \ \alpha \neq \beta \\ \alpha \cdot M \cdot t \cdot \mathrm{e}^{-\alpha \cdot t}: \ \alpha = \beta \end{cases} \tag{11}$$

Note that Eq. 11 must be considered in the response model of Eq. 2. In addition, M=20 mg and $s_0 = 0$, due to the individual has not consumed methylphenidate for very long. The calibration of the response model for the individual GFP-FAS outcomes, observed every 7.5 minutes during 3 hours, consists in finding the optimal parameter values that minimize the square sum of the difference between the experimental values and the theoretical ones. The strength of the calibration is measured by the determination coefficient (R2). In addition, the residuals' randomness is provided by the p-value of the Anderson-Darling test, which reports if the residuals distribute as a N(0,std), i.e., as a Normal distribution of zero mean and constant standard deviation (std), being std the standard deviation of the residuals. Fig. 1 provides this calibration. Note that the visual observation provides a very good theoretical prediction of the response model given by Eqs. 2 and 11, supported by a determination coefficient close the unit (R2=0.94) and an Anderson-Darling's p-value=0.58 that confirms the residuals' randomness.

Fig. 2 represents the dynamics of the Hamiltonian given by Eq. 9. In it, the additive constant has been taken $H_0 = 4$ and $u_0 = 1$ in order to present its evolution in the range of positive values. Note that after a strong increase, it reaches its maximum and, after a slight decrease, it tends to a constant value.
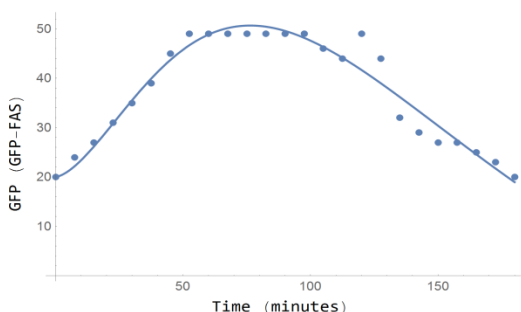
Fig. 1. General Factor of Personality response to 20 mg of methylphenidate intake versus time. Experimental values (dots) and theoretical values (line). R²=0.94. P-value=0.58.
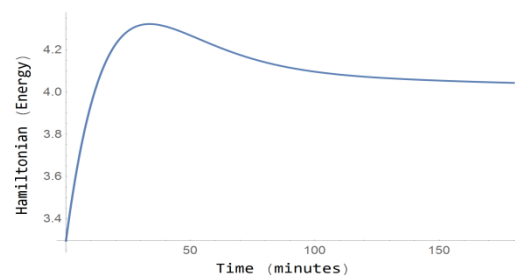
Fig. 2. Hamiltonian versus time with $H_0 = 4$ and $u_0 = 1$.

Conclusions

Fig. 1 confirms that the response model, given by Eqs. 2 and 11 (when a stimulant drug is being modelled) is suitable to reproduce the dynamical response of the GFP as a consequence of a dose intake of a stimulant drug. The works [7, 8] points out clearly the utility of the response model to reproduce the dynamics of the GFP and the changes in its biological bases. Also it is a tool to describe the personality change with the help of drugs and the support of the self-regulation therapy there provided.

In addition, the analytical approach to the problem given by the minimum action principle, the Langrangian and the Hamiltonian is a success theoretical approach because it connects the behavioral sciences with a consolidated formalism of physics. This connection represents an objective of General Systems Theory: provide to the human disciplines of the same epistemological status than positive sciences have. However, no utility has been found by the moment, which must be still investigated. However, the motivation of the authors is given by the search of an "energy function" to explain the human personality. If mathematically it has been found in the context of the short term dynamics response of the GFP to a stimulant drug, it must be also sought for other kinds of stimuli and for long term dynamics responses. In addition, the utility of the Hamiltonian could be speculated that consists in analyzing the Hamiltonian terms searching for the processes of sensitization and habituation of personality. Another use could consist in providing a quantum approach to brain.

References

[1] L. D. Landau, E. M. Lifshitz, Mecánica (Vol. 1 del Curso de Física Teórica), Ed. Reverté, 2002.

[2] P. A. M. Dirac, Lectures on Quantum Dynamics, Ed. Yeshiva University, 1964.

[3] S. Amigó, J.C. Micó, A. Caselles, Five adjectives to explain the whole personality: a brief scale of personality, Rev. Int. Sist. 16 (2009) 41–43.

[4] S. Amigó, A. Caselles, J.C. Micó, The General Factor of Personality Questionnaire (GFPQ): Only one factor to understand the personality?, Span. J. Psychol. (2010) 5–17.

[5] J.C. Micó, S. Amigó, A. Caselles, From the Big Five to the General Factor of Personality: a Dynamic Approach, Span. J. Psychol. 17 (2014) E74 1-18.

[6] J.C. Micó, S. Amigó, A. Caselles, Advances in the General Factor of Personality Dynamics, Revista Internacional de Sistemas 22 (2108) 34-38.

[7] J.C. Micó, S. Amigó, A. Caselles, Changing the General Factor of Personality and the c-fos Gene Expression with Methylphenidate and Self-Regulation Therapy, Span. J. Psychol. 15 (2012) 850–867.

[8] S. Amigó, A. Caselles, J.C. Micó, The self-regulation therapy to reproduce drug effects: a suggestion technique to change personality and the DRD3 gene expression, Int. J. Clin. Exp. Hypn. 61 (2013) 3 282–304.

A Pattern Recognition Bayesian Model for the appearance of Pathologies in Automated Systems

M.Alacreu$^{\flat}$, N.Montes$^{\flat}$ *, E.Garcia† and A.Falco$^{\flat}$

($\flat$) UniversityCEU Cardenal Herrera,

C/San Bartolome 55, Alfara del Patriarca, Valencia (Spain),

($\dagger$) Ford Spain,

Polgono industrial Ford S/N, Almussafes, (valencia),

November 30, 2018

1   Introduction

Most of the multitasking robots used in large factories are characterized by the performance of repetitive and highly accurate actions. Its correct operation is crucial to obtain an excellent, competitive and sustainable product. Often, machines act in a chain, so that the tasks that fall on a machine affect the operation of the next. During their lifespan, machine components suffer deterioration motivated by multiple pathologies that do not necessarily imply a breakdown. These deteriorations could be unnoticed during time and have negative consequences like for instance, energy waste, slowed down the production process, breakdowns in contiguous components, etc. With the aim of minimizing the impact on the cost of production that causes deterioration in automated systems, the present paper shows a pattern recognition Bayesian model able to detect the deviation that the machine has compared with correct behavior. This deviation is measured in terms of the technical

---

*e-mail: nicolas.montes@uchceu.es

cycle time of the machine. As a result, the posterior distribution of probabilities of each one of the considered machine pathologies is obtained, in a mixture of Gaussian distributions. As an example of the proposed methodology, a welding unit located at Ford factory in Almussafes was isolated and tested for some pathologies. These pathologies are in a proportional valve, a cylinder, an electrical transformer, the robot speed and the loss of pressure. As a result, and based on a real-time cycle time monitorization, a "ranking" of pathologies based on probabilities of its occurrence is established.

2 Previous works. From micro-term to long-term
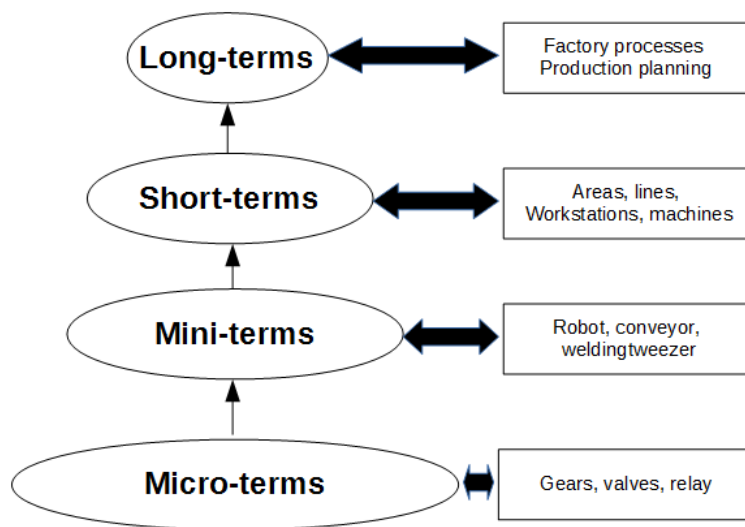
Figure 1: A pyramid of terms.

The data used in the analysis of the production lines is classified into *long-term* and *short-term*. *Long-term* is mainly used for process planning, while *short-term* focuses, primarily, on process control. Following the definition in [2], *short-term* is referred to an operational period not large enough for machine failure period to be described by a statistic distribution. The machine cycle time is considered *short-term*. In [3], [4] redefines *short-term* into two new terms, *mini-term* and *micro-term*, see figure 1. A *mini-term*

could be defined as a machine part, in a predictive maintenance policy or in a breakdown, replaceable easier and faster than another machine part subdivision. Furthermore, a *mini-term* could be defined as a subdivision that allows us to understand and study the machine behavior. These sub-cycle times (*micro-terms* and *mini-terms*) are not the same at each repetition and they follow a probabilistic distribution, mean value μ and standard deviation σ. In addition to that, the probabilistic sub-cycle time for each machine component varies during the lifespan of the component. In other words, the deterioration indicators that can be measured with thermal cameras, vibration and ultrasonic devices have an effect on the machine cycle time. In most cases, the measurement of these cycle times does not imply any additional costs because the actuators that allow the sub-cycle time measurement were installed in the machine and are used for their automated work.

3   A test bench for a welding station.

The behavior of the welding station, figure 2, is simple. First, the robot arm moves the welding clamp to the point to weld. Then, a pneumatic cylinder moves the welding clamp in two phases: One to bring closer the clamp and a second one to weld. The pressure applied by the clamp is controlled by a control system.
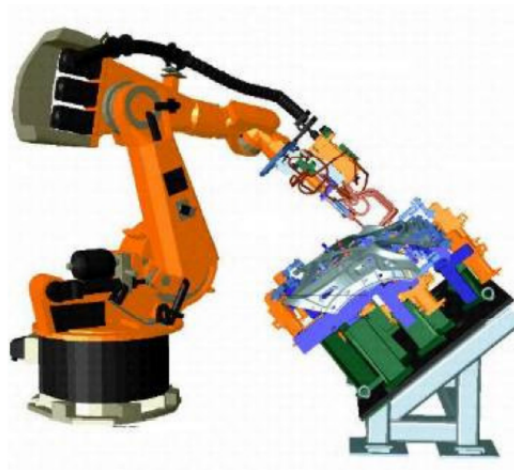
Figure 2: Welding station

The Robot Arm and Welding Clamp need a certain time to develop their

task and their components also need a certain time to develop their own tasks. In [5] the welding station was divided in three *mini-terms*, the robot arm, the welding clamp motion and the welding task. .

3.1 Pathologies analysed.

The welding station, as well as other stations in the industry, is bound to suffer from pathologies that produce an effect on the cycle time. Based on the operator's experience, In [5] were selected the most common ones for the experimental welding station. These pathologies produce a cycle time modification but do not produce failure of the component, going unnoticed for maintenance workers and also for the control system, in other words, after the change point and before the failure of the component, see Figure **??**. The pathologies rated are; for the welding clamp *mini-term*: the proportional valve, the cylinder stiffness, welding failure produced by the transformer and pressure loss, and for the robot arm *mini-term*; the robot arm speed.

3.2 Rules definition and Bayesian mixture model.

In [5] was analyzed the experimental samples to understand how the pathologies affect the cycle time and to generate rules that allow us to define it. The test developed without pathology is called as "C" (control test) and the behaviour with one of the pathologies $(P_1...P_5)$, that is, six different situations for each *mini-term*. It is obvious that the cycle time for the *mini-terms* with pathology are different compared with the control or bassal situation.

The statistical tests used in [5] was Shaphiro-Wilk, Levene, ANOVA, Kruskal-Wallis and a variance. For all the tests, the significance level is $\alpha = 0.05$. With this, statistical rules was obtained, allowing to diferenciate between pathologies. In Table 1 is shown a resume of the rules. the first two rows are rules that classify mean and variance values according to the pathology. Third row shows threshold values to determine if there are pathologies or not and the last row shows extra rules like for instance, when pathology 4 occurs, the data do not pass the normality test.

By means of these rules, A Bayesian model that mix the gausians is used to determine which pathology occurs in real-time. [6].

Table 1: Rules for the Knowledge-driven MSS. Welding station case.

Robot Motion *mini-terms*	
Mean rule	$\mu_C = \mu_{P1} = \mu_{P2} = \mu_{P3} = \mu_{P4} < \mu_{P5}$
Variance rule	$---$
Stand. desv rule.	$S > 25.4 \cdot 10^{-3} \rightarrow All$
Normality rule	$----$
Welding Motion *mini-terms*	
Mean rule	$\mu_C = \mu_{P5} < \mu_{P1} < \mu_{P3} < \mu_{P2} < \mu_{P4}$
Variance rule	$\sigma_C^2 = \sigma_{P1}^2 = \sigma_{P3}^2 = \sigma_{P5}^2 < \sigma_{P2}^2 < \sigma_{P4}^2$
Stand. desv. rule	$S \notin [47 \cdot 10^{-4} 74 \cdot 10^{-4}] \rightarrow P_2, P_4$
Normality rule	$P4 fail$
Welding task *mini-terms*	
Mean rule	$\mu_{P2} < \mu_{P4} < \mu_C = \mu_{P3} < \mu_{P5} < \mu_{P1}$
Variance rule	$\sigma_C^2 = \sigma_{P3}^2 = \sigma_{P5}^2 < \sigma_{P2}^2 = \sigma_{P4}^2 < \sigma_{P1}^2$
Stand. desv. rule	$S > 12.9 \cdot 10^{-3} \rightarrow P_1, P_2, P_4$
Normality rule	$P1 fail$

4   Conclusions and actual developments.

This paper shows how to design real-time Maintenance Support System to prognosticate breakdowns in production lines. The system is based on the sub-cycle time (*mini-terms*) monitorization, statistical analysis, learning of the data obtained for the real production lines, defining the density functions that govern the decisions, based on Bayesian model. The system is nowadays on an standarization process in a Ford Motor company where thousands of *mini-terms* and their pathologies are reported and analyzed. The system is well known as *mini-term 4.0*, see, [7].

References

[1] AutorA NameA, and AutorB NameB. Title of the paper *Name of the Journal*, Volume(Number):Initial Page–Final Page, Year.

[2] L.Li, Q.Chang, J.Ni.S.Biller: Real time production improvement through bottleneck control. *International Journal of production research*, 47(21):6145-6158, 2009

[3] E.Garcia: Análisis de los sub-tiempos de ciclo técnico para la mejora del rendimiento de las líneas de fabricación. PhD (2016).

[4] E.Garcia, N.Montes: A Tensor Model for Automated Production Lines based on Probabilistic Sub-Cycle Times. Nova Science Publishers, 18(1). 221-234, 2017

[5] E.Garcia, N.Montes,M.Alacreu: Towards a Knowledge-driven Maintenance Support System for Manufacturing Lines. In Proceedings of the 15th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2018) , 1 43-54, 2018.

[6] A. R.Webb, K.D.Copsey, Statistical Pattern Recognition. New Jersey, Willey, 2011.

[7] E.Garcia, N.Montes, M.Alacreu Towards a novel maintenance support systems based on *mini-terms. Mini-term 4.0.* Springer, 2019.

A study of the seasonal forcing in SIRS models for Respiratory Syncytial Virus (RSV) using a constant period of temporary immunity

L. Acedo, J. A. Moraño [*], and R. J. Villanueva

Instituto de Matemática Multidisciplinar. Universitat Politècnica de València

November 30, 2018

1 Introduction

Respiratory Syncytial Virus (RSV) is an acute respiratory infection that infects millions of children and infants worldwide. It is the major cause of their hospitalizations, especially for bronchiolitis and pneumonia [1] and its impact on health services is increasing [2]. The impact on adults also is studied because up to 18% of the pneumonia hospitalizations in patients older than 65 years are due to RSV [3].

The cost of pediatric hospitalization for the Valencian Health Service is about €3.5 million per year being the RSV the cause of annual seasonal epidemics with minor variations each year but its coincidence with other common viral infections such as influenza or rotavirus produces an overstretching of the health service.

Recent research clearly shows the possibility of developing a variety of effective vaccines on RSV. Some of those types are already in preclinical development or even in clinical trials. These vaccines might be available in the near future so, planning of vaccination strategies is urgently required.

Mathematical models are powerful tools to analyse the epidemiology of infectious illnesses, to understand their behaviour, to predict their social

---

[*]jomofer@mat.upv.es

impact and to discover how external actors change the impact of disease. In the case of RSV, the building of a reliable model is a priority objective to predict the medical care requirements needed in the following seasons [4].

Different mathematical models have been studied to simulate the propagation of RSV. For instance, Weber et al. [5] proposed and studied a classical SIRS and its variation MSEIRS (including maternal immunity and latent period). Some of SIRS models of RSV use differential equations [4, 5, 6, 7, 8, 9] and others are based on network models [10] or even the SIRS model has been studied from a Bayesian perspective [11].

Although the network model [4] achieved fitting the seasonality without that forcing, most of models reproduce the seasonality forcing it by means a periodic transmission function β varying with a cosine with higher transmission during winter months, such as appears in literature [4, 5]:

$$\beta(t) = b_0 + b_1 cos(2\pi t + \phi) \tag{1}$$

In this study, we consider this type of SIRS model but with fixed temporal immunity, w, as Brauer et al. proposes [12]. We analyze if the cosine coefficient, b_1, is small enough to remove the seasonal forcing when we fit the model to data. In order to fit the parameters of the model we use data from children hospitalizations in the Spanish Region of Valencia from November 2001 to 2004, which are the same ones as we used in the classical model [4].

2 Data

In this work we use the same data as were used in our previos work [4]:

- Weekly data between 2001 and 2004 (208 weeks) with the number of hospitalizations of children less than one year.

- Population data are obtained from the IVE (Valencian Institute of Statistics).

- We initially consider an infectivity period of 10 days [1, 4, 5] and a recovered period of 200 days [4, 5] to do comparisons with the results obtained in [4].

- In the same way we use at first the proportion of infected children who are hospitalized, $s = 0.022$, that was obtained in [4].

3 Model

Usually in SIRS models, individuals remain in the Recovered state for a period of time after which they return to susceptible class, but we consider a constant period of temporary immunity, w, following recovery from the infection, such as proposed Brauer in [12]. This type of SIRS model can have abiding oscillations even if the seasonal forcing in function β is not used.

3.1 Description

Therefore, in our model there is a temporary immunity period of fixed length, w, after which recovered infectives revert to the susceptible class, what means that w is a delay in the lost of immunity. See Figure 1.
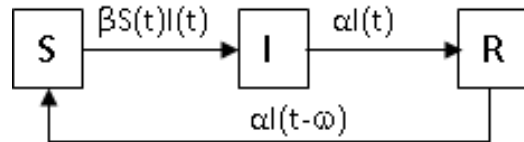
Figure 1: *Model with temporary immunity period of fixed length (w).*

If we use discrete-time evolution, the transmission dynamics of RSV can be modeled by only two equations [12]:

$$
\begin{aligned}
S(t+1) &= S(t) - \beta S(t)I(t) + \alpha I(t-w) \\
I(t+1) &= I(t) + \beta S(t)I(t) - \alpha I(t)
\end{aligned}
\tag{2}
$$

3.2 Two age groups and constant population

In our calculations we use constant population ($N = S + I + R$) and as well as the available data are only for children younger than one year we divide the population into two age groups: [0-1[and older than one year old.

In order to have an adequate distribution of the age groups we consider the data from IVE averaged between 2001 and 2004 having adjusted the mortality rate of the second group for maintaining of constant population.

3.3 Calibration

We have adjusted the parameters of the seasonal term (b_0, b_1, ϕ) and the proportion of hospitalized patients (s) to minimize the Root Mean Square Error (RSME) between data and the model output every week.

The Particle Swarm Optimization (PSO) algorithm has been used for this calibration [13, 14].

As initial conditions, it has been necessary to estimate 28 values of I_1 and I_2. The immunity rate α is the related to a period of immunity of 10 days (or 1.43 weeks) and as delay w in the loss of immunity we consider 28 weeks (200 days). In addition, we have left a transitory period of stabilization of the disease of 520 weeks before beginning the calibration.

4   Results

The first result has been obtained with the parameters $\alpha \sim 10$ days and $w = 28$ weeks that were already used in [4]. Values obtained are:

$$b_0 = 0.922; b_1 = 0.143; \phi = 1.627; s = 0.0339; RSME = 1.1645 \qquad (3)$$

and the graph representing the fitting is:
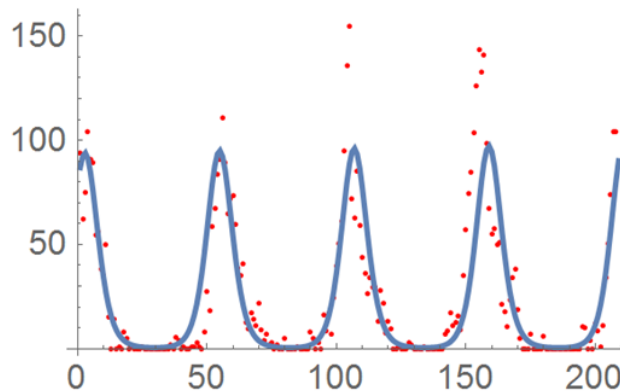
Figure 2: *Comparison between the weekly data of children under one year that were hospitalized by RSV (red dots) and the results obtained in the simulation made using the values $b_0 = 0.922; b_1 = 0.143; \phi = 1.627$ and $s = 0.0339$ (blue line) during the 208 weeks.*

The parameter b_1 is the coeficient that extends the periodic part of $\beta(t)$, then to study its relevance we must compare its proportion with b_0. If we

compare the new values to the ones obtained in [4] we can observe that the new b_1 is less relevant that the previous one because

$$\frac{b_1(0.143)}{b_0(0.922)} < \frac{b_1(14.31)}{b_0(69.52)}^{[4]}$$

On the other hand, one of the objectives of incorporating this delay was to see if the oscillations of the data can be reproduced by reducing the seasonal forcing of the transmission rate. So we have made new calibrations with $b_1 = 0$ but allowing a different infected period $\alpha \nsim 10$ days and a different period of losing immunity $w \neq 28$ weeks.

In this case we show two interesting results:

- The first one has been selected because it is capable of reproducing seasonal maxima very well although the RSME is larger than in the previous result

$$b_0 = 0.860; \alpha \sim 15 \text{ days}; w = 26 \text{ weeks}; s = 0.016; RSME = 1.4165 \quad (4)$$

and their graph representation is:
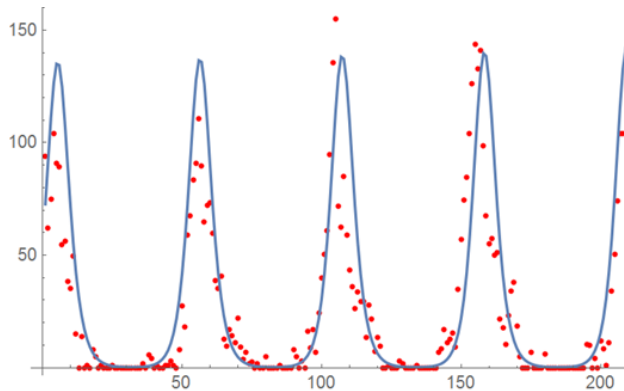
Figure 3: *Same as Figure 2 with $b_0 = 0.860; b_1 = 0.0; \alpha \sim 15$ days; $w = 26$ weeks and $s = 0.016$*

- The second result is the best one (minimum RSME) obtained with $b_1 = 0$ and without constraints for α and w:

$$b_0 = 0.865; \alpha \sim 14 \text{ days}; w = 26 \text{ weeks}; s = 0.012; RSME = 1.0574 \quad (5)$$
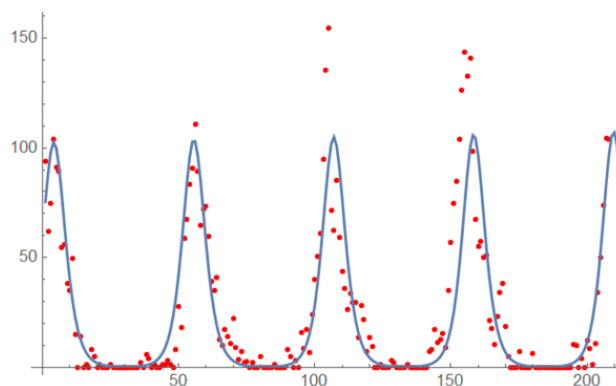
Figure 4: *Same as Figure 2 with* $b_0 = 0.865; b_1 = 0.0; \alpha \sim 14$ *days;* $w = 26$ *weeks and* $s = 0.012$

5   Conclusions

We have adapted the SIRS model of RSV [4] incorporating a temporary immunity period of fixed length (or delay) following the approach proposed by Brauer and Castillo-Chavez in [12].

- Periodic oscillations appear naturally in this model based on equations in differences:

  - With 28 weeks of immunity and 10 days as a period of infectivity we found good values of parameters where b_1 reduces its relevance.
  - Allowing other values for these parameters we can obtain a calibration of the model without seasonal forcing $(b_1 = 0)$.

- The seasonal forcing in these models can be reduced or even obviated by using a delay in the loss of immunity.

- This reduction in the seasonal forced also appeared in the network model [10] but not as a consequence of the delay in the loss of immunity.

- Oscillations in models do not appear related only to external reasons, they could be related to intrisic values of illness.

References

[1] CB. Hall, Textbook of Pediatric Infectious Diseases (Respiratory syncytial virus and human metapneumovirus). Philadelphia, PA, Saunders, 2004, pp. 2315–2341

[2] Langley JM., LeBlanc JC., Smith B. and Wang EEL. Increasing Incidence of Hospitalization for Bronchiolitis among Canadian Children, 1980–2000 *The Journal of Infectious Diseases*, Volume(188), Issue 11: 1764–1767, 2003. DOI: https://doi.org/10.1086/379740

[3] Han L., Alexander J. and Anderson L. Respiratory syncytial virus pneumonia among the elderly: An assessment of disease burden. *Journal of Infectious Diseases*, Volume(179): 25–30, 1999. DOI: https://doi.org/10.1086/314567

[4] Acedo L., Díez-Domingo J., Moraño JA. and Villanueva RJ. Mathematical modelling of respiratory syncytial virus (RSV): Vaccination strategies and budget applications. *Epidemiology and Infection*. Volume (138), Issue 6: 853–860, 2010. DOI: https://doi.org/10.1017/S0950268809991373

[5] Weber A, Weber M, and Milligan P. Modeling epidemics caused by respiratory syncytial virus (RSV). *Mathematical Biosciences* Volume 172: 95–113, 2001. DOI: https://doi.org/10.1016/S0025-5564(01)00066-9

[6] Hogan AB., Glass K., Moore HC. and Anderssen RS. Applications + Practical Conceptualization + Mathematics = Fruitful Innovation. Mathematics for Industry, vol 11 (Age Structures in Mathematical Models for Infectious Diseases, with a Case Study of Respiratory Syncytial Virus). Tokyo. Anderssen R. et al. (eds) Springer, 2016. DOI: https://doi.org/10.1007/978-4-431-55342-7_9

[7] Moore HC., Jacoby P., Hogan AB., Blyth CC. and Mercer GN. Modelling the Seasonal Epidemics of Respiratory Syncytial Virus in Young Children. *PLoS ONE* Volume (9) Issue 6: e100422, 2014. DOI: https://doi.org/10.1371/journal.pone.0100422

[8] Aranda-Lozano DF., González-Parra GC. and Querales J. Modelling respiratory syncytial virus (RSV) transmission children aged less than five years-old. *Rev Salud Publica*, Volume(15)-4: 689–700, 2013.

[9] Paynter S, Yakob L, Simões EAF, Lucero MG, Tallo V, Nohynek H et al. Using Mathematical Transmission Modelling to Investigate Drivers of Respiratory Syncytial Virus Seasonality in Children in the Philippines. *PLoS ONE*, Volume(9)-2: e90094, 2014. DOI: https://doi.org/10.1371/journal.pone.0090094

[10] Acedo L., Moraño JA. and Diez-Domingo J. Cost analysis of a vaccination strategy for respiratory syncytial virus (RSV) in a network model *Mathematical and Computer Modelling*, Volume(52): 1016–1022, 2010. DOI: https://doi.org/10.1016/j.mcm.2010.02.041

[11] Jornet-Sanz M., Corberán-Vallet A., Santonja FJ. and Villanueva RJ. A Bayesian stochastic SIRS model with a vaccination strategy for the analysis of respiratory syncytial virus. *SORT*, Volume(41)-1: 159–176, 2017. DOI: https://doi.org/10.2436/20.8080.02.56

[12] Brauer F. and Castillo-Chavez C., Mathematical Models in Population Biology and Epidemiology (Models for Endemic Diseases). Texts in Applied Mathematics, Volume(40). New York, Springer, 2012. DOI: https://doi.org/10.1007/978-1-4614-1686-9_10

[13] Khemka N. and Jacob C. Exploratory toolkit for evolutionary and swarm-based optimization *The Mathematical Journal*, Volume(11)-3: 376–391, 2010. DOI: https://doi.org/10.3888/tmj.11.3-5

[14] Acedo L., Burgos C., Hidalgo JI., Snchez-Alonso V., Villanueva RJ. and Villanueva-Oller, J. Calibrating a large network model describing the transmission dynamics of the human papillomavirus using a particle swarm optimization algorithm in a distributed computing environment. *The International Journal of High Performance Computing Applications* Volume(32)-5, 721–728, 2018. https://doi.org/10.1177/1094342017697862

Improving urban freight distribution through techniques of multicriteria decision making. An AHP-GIS approach

Victoria Muerza[1,2], Carina Thaller[3], Emilio Larrodé[1,4]

[1]ARAGON INSTITUTE OF ENGINEERING RESEARCH (i3A). Edificio de I+D+i, C/ Mariano Esquillor s/n – 50018, Zaragoza (Spain). Tel: +34 976 761888. vmuerza@unizar.es

[2]MIT INTERNATIONAL LOGISTICS PROGRAM. ZARAGOZA LOGISTICS CENTER. C/ Bari 55, Edificio Náyade 5 – 50197, Zaragoza (Spain). Tel: +34 976 077604. vmuerza@zlc.edu.es

[3]INSTITUTE OF TRANSPORT LOGISTICS, TU DORTMUND UNIVERSITY. Leonhard-Euler-Straße 2, Dortmund – 44227, Germany. Tel.: +49 231 755-8131*

[4]UNIVERSITY OF ZARAGOZA. Dpto. de Ingeniería Mecánica. C/ María de Luna, 3 – 50018. Zaragoza (Spain). Tel: +34 976 762319. elarrode@unizar.es

1. Introduction

Transportation is a key factor in the social and economic development of a country; it influences the activities of the inhabitants in terms of mobility, and has an impact on the activities of all economic sectors. Transportation is a complex domain, and involves several stakeholders and levels of decision-making processes. In addition, investments are capital-intensive and usually require long implementation delays.

This paper proposes a model based on the Analytic Hierarchy Process (AHP) for the selection of the best solution (type of Urban Distribution Center- UDC) to improve the distribution of goods in different urban environments. Three typologies of distribution centers are considered: (i) Large size UDCs, characterized by the existence of an own distribution electric vehicles fleet, the possibility to manual or semi- motorized distribution to perform the last mile delivery, and self-service collection (ii) Small size UDCs, characterized by a manual or semi- motorized distribution, and the availability of a self-service collection; and (iii) Automated or self-service UDCs. In addition, a Geographic Information System (GIS) is used to identify those urban areas where it is possible to locate the UDC according to its typology. Different layers are applied to make the decision (e.g. number of inhabitants, population dispersion, and commerce location).

2. Background

2.1. The location problem

There have been proposed some methodologies for the location selection's problem by using different approaches (e.g. Analytic Hierarchy Process, Topsis, Artificial Neural Network, Genetic Algorithm and Fuzzy Logic).

Thus, for example, Kayikci (2010) presents a model based on a combination of the fuzzy-analytic hierarchy process (AHP) and artificial neural networks (ANN) method to identify criteria in a framework of an empirical survey. Rao et al. (2015) integrate the economic, environmental and social dimensions of sustainability. Li et al. (2011) outline a hybrid method, which incorporates Axiomatic Fuzzy Set (AFS) and TOPSIS techniques into an evaluation process, in order to select competitive regions in logistics. Bozorgi-Amiri and Asvadi (2015) locate relief logistics centers using AHP. The study is focused on availability,

---

* Present address: German Aerospace Center (DLR), Institute of Transport Research. Deutsches Zentrum für Rutherfordstraße 2, 12489 Berlin. carina.thaller@dlr.de

risk, technical issues, cost and coverage criteria. More recently, Di Matteo et al. (2016) propose a multicriteria hybrid model AHP-Electre for optimal locating emergency operation centers.

The typology of logistics center in urban freight distribution is important. In this regard, according to Marcucci and Danielis (2008), the lack of knowledge on the cost structure, and the potential freight transport demand for the urban freight consolidation centre, are revealed as causes of failure of many European schemes.

2.2. The Analytic Hierarchy Process (AHP)

AHP is a multicriteria method used in decision-making processes. It considers a discrete number of alternatives which can be explicitly treated. AHP allows the consideration of multiple actors, factors, criteria and scenarios. The method considers four-steps (Saaty, 1980; 1994): (i) modeling, (ii) valuation, (iii) prioritization and (iv) synthesis. One of the characteristics of this approach is that it requires the translation of perceptions into numerical scales. For doing this, it can be used the Saaty's fundamental scale in pairwise comparisons (Saaty, 1980).

The priorities of the model ($w_i=1,\ldots,n$) can be obtained using different methods. One of the most used is the eigenvector problem, which considers the following expression:

$$Aw = \lambda_{max}w \qquad \sum_{i=1}^{n} w_i = 1 \tag{1}$$

where $A=(a_{ij})$ is the reciprocal pairwise comparison matrix, λ_{max} is the principal eigenvalue of A and w is the vector of priorities. The measure of inconsistency in judgements is obtained applying the *Consistency Index (CI)* (Saaty, 1980), expressed as:

$$CI = \frac{\lambda_{max} - n}{n-1} \tag{2}$$

where λ_{max} is the principal eigenvalue of the judgements matrix, and n its order. When the reciprocal comparison matrix is consistent $\lambda_{max} = n$, and $CI=0$. Saaty proposed the *Consistency Ratio (CR)* to normalize the measurement. It is calculated as:

$$CR = \frac{CI}{RI(n)} \tag{3}$$

where $RI(n)$ is the Random Consistency Index for matrices of order n, obtained by simulating 100,000 reciprocal matrices randomly generated (Aguarón and Moreno-Jiménez, 2003).

3. Development and assessment of the AHP model

A group of five experts in urban freight transport and urban mobility defined and assessed the elements of the model (decision by consensus). The problem was structured in three levels: Goal, criteria (4), subcriteria (18). The elements of the model can be seen in Figure 1. Local priorities and global priorities of the model can be seen in Table 1. All the paired comparison matrices presented acceptable inconsistencies (CR<0.10).
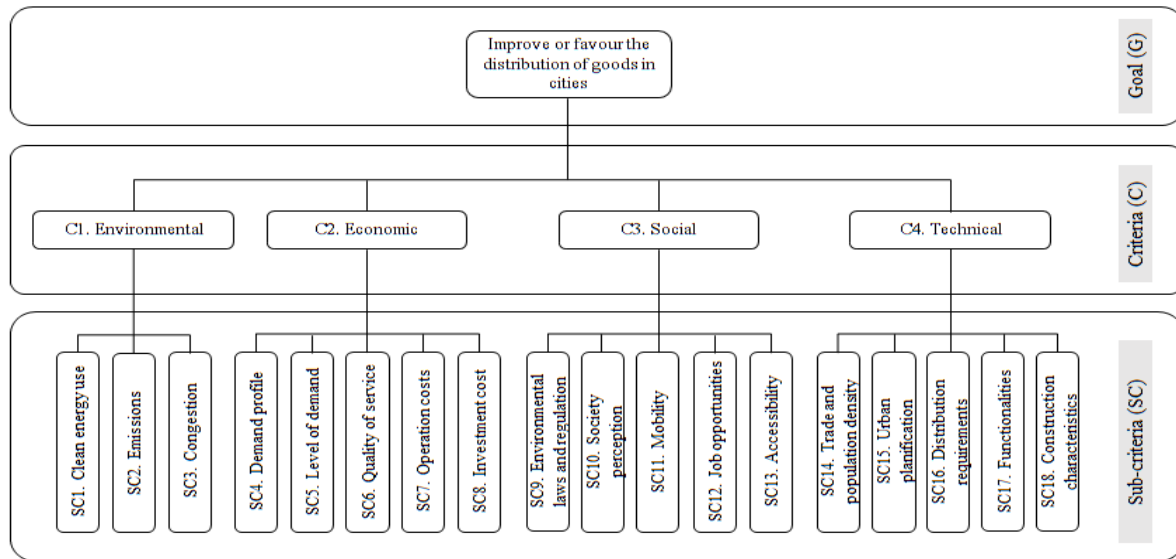
Figure 1. Elements of the proposed model

Table 1. Model prioritization

	C1. Environmental			C2. Economic					C3. Social					C4. Technical				
	L: 0.076; G: 0.076			L: 0.191; G: 0.191					L: 0.076; G: 0.076					L: 0.657; G: 0.657				
	SC1	SC2	SC3	SC4	SC5	SC6	SC7	SC8	SC9	SC10	SC11	SC12	SC13	SC14	SC15	SC16	SC17	SC18
L	0.075	0.324	0.602	0.151	0.104	0.264	0.387	0.094	0.059	0.498	0.168	0.056	0.219	0.321	0.078	0.258	0.219	0.125
G	0.006	0.025	0.046	0.029	0.020	0.050	0.074	0.018	0.004	0.038	0.013	0.004	0.017	0.211	0.051	0.169	0.144	0.082

The application of the model has been carried out by assessing three alternatives (A): (i) Large size UDCs, characterized by the existence of an own distribution electric vehicles fleet, the possibility to manual or semi- motorized distribution to perform the last mile delivery, and self-service collection (ii) Small size UDCs, characterized by a manual or semi- motorized distribution, and the availability of a self-service collection; and (iii) Automated or self-service UDCs. The valuation of the three alternatives was made by the same group of experts participating in the definition of the model based on their knowledge and expertise, and the city of Zaragoza, and specifically the "Las Fuentes" district was selected as case study. From a global point of view, it can be seen the total or final priorities of the alternatives: $w(A1) = 0.391$; $w(A2) = 0.263$; $w(A3) = 0.346$. The ranking of alternatives shows that Small size UDCs (A1>A3> A2) is the preferred alternative in terms of improvement of the distribution of goods in a city (Figure 2).

A1: Large size UDCs	.391
A2: Small size UDCs	.263
A3: Automated or self-s	.346

Figure 2. Final priorities of the analyzed alternatives

4. Location of terminals through GIS

The location of terminals has been performed through the application of the software QGIS, a free-of-charge Geographical Information System (GIS). The methodology followed consisted of the following steps:

- Step 1. Study of the district's demand "Las Fuentes".
- Step 2. Number of distribution centers: current situation versus future situation.
- Step 3. Determination of the location of distribution centers.

4.1. Study of the district's demand "Las Fuentes"

The demand for parcels in an area is influenced by many variables such as the age of the population, purchasing power, sex, immigration rate, commercial activity, among others. In this study we simplify and consider two factors: age of population (with high possibility to use e-commerce (15-65 years old) which highly determine demand), and commercial activity.

A Spanish delivery company provided us the data of the average volume delivery: 17,000 parcels distributed in small carriages; 2,000 larger parcels distributed by van; 2,000 parcels collected in the office at request of the customer; 3,000 large ordinary parcels collected in the office because they do not fit in the address box; and 6,000 collected in the office due to the absence of the customer at home. Thus, it can be estimated that the demand of the district "Las Fuentes" is 30,000 monthly parcels. Table 2 shows the main characteristics of the district. For the determination of the future demand it has been considered an increment of 25 %[2].

Table 2. Main characteristics district "Las Fuentes"

N. Inhabitants	Area (sqm)	Inhabitants/ha	Population use e-commerce	Monthly demand (parcels)	Monthly demand forecast (parcels)
42,610	6.31	67.52	27,527	30,000	37,500

4.2. Number of distribution centers: current situation versus future situation

Currently, the district "Las Fuentes" has only one large size UDC used for the delivery of parcels.

According to the results obtained in the AHP model, for this district A1 is the preferred alternative, very close to A3 (4.5 % difference). In addition, from the results obtained it can be estimated that a large size UDC has a capacity of 30,000 parcels/ month. According to this information it is intended to cover the future demand (7,500 extra parcels) by using Automated or self-service UDCs (HP). An HP has a capacity of 60 parcels/ month. Thus, it is estimated the use of 125 HP.

4.3. Determination of the location of distribution centers

The location of HPs will depend on the population with the highest possibility to use e-commerce and the concentration of commerce in the area. The population is concentrated in one area with high density (a higher number of HPs will be located) and one area with lower concentration. The typology of commerce selected for this study considers the possibility of transportation of products by means of parcels and lacks of own transportation service. Using the information provided by the City Council of Zaragoza four typologies of commerce have been selected (see Table 3) and can be depicted (see Figure 3): second hand, haberdasheries, bookstores and electronics. The current UDC is represented by a red star.

Table 3. Typology of commerce under study district "Las Fuentes"

Second hand	Haberdasheries	Bookstores	Electronics	Total
2	5	17	30	54

---

[2] https://ecommerce-news.es/comercio-electronico-espana-2017-71707

(a) "Las Fuentes " district   (b) Location of commerce and population distribution
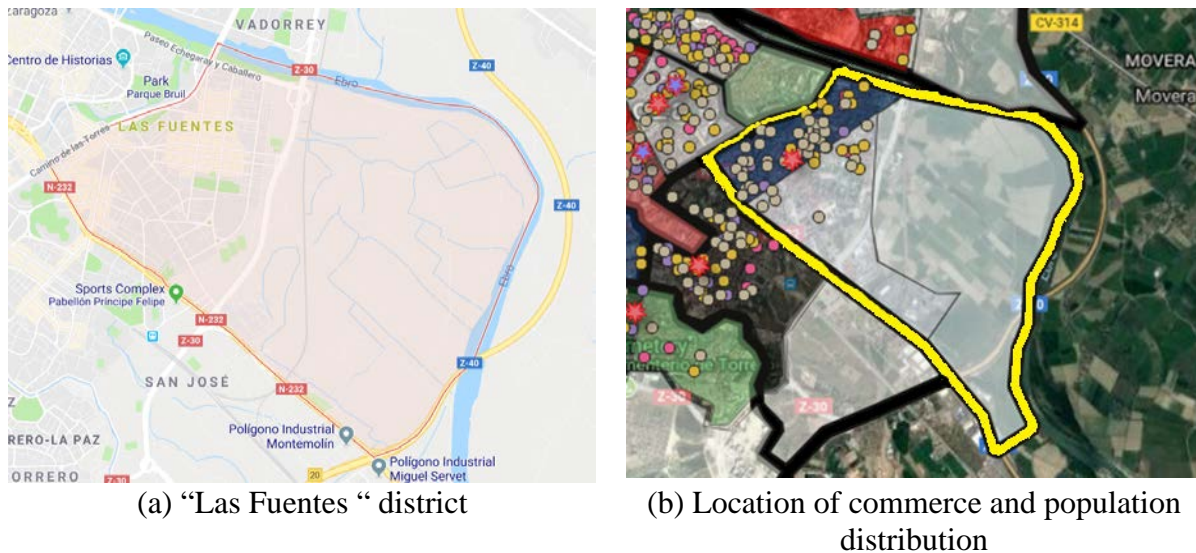
Figure 3. Main characteristics in terms of population and commerce of "Las Fuentes" district

Once it is know the quantity and location of commerce, the highest concentration of commerce can also be represented (see Figure 4, grated area).
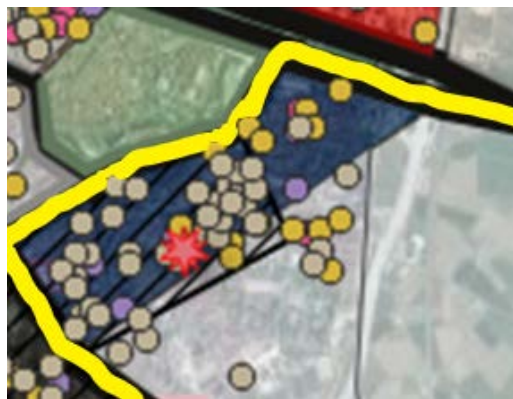
Figure 4. Higher concentration of commerce in "Las Fuentes" district

Due to the total number for the district is of 125 HPs, and a 90 % of the population is in this area, it is proposed to locate 113 HPs in this area.

5. Conclusions

This paper proposes an AHP-GIS approach for the location of UDCs according to three typologies: (i) Large size UDCs, characterized by the existence of an own distribution electric vehicles fleet, the possibility to manual or semi- motorized distribution to perform the last mile delivery, and self-service collection (ii) Small size UDCs, characterized by a manual or semi- motorized distribution, and the availability of a self-service collection; and (iii) Automated or self-service UDCs. The prioritization of alternatives is analysed for a district in the city of Zaragoza. In addition, a GIS is proposed for its location. Two main criteria has been used: (i) age of population (with high possibility to use e-commerce (15-65 years old) which highly determine demand), and (ii) commercial activity. This procedure allows the calculation of the number of UDCs to locate to comply with the future demand (considering an increment of 25 %) and the distribution area. Future research will focus on determining the place (shop, mall, bus stop…).

REFERENCES

Aguarón, J., and J. M. Moreno-Jiménez. 2003. The Geometric Consistency Index: Approximated Thresholds. *European Journal of Operational Research*, 147 (1), pp. 137–145.

Bozorgi-Amiri, A. and Asvadi, S. (2015). A prioritization model for locating relief logistic centers using analytic hierarchy process with interval comparison matrix. *Knowledge-Based Systems*, Vol. 86, pp.173–181.

Kayikci, Y. (2010). A conceptual model for intermodal freight logistics centre location decisions. *Procedia Social and Behavioral Sciences*. Vol. 2, pp. 6297–6311.

Li, Y., Liu, X., and Chen, Y. (2011). Selection of logistics center location using Axiomatic Fuzzy Set and TOPSIS methodology in logistics management. *Expert Systems with Applications*, Vol. 38, pp. 7901–7908

Marcucci, E., and Danielis, R. (2008). The potential demand for a urban freight consolidation centre. *Transportation*, Vol. 35 (2), pp. 269-284.

Matteo, U. D., Pezzimenti, P.M. and Garcia, D.A. (2016). Methodological Proposal for Optimal Location of Emergency Operation Centers through Multi-Criteria Approach. *Sustainability*, Vol. 8(1), 50; doi:10.3390/su8010050

Rao, C., Goh, M., Zhao, Y., and Zheng, J. (2015). Location selection of city logistics centers under sustainability. *Transportation Research Part D*, Vol. 36, pp. 29–44.

Saaty, T.L. (1980). *The Analytic Hierarchy Process*. McGraw-Hill, New York.

Saaty, T.L. (1994). *Fundamentals of Decision Making and Priority Theory with the Analytic Hierarchy Process*. RWS Publications, Pittsburgh, PA.

Nonlinear transport through thin heterogeneous membranes

Adrian Muntean [*]

Department of Mathematics and Computer Science, Karlstad University, Sweden.

October 22, 2018

1 Introduction

Our main motivation is to develop multiscale mathematical modelling strategies of gas transport processes through paperboard that can describe, on several space scales, how internal structural features and local defects affect the permeability of the material, perceived as a thin long permeable membrane.

To this end, we study the diffusion of particles through a thin heterogenous membrane under a one–directional nonlinear drift. Using mean-field equations derived from a Monte Carlo lattice dynamics for the problem at hand (for details, see [2]), we study the possibility to upscale the system and to compute the effective transport coefficients accounting for the presence of the membrane, adding this way analytic results to our simulation study [3]. For a special scaling regime, we perform a simultaneous homogenization asymptotics and dimension reduction, allowing us not only to replace the heterogenous membrane by an homogeneous obstacle line, but also to provide the effective transmission conditions needed to complete the upscaled model equations. The heterogeneities we account for in this context are assumed to be arranged periodically, but the same methodology can be adapted to cover also the locally periodic case. As working techniques, we employ scaling arguments as well as two-scale homogenization asymptotics expansions

---

[*]e-mail: adrian.muntean@kau.se

to guess the structure of the model equations and the corresponding effective transport coefficients.

The research presented here goes on the line open by M. Neuss-Radu and W. Jäger in [5] by adding to the discussion the presence of nonlinear transport terms and is remotely related to our work on filtration combustion through heterogeneous thin layers; compare [4].

2   Results

We apply simultaneously two conceptually different limiting processes – a periodic homogenization upscaling designed for a thin layered composite material and the dimension reduction of this layer to a sharp interface. Depending strongly on the choice of the microstructure model, a typical result of this procedure is a macroscopic model with nonlinear transmission boundary conditions. Hinting to the results reported in [1], a typical outcome is the following reduced upscaled model:

Find the triplet $(U^{\mathrm{l}}, u_0^{\mathrm{m}}, U^{\mathrm{r}})$ satisfying the following set of mass-balance equations:

$$\frac{\partial U^{\mathrm{i}}}{\partial T} - \nabla \cdot D^{\mathrm{i}}[\nabla U^{\mathrm{i}} + G(U^{\mathrm{i}})] = F^{\mathrm{i}} \ \text{ in } \ \Omega_{\mathrm{i}} \ (i \in \{\mathrm{l, r}\}), \qquad (1)$$

where Ω_{i} are two domains separated by a sharp interface (the flat support of the collapsed thin layer the thin layer). "Inside" the sharp interface, in the lower dimensional domain it holds

$$\frac{\partial u_0^{\mathrm{m}}}{\partial T} - \nabla_{y_2} \cdot D^{\mathrm{m}}[\nabla_{y_2} u_0^{\mathrm{m}} + G(u_0^{\mathrm{m}})] = F^{\mathrm{m}}. \qquad (2)$$

Further, boundary and initial conditions complete the model equations, viz.

$$u_0^{\mathrm{m}} \text{ is periodic in } y_2, \qquad (3)$$

$$u_0^{\mathrm{m}}(z_2, y_2, T) = U^{\mathrm{i}}(0, z_2, T), \ \text{ for i = l,r and } u_0^{\mathrm{m}}(z_2, y_2, 0) = V^{\mathrm{m}}(X_1, z_2), \qquad (4)$$

$$-D^{\mathrm{l}}(\nabla U^{\mathrm{l}} + G(U^{\mathrm{l}}) \cdot n = \varnothing_{\Gamma_{\mathrm{l}}} \ D_{11}^{\mathrm{m}} \frac{\partial u_0^{\mathrm{m}}}{\partial z_1} + D_{12}^{\mathrm{m}} \frac{\partial u_0^{\mathrm{m}}}{\partial y_2} + D_{11}^{\mathrm{m}} g(u_0^{\mathrm{m}}), \qquad (5)$$

$$-D^{\mathrm{r}}(\nabla U^{\mathrm{r}} + G(U^{\mathrm{r}})) \cdot n = \varnothing_{\Gamma_{\mathrm{r}}} -D_{11}^{\mathrm{m}} \frac{\partial u_0^{\mathrm{m}}}{\partial z_1} - D_{12}^{\mathrm{m}} \frac{\partial u_0^{\mathrm{m}}}{\partial y_2} - D_{11}^{\mathrm{m}} g(u_0^{\mathrm{m}}), \qquad (6)$$

$$U^{\mathrm{l}}(X,T) = u_{\mathrm{l}} \quad \text{on } \Gamma_{\mathrm{v}} \cap \Gamma_{\mathrm{l}} \quad \text{and} \quad U^{\mathrm{r}}(X,T) = u_{\mathrm{r}} \quad \text{on } \Gamma_{\mathrm{v}} \cap \Gamma_{\mathrm{r}}, \quad (7)$$

$$J^{\mathrm{i}}(X,T) \cdot n = 0 \quad \text{on } \Gamma_{\mathrm{h}} \cap \Omega_{\mathrm{i}} \quad \text{for } \mathrm{i} = \mathrm{l,r}, \quad (8)$$

$$U^{\mathrm{i}}(X,0) = V^{\mathrm{i}}(X) \quad \text{in } \Omega_{\mathrm{i}} \quad \text{for } \mathrm{i} = \mathrm{l,r}. \quad (9)$$

The coupling between the equations (1) and (2) is done via the micro-macro boundary conditions (4), (5), and (6). Note that the structure of the nonlinearity in the transmission conditions (5) and (6) depends on both the initial geometry of the thin later as well as of the balance laws incorporated in the microscopic model.

Currently, we work on estimating the quality of such nonlinearly coupled model from a multiple points of view: mathematical analysis, multiscale approximation perspective as well as validity with respect to the physical transport scenario supposed to be described.

3   Collaboration

References

[1] E. N. M. Cirillo, I. de Bonis, A. Muntean, O. Richardson, "Driven particle flux through a membrane: Two-scale asymptotics of a diffusion equation with polynomial drift.", *arXiv*:1804.08392 [math.AP] (2018).

[2] E.N.M. Cirillo, O. Krehel, A. Muntean, R. van Santen, A. Sengar, "Residence time estimates for asymmetric simple exclusion dynamics on strips." *Physica A* **442**, 436–457 (2016).

[3] E.N.M. Cirillo, O. Krehel, A. Muntean, R. van Santen, "A lattice model of reduced jamming by barrier." *Physical Review E* **94**, 042115 (2016).

[4] E.R. Ijioma, T. Ogawa, A. Muntean, T. Fatima, "Homogenization and dimension reduction of filtration combustion in heterogeneous thin layers". *Networks and Heterogeneous Media* **9**, 4, 709–737 (2014).

[5] M. Neuss-Radu, W. Jäger, "Effective transmission conditions for reaction-diffusion processes in domains separated by an interface". *SIAM Journal of Mathematical Analysis* **9**, 4, 709–737 (2007).

Application of the transfer matrix method for modelling Cardan mechanism of a real vehicle

Petr Hrubý[♭] , Tomáš Náhlík[†] [*]

(♭) Department of Mechanical Engineering, Institute of Technology and Business,

Okružní 517/10, České Budějovice, Czech Republic,

(†) Department of Informatics and Natural Sciences, Institute of Technology and Business,

Okružní 517/10, České Budějovice, Czech Republic.

November 30, 2018

1   Introduction

In engineering constructions, the most endangered parts are rotating components. Reliability of shaft endangers in particular two limit states. In the vicinity of resonance there is an enormous increase in the amplitudes of the state variables and the achievement of the yield strength of the material. These conditions often occur with the coupling shafts of cardan mechanisms. The torque is transmitted here over long distances. Shafts are long and slender and are prone to transverse bending. The gearbox shafts are compact and operate at a sufficient distance from the resonant area. In this case, they are threatened by fatigue fractures; they need to be checked for safety to fatigue. A similar situation to gearboxes is with gear pump shafts. The authors have long-term cooperation with engineering companies in questions of design and modelling of joint shafts and gear pumps. Mathematical models of gear pumps lead to solutions from the field of linear algebraic equations. In the case of bending oscillations, the motion equation of the basic element is a partial differential equation of the 4th order for the variables x and t. An

---
[*]e-mail: nahlik@mail.vstecb.cz

analytical solution for simpler cases can be used. A real shaft profile must be resolved by using one of the most sophisticated methods. In the cases solved by us, the finite element method and transfer matrix method proved to be successful. The transfer matrix method does not increase the matrix size (matrix 4x4 for planar oscillation, 8x8 for spatial oscillation), resulting in lower hardware requirements. Transfer matrix method uses a combination of analytical and numerical methods. Transfer matrix method also gives us the possibility to calculate the deformations caused by external excitation and dynamic deformation and stress analysis. Using the transfer matrix method is relatively easy to obtain a solution of the whole system (the whole Cardan mechanism). Another advantage is that it can be combined with the method of the imaginary slice which analytically solve the differential equations of motion for a smooth shaft (smooth continuum - a constant diameter), we derive the transfer matrices for shaft, matrices of concentrated mass and the elastic bearing, which are the basic structural elements of a dynamic model of shafts. The aim of our work is to provide designers of small mediums and businesses with whom we work together for a long time and who do not have specialized hardware and software equipment to facilitate their work. The aim of our work is to provide designers of small mediums and businesses with whom we work together for a long time and who do not have specialized hardware and software equipment to facilitate their work. Our aim is to provide tools to help engineers, which are working in small and medium-sized companies, with which we co-operate and who do not have specialized hardware and software, with their problems. An exemplary demonstration of the application of the transfer matrix method is the solution of the dynamic deformation analysis of the Cardan coupling shaft in the real vehicle drive, where permanent bending deformations occurred.

2   Transfer Matrix Method

Transfer Matrix Method (TMM) is combination of numerical and analytical methods and comes from the exact analytical solution (PDE of 4th order). Advantage of TMM is that it does not increase size of matrix (Planar oscillations - 4x4 matrix, Spatial oscillations - 8x8 matrix). That means lower hardware requirements. It allows us to calculate the deformation caused by external excitation, deformation caused by dynamics and also to make stress analysis. TMM can be used for solving of the whole system (whole Cardan

Mechanism).

It is necessary to derive transfer matrices for each of basic structural elements: shaft **H**, concentrated mass **M** and elastic bearing **K**.

We need to define vector of state $\mathbf{V}_i$ on the edge cuts of each elements based on amplitudes of state variables.
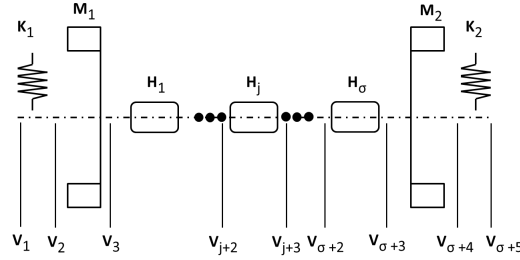
Vector of state:

$$\mathbf{V}(x,t) = \mathbf{V}(x)e^{i\bar{\omega}t}, \mathbf{V}(x) = [\mathbf{Y}(x)|\mathbf{Z}(x)]$$

$$\mathbf{Y}(x) = \begin{bmatrix} y(x) \\ y'(x) \\ -M_z(x) \\ -Q_y(x) \end{bmatrix}, \mathbf{Z}(x) = \begin{bmatrix} z(x) \\ z'(x) \\ -M_y(x) \\ -Q_z(x) \end{bmatrix}$$

,

where $y(x)$ is the amplitude of deflection, $y'(x)$ is slope of deflection, $M_z(x)$ is the amplitude of the bending moment and $Q_y(x)$ is moving force.

2.1   Derivation of transfer matrix

$\mathbf{V}_{i+1} = \mathbf{H}_j\mathbf{V}_i$, where $j = 1, \ldots, \sigma, i = j + 2$, $\mathbf{V}_2 = \mathbf{K}_1\mathbf{V}_1, \mathbf{V}_3 = \mathbf{M}_1\mathbf{V}_2 \ldots$, $\mathbf{V}_f = \mathbf{P} \cdot \mathbf{V}_1$, where $\mathbf{P}$ is the Transfer Matrix

$$\mathbf{P} = \mathbf{K}_2 \cdot \mathbf{M}_2 \cdot \mathbf{H}_\sigma \ldots \mathbf{H}_j \ldots \mathbf{H}_1 \cdot \mathbf{M}_1 \cdot \mathbf{K}_1$$

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_y & 0 \\ 0 & \mathbf{P}_z \end{bmatrix}, \mathbf{P}_y = \mathbf{P}_z = [p_{ij}]_1^4$$

After entering the edge vectors, transfer matrix and matrix multiplication

$$\mathbf{A}_y \cdot \mathbf{B}_y = \mathbf{D}_y$$

$$\mathbf{A}_z \cdot \mathbf{B}_z = \mathbf{D}_z$$

2.2   Solution of the set of equation

$$\mathbf{A}_y = \mathbf{A}_z = \begin{bmatrix} p_{11} & p_{12} & -1 & 0 \\ p_{21} & p_{22} & 0 & -1 \\ p_{31} & p_{32} & 0 & 0 \\ p_{41} & p_{42} & 0 & 0 \end{bmatrix}$$

$$\mathbf{B}_y = \begin{bmatrix} y_1(0) \\ y_1'(0) \\ y_0(l_0) \\ y_0'(l_0) \end{bmatrix}, \mathbf{B}_z = \begin{bmatrix} z_1(0) \\ z_1'(0) \\ z_0(l_0) \\ z_0'(l_0) \end{bmatrix}$$

$$\mathbf{D}_y = M_{1z} \cdot \begin{bmatrix} p_{13} \\ p_{23} \\ p_{34} - M_{2z}/M_{1z} \\ p_{43} \end{bmatrix}, \mathbf{D}_z = M_{1y} \cdot \begin{bmatrix} p_{13} \\ p_{23} \\ p_{34} - M_{2y}/M_{1y} \\ p_{43} \end{bmatrix}$$

Solution of left edge of the joint shaft

$$y_1(0) = \frac{p_{42}\left(p_{33}M_{1z} - M_{2z}\right) - p_{32}p_{43}M_{1z}}{p_{31}p_{42} - p_{32}p_{41}}$$

$$y_1'(0) = \frac{p_{31}p_{43}M_{1z} - p_{41}\left(p_{33}M_{1z} - M_{2z}\right)}{p_{31}p_{42} - p_{32}p_{41}}$$

$$z_1(0) = \frac{p_{42}\left(p_{33}M_{1y} - M_{2y}\right) - p_{32}p_{43}M_{1y}}{p_{31}p_{42} - p_{32}p_{41}}$$

$$z_1'(0) = \frac{p_{31}p_{43}M_{1y} - p_{41}\left(p_{33}M_{1y} - M_{2y}\right)}{p_{31}p_{42} - p_{32}p_{41}}$$

2.3   Transfer Matrix for Shaft

$$\mathbf{H}(x) = \begin{bmatrix} \mathbf{H}_y(x) & 0 \\ 0 & \mathbf{H}_z(x) \end{bmatrix}, \mathbf{H}_y = \mathbf{H}_z = [\mathbf{H}_{11}|\mathbf{H}_{12}|\mathbf{H}_{13}|\mathbf{H}_{14}]$$

$$\mathbf{H}_{11} = \frac{1}{\beta_1^2 + \beta_2^2} \begin{bmatrix} \beta_2^2 \cosh \beta_1 l + \beta_1^2 \cos \beta_2 l \\ \beta_1 \beta_2 \left(\beta_2 \sinh \beta_1 l - \beta_1 \sin \beta_2 l\right) \\ EJ\beta_1^2\beta_2^2 \left(\cosh \beta_1 l - \cos \beta_2 l\right) \\ EJ\beta_1^2\beta_2^2 \left(\beta_1 \sinh \beta_1 l + \beta_2 \sin \beta_2 l\right) \end{bmatrix}$$

$$\mathbf{H}_{12} = \frac{1}{\beta_1^2 + \beta_2^2} \begin{bmatrix} \beta_2^2/\beta_1 \sinh \beta_1 l + \beta_1^2/\beta_2 \sin \beta_2 l \\ \beta_2^2 \cosh \beta_1 l + \beta_1^2 \cos \beta_2 l \\ EJ\beta_1\beta_2 \left(\beta_2 \sinh \beta_1 l - \beta_1 \sin \beta_2 l\right) \\ EJ\beta_1^2\beta_2^2 \left(\cosh \beta_1 l - \cos \beta_2 l\right) \end{bmatrix}$$

$$\mathbf{H}_{13} = \frac{1}{\beta_1^2 + \beta_2^2} \begin{bmatrix} 1/EJ \left(\cosh \beta_1 l - \cos \beta_2 l\right) \\ 1/EJ \left(\beta_1 \sinh \beta_1 l + \beta_2 \sin \beta_2 l\right) \\ \beta_1^2 \cosh \beta_1 l + \beta_2^2 \cos \beta_2 l \\ \beta_1^3 \sinh \beta_1 l - \beta_2^3 \sin \beta_2 l \end{bmatrix}$$

$$\mathbf{H}_{14} = \frac{1}{\beta_1^2 + \beta_2^2} \begin{bmatrix} 1/EJ \left(1/\beta_1 \sinh \beta_1 l - 1/\beta_2 \sin \beta_2 l\right) \\ 1/EJ \left(\cosh \beta_1 l - \cos \beta_2 l\right) \\ \beta_1 \sinh \beta_1 l + \beta_2 l \\ \beta_1^2 \cosh \beta_1 l + \beta_2^2 \cos \beta_2 l \end{bmatrix}$$

where:

$$J = \frac{\pi}{4} \left(r_2^4 - r_1^4\right)$$

$$\beta_1 = \left\{ -\frac{\rho}{2E} \left(\overline{\omega^2} - \omega^2\right) + \left[\frac{\rho^2}{4E^2} \left(\overline{\omega^2} - \omega^2\right)^2 + \frac{4\rho \left(\overline{\omega} - \omega\right)^2}{E \left(r_2^2 + r_1^2\right)}\right]^{\frac{1}{2}} \right\}^{\frac{1}{2}}$$

$$\beta_2 = \left\{ \frac{\rho}{2E} \left(\overline{\omega^2} - \omega^2\right) + \left[\frac{\rho^2}{4E^2} \left(\overline{\omega^2} - \omega^2\right)^2 + \frac{4\rho \left(\overline{\omega} - \omega\right)^2}{E \left(r_2^2 + r_1^2\right)}\right]^{\frac{1}{2}} \right\}^{\frac{1}{2}}$$

2.4 Transfer matrix for concentrated mass

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_y & 0 \\ 0 & \mathbf{M}_z \end{bmatrix}$$

$$\mathbf{M}_y = \mathbf{M}_z = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -J_1\omega^2 + (J_0 - J_1)\omega^2 & 1 & 0 \\ m\left(\overline{\omega} + \omega\right)^2 & 0 & 0 & 1 \end{bmatrix}$$

2.5 Transfer matrix for elastic bearing

$$\mathbf{K}(x) = \begin{bmatrix} \mathbf{K}_y & 0 \\ 0 & \mathbf{K}_z \end{bmatrix}$$

$$\mathbf{K}_y = \mathbf{K}_z = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -k & 0 & 0 & 1 \end{bmatrix}$$

3   Conclusion

- We are able to calculate vector of state in every part of the shaft based on known physical characteristics

- Calculation of transfer matrix **P** and all other amplitude-frequency characteristics of state quantity are done in Octave 4.2.0

- By application of this method we are able to assess the resistance of the shaft to transverse oscillation during the design

4   Acknowledgement

The RVT method to solve random non-autonomous second-order linear differential equations about singular-regular points

J.-C. Cortés[♭] [*], A. Navarro-Quiles[†],
J.-V. Romero[♭] and M.-D. Roselló[♭].

(♭) Instituto Universitario de Matemática Multidisciplinar,

Universitat Politècnica de València, Spain,

(†) DeustoTech, Fundación Deusto

Univesidad de Deusto, Bilbao, Spain

1 Introduction

In this contribution we solve, from a probabilistic point of view, the following second-order random linear differential equation

$$\left.\begin{array}{l} X''(t) + p(t; A)X'(t) + q(t; A)X(t) = 0, \\[2mm] X(t_1) = Y_0, \quad X'(t_1) = Y_1, \quad t > t_1 > t_0 \in \mathbb{R}, \end{array}\right\} \tag{1}$$

where A, Y_0 and Y_1 are assumed to be absolutely continuous dependent random variables (RVs) defined on a common complete probability space, $(\Omega, \mathfrak{F}, \mathbb{P})$. Notice that, in IVP (1) t_0 is a singular-regular point and t_1 belongs to a neighbourhood of t_0, $t_1 \in \mathcal{N}(t_0)$.

---

[*]e-mail: jccortes@imm.upv.es

The results obtained in this contribution are a continuation of those established in a previous work [1], where approximations of the first probability density function (1-PDF) of the solution SP of the second-order random differential equation (1) about an ordinary or regular point are computed. Then, in this work the objective is to obtain an expression to approximate the 1-PDF of the solution SP of the IPV (1). The 1-PDF gives us a full probabilistic description of the solution, $X(t)$, in every instant time t. In addition, from it some interesting properties of the SP can be derived, like the mean and the variance,

$$\mathbb{E}\left[X(t)\right] = \int_{\mathbb{R}} x f_1(x,t)\mathrm{d}x, \quad \mathbb{V}\left[X(t)\right] = \int_{\mathbb{R}} x^2 f_1(x,t)\mathrm{d}x - \mathbb{E}\left[X(t)\right]^2.$$

The 1-PDF also allows to obtain confidence intervals. Additionally, the asymmetry and the kurtosis functions can be obtained from the 1-PDF too since any higher one-dimensional moment of the solution SP can be calculated via the 1-PDF:

$$\mathbb{E}\left[X^k(t)\right] = \int_{\mathbb{R}} x^k f_1(x,t)\mathrm{d}x, \quad k = 1, 2, \ldots$$

The key tool to achieve the aforementioned goals is the Random Variable Transformation (RVT) technique. In the multi-dimensional version, this result is stated in Theorem 1.

Theorem 1 (Multidimensional RVT method) *[2, p.25].* *Let us consider* $\mathbf{X} = (X_1, \ldots, X_n)^\mathsf{T}$ *and* $\mathbf{Z} = (Z_1, \ldots, Z_n)^\mathsf{T}$ *two n-dimensional absolutely continuous random vectors defined on a probability space* $(\Omega, \mathfrak{F}, \mathbb{P})$. *Let* $\mathbf{r} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ *be a one-to-one deterministic transformation of* $\mathbf{X}$ *into* $\mathbf{Z}$, *i.e.,* $\mathbf{Z} = \mathbf{r}(\mathbf{X})$. *Assume that* $\mathbf{r}$ *is continuous in* $\mathbf{X}$ *and has continuous partial derivatives with respect to each* X_i, $1 \leq i \leq n$. *Then, if* $f_{\mathbf{X}}(\mathbf{x})$ *denotes the joint probability density function of the random vector* $\mathbf{X}$, *and* $\mathbf{s} = \mathbf{r}^{-1} = (s_1(z_1, \ldots, z_n), \ldots, s_n(z_1, \ldots, z_n))^\mathsf{T}$ *represents the inverse mapping of* $\mathbf{r} = (r_1(x_1, \ldots, x_n), \ldots, r_n(x_1, \ldots, x_n))^\mathsf{T}$, *the joint probability density function of random vector* $\mathbf{Z}$ *is given by*

$$f_{\mathbf{Z}}(\mathbf{z}) = f_{\mathbf{X}}\left(\mathbf{s}(\mathbf{z})\right)|J|,$$

where $|J|$, *which is assumed to be different from zero, is the absolute value of*

the Jacobian defined by the determinant

$$J = \det\left(\frac{\partial \mathbf{s}^{\mathsf{T}}}{\partial \mathbf{z}}\right) = \det \begin{pmatrix} \frac{\partial s_1(z_1,\ldots,z_n)}{\partial z_1} & \cdots & \frac{\partial s_n(z_1,\ldots,z_n)}{\partial z_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial s_1(z_1,\ldots,z_n)}{\partial z_n} & \cdots & \frac{\partial s_n(z_1,\ldots,z_n)}{\partial z_n} \end{pmatrix}.$$

2   Computing the 1-PDF

By the Fröbenious method [3], the solution SP of the IVP (1) can be written as

$$X(t) = K_0(A, Y_0, Y_1)X_1(t; A) + K_1(A, Y_0, Y_1)X_2(t; A), \quad t \geq t_1 > t_0, \quad (2)$$

where the random power series $X_1(t; A)$ and $X_2(t; A)$ are determined taking into account the values of the roots (r_1 and r_2) of the associated indicial equation

$$r(r-1) + p_0 r + q_0 = 0, \quad \text{where } p_0 = p(t_0; A), \ q_0 = q(t_0; A).$$

Notice that, in the deterministic setting the Fröbenious method considers the following three cases: $r_1 - r_2 \neq 0$, $r_1 - r_2 = N \in \mathbb{N}$, $r_1 = r_2$. As in our context, the roots r_1 and r_2 depend on the absolutely continuous RV A, the two latter cases occur with probability 0. Therefore, both scenarios will be neglected hereinafter to conduct our subsequent analysis. As a consequence, the random power series $X_1(t; A)$ and $X_2(t; A)$ are given, respectively, by

$$X_1(t; A) = \sum_{n=0}^{\infty} C_n(A)|t - t_0|^{n+r_1(A)}, \quad \text{and} \quad \sum_{n=0}^{\infty} D_n(A)|t - t_0|^{n+r_2(A)}, \quad (3)$$

where the coefficients $C_n(A)$ are determined by adequate recurrences From a computational standpoint, the infinite series in (3) must be truncated to keep the computational burden affordable. So, we consider

$$X^N(t) = K_0(A, Y_0, Y_1)X_1^N(t; A) + K_1(A, Y_0, Y_1)X_2^N(t; A),$$

$$X_1^N(t; A) = \sum_{n=0}^{N} C_n(A)|t - t_0|^{n+r_1(A)}, \sum_{n=0}^{N} D_n(A)|t - t_0|^{n+r_2(A)}, \quad (4)$$

being N a positive integer. By applying Th.1, it can be proved that the PDF of $X^N(t)$ is given by

$$f_1^N(x,t) = \int_{\mathbb{R}^2} f_{A,Y_0,Y_1}\left(a, \frac{x - y_1 S_2^N(t;a)}{S_1^N(t;a)}, y_1\right)\left|\frac{1}{S_1^N(t;a)}\right|, \tag{5}$$

where

$$S_1^N(t;A) = G_{0,1}(A)X_1^N(t;A) + G_{1,1}(A)X_2^N(t;A),$$

$$G_{0,1} = \frac{X_2'(t_1;A)}{E(A)}, \quad G_{1,1} = \frac{-X_1'(t_1;A)}{E(A)}$$

$$S_2^N(t;A) = G_{0,2}(A)X_1^N(t;A) + G_{1,2}(A)X_2^N(t;A), \tag{6}$$

$$G_{0,2} = \frac{-X_2(t_1;A)}{E(A)}, \quad G_{1,2} = \frac{X_1(t_1;A)}{E(A)}.$$

Finally, we point out that assuming some mild conditions on the random vector (Y_0, Y_1, A), and on its PDF, it can be shown that the approximation $f_1^N(x,t)$ given in (5) will converge to the exact PDF $f_1(x,t)$ of the exact solution SP (2)–(3), i.e.,

$$\lim_{N \to +\infty} f_1^N(x,t) = f_1(x,t), \quad \text{for each } (x,t) \in \mathbb{R} \times [t_1, +\infty[\text{ fixed.} \tag{7}$$

3 An illustrative example

In this section we consider the particular random IVP

$$\left.\begin{array}{l} At^2 X''(t) + t(t+1)X'(t) - X(t) = 0, \\[2mm] X(1) = Y_0, \quad X'(1) = Y_1, \quad t > 1 > 0, \end{array}\right\} \tag{8}$$

where A, Y_0 and Y_1 are assumed to be independent RVs with the following distributions: A is a uniform RV on the interval $[1,2]$, i.e., $A \sim \mathrm{U}([1,2])$; Y_1 is a Beta RV with parameters 2 and 3, i.e., $B \sim \mathrm{Be}(2;3)$; Y_0 is a Gaussian RV with 0 mean and variance 0.1, i.e., $Y_0 \sim \mathrm{N}(0;0.1)$. In Figures 1 and 2, we show $f_1^N(x,t)$ at the time instants $t = 1.1$ and $t = 1.5$, respectively, for different values of N (truncation order). On the left, for $N \in \{1,\ldots,5\}$ and $N \in \{1,\ldots,6\}$, and on the right for $N \in \{4,5\}$ and $N \in \{5,6\}$. We can observe that when the truncation increases the 1-PDF of the approximate solution tends to the exact 1-PDF. In addition, when the time instant t is close the initial time $t = 1$ the convergence is faster. For sake of clarity in
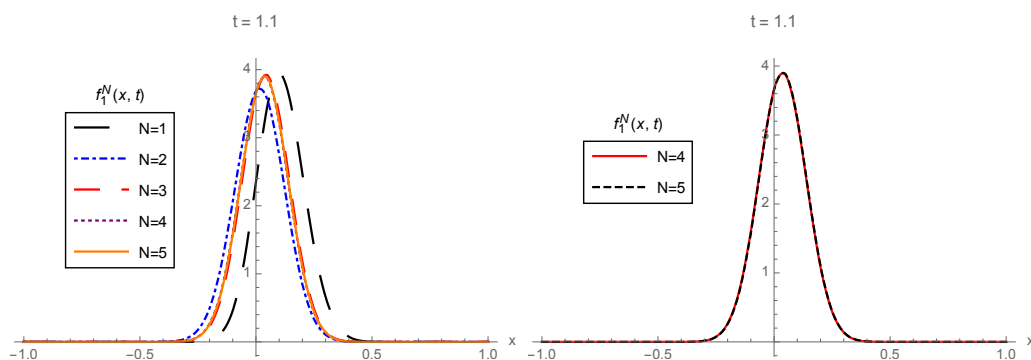
Figure 1: Plot of $f_1^N(x, 1.1)$ given by (5)–(6) for different values of N: $N \in \{1, \ldots, 5\}$ (left), $N \in \{4, \ldots, 5\}$ (right).

Table 1 the error between the approximate and the exact distributions is shown

$$e_N = \int_{-\infty}^{+\infty} |f_1^{N+1}(x, t) - f_1^N(x, t)| \, \mathrm{d}x, \quad N \geq 1, \ t \geq 1 \ \text{fixed.} \qquad (9)$$

e_N	$N = 1$	$N = 2$	$N = 3$	$N = 4$	$N = 5$
$t = 1.1$	0.519832	0.166420	0.039233	0.007228	0.001129
$t = 1.5$	0.437495	0.219520	0.071847	0.018712	0.004031

Table 1: Error measure e_N defined by (9) for different time instants, $t \in \{1.1, 1.5\}$, and series truncation orders, $N \in \{1, 2, 3, 4, 5\}$.

Acknowledgements
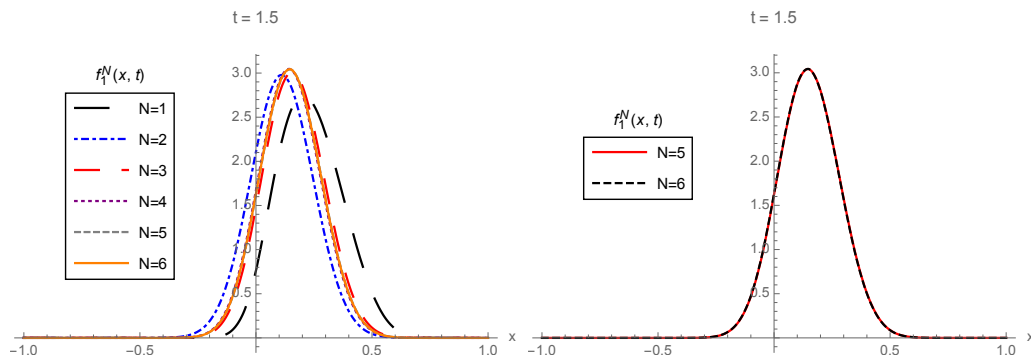
Figure 2: Plot of $f_1^N(x, 1.5)$ given by (5)–(6) for different values of N: $N \in \{1, \ldots, 6\}$ (left), $N \in \{5, \ldots, 6\}$ (right).

References

[1] J.-C. Cortés, A. Navarro-Quiles, J.-V. Romero, M.-D. Roselló. Solving second-order linear differential equations with random analytic coefficients about ordinary points: A full probabilistic solution by the first probability density function. *Applied Mathematics and Computation*, 331:33–45, 2018.

[2] T.T. Soong, Random Differential Equations in Science and Engineering. New York, Academic Press, 1973.

[3] S.L. Ross. Differential Equations. John Wiley & Sons. New York, 1984.

On some properties of the PageRank versatility

F. Pedroche[♭*], R. Criado[†], E. García[†], and M. Romance[†]

(♭) Institut de Matemàtica Multidisciplinària, Universitat Politècnica de València,

Camí de vera s/n, 46022 València. Espanya,

(†) Department of Applied Mathematics, Rey Juan Carlos University,

C/Tulipán s/n, 28933 Móstoles (Madrid), Spain.

November 30, 2018

1   Introduction

The study of *multilayered* networks is a major area of research within the field of Complex Networks. In the literature, one can find some studies that failed to satisfactorily describe the behaviour of the systems by using classical techniques of monoplex networks; See, e.g, [7] for fails about detection of communities, [11] for misunderstandings when combining different interactions on social networks, [9] for ranking differences when ignoring the multilayered nature of a metro system, and [10] for an analysis of the transition from a collection of independent networks to a whole multiplex. As a consequence, it was necessary to implement new concepts and techniques to cope with the heterogeneity of links shown by these networks (see, e.g, [1], [12], [2] for more references). In particular, the concept of multiplex networks (formed by some layers with the same nodes and such that the only allowed interlayer links are those corresponding to nodes connected with themselves, see Fig. 1) has been used extensively.

---

*e-mail:pedroche@mat.upv.es

2 Classic PageRank

For the sake of simplicity, let us consider an undirected connected graph with n nodes, and with adjacency matrix $A \in \mathbb{R}^{n \times n}$, where

$$a_{ij} = \left\{ \begin{array}{ll} 1 & \text{if node } i \text{ is connected with node } j \\ 0 & \text{otherwise} \end{array} \right.$$

and let P_A be the row stochastic matrix defined as $P_A = (p_{ij}) \in \mathbb{R}^{n \times n}$ such that

$$p_{ij} = \frac{a_{ij}}{\sum_{k=1}^{n} a_{ik}}$$

In this case, the Google matrix (see, e.g., [8]) is defined as

$$G = \alpha P_A + (1 - \alpha)\mathbf{e}\mathbf{v}^T \in \mathbb{R}^{n \times n} \qquad (1)$$

where α is a probability, $\mathbf{e}$ is the column vector of all ones and $\mathbf{v}$ is a probability distribution vector, that is, is nonnegative and $\mathbf{v}^T\mathbf{e} = 1$. The *classic* PageRank vector $\widehat{\pi}$ is the unique positive left eigenvector of G associated with the eigenvalue 1 and normalized such that $\widehat{\pi}^T\mathbf{e} = 1$.

3 PageRank versatility

In this paper, we focus on a centrality measure called *PageRank versatility* that was introduced in [4] where the authors make use of the tensor notation for multilayer networks developed in [3]. Formally, a multilayer network is characterized by a *multilayer adjacency tensor* $M_{\beta\tilde{\delta}}^{\alpha\tilde{\gamma}}$, where indices with tilde refer to layers. Let us denote by n the number of nodes of each layer, and by k the number of layers. For example, when handling with a multiplex network like the one shown in Fig. 1, we have $n = 3$ and $k = 3$ and the tensor $M_{\beta\tilde{\delta}}^{\alpha\tilde{\gamma}}$ can be represented in matrix notation (without explicitly show the indices of the nodes) by a matrix $\mathbb{M}$ of size $nk \times nk$ in the following form

$$M_{j\beta}^{i\alpha} \equiv \mathbb{M} = \sum_{\alpha,\beta=1}^{k} \mathbb{E}(\alpha, \beta) \otimes \mathbb{C}(\alpha, \beta)$$

where $\otimes$ denotes the Kronecker product (see, e.g., [6]) and the matrix $\mathbb{E}(\alpha, \beta) \in \mathbb{R}^{k \times k}$ is given by

$$\mathbb{E}(\alpha, \beta) = \mathbf{e}_\alpha^k \otimes (\mathbf{e}_\beta^k)^T$$
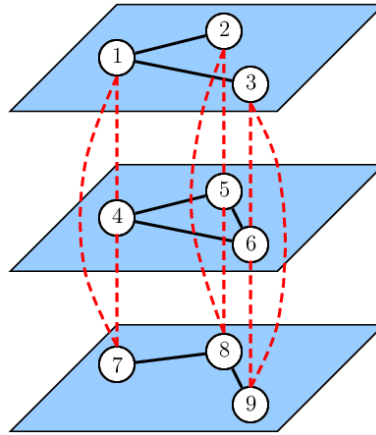
Figure 1: A multiplex with three layers and three nodes on each layer. Red dashed lines represent *inter-layer* links.

where $\mathbf{e}_\alpha^k$ is the α-th (column) vector of the canonical basis of $\mathbb{R}^{k\times 1}$ and the superscript T means transposition. Note, for example, that

$$\mathbb{E}(1,3) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \otimes (0\,0\,1) = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The matrices $\mathbb{C}(\alpha,\beta) \in \mathbb{R}^{n\times n}$ represent both the adjacency matrices of the layers and the matrices accounting for the links between layers, that is, in the example of Fig 1:

$$\mathbb{C}(\alpha,\beta) = \begin{cases} I_n & if \quad \alpha \neq \beta \\ A_\alpha & if \quad \alpha = \beta \end{cases}$$

where I_n is the identity matrix of size n and A_α is the adjacency matrix of layer α.

Once $\mathbb{M}$ is defined, the *PageRank versatility* can be defined following a similar procedure as in the *classic PageRank* [8]. For the sake of clearness we use matrix notation. Note that in a multiplex framework with undirected links, the matrix $\mathbb{M}$ is a symmetric matrix and has no zero rows. Let us denote m_{ij}, $i,j \in \{1,2,\dots,nk\}$ each element of $\mathbb{M}$. Hence we can define the row stochastic matrix $\mathbb{T}$ with elements t_{ij} as follows

$$t_{ij} = \frac{m_{ij}}{\sum_{j=1}^{nk} m_{ij}}$$

and define a matrix $\mathbb{G} \in \mathbb{R}^{nk \times nk}$ (analogous to the Google matrix 1) as follows

$$\mathbb{G} = \alpha \mathbb{T} + (1 - \alpha) \mathbf{e}^{nk} \mathbf{v}^T$$

where $\mathbf{e}^{nk} \in \mathbb{R}^{nk \times 1}$ is the vector of all ones, and $\mathbf{v}^T = \frac{1}{k}[\mathbf{v_1}^T \, \mathbf{v_2}^T, \ldots \mathbf{v_k}^T]$ is a *personalization vector* formed by staking the personalization vector of each layer $\mathbf{v}_i \in \mathbb{R}^{n \times 1}$. We remark that by taking $\mathbf{v}_i = \frac{1}{n}[1, 1, \ldots, 1]$ for all i, the term $\mathbf{e}^{nk}\mathbf{v}^T$ is $\frac{1}{nk}$ multiplied by a square matrix of size $nk \times nk$ with all its elements equal to one.

By construction, $\mathbb{G}$ is row stochastic and positive and therefore by using the Perron Theorem for positive matrices it is known that $\mathbb{G}$ has a unique positive left eigenvector $\Pi \in \mathbb{R}^{nk \times 1}$ with norm equal to 1 associated to the eigenvalue 1 of $\mathbb{G}$. This vector can be *folded* to obtain a vector of size $\mathbb{R}^{n \times 1}$ by doing the following. First, we define

$$p_i = \mathbf{e}^k \otimes \mathbf{e}_i^n, \quad i = 1, 2, \ldots, n$$

where $\mathbf{e}^k$ is the vector of all ones in $\mathbb{R}^{k \times 1}$ and $\mathbf{e}_i^n$ is the i-th column of the identity matrix of size n. Second, we define

$$\pi_i = p_i^T \Pi \in \mathbb{R}$$

and finally, the *PageRank versatility* is the vector of $\mathbb{R}^{n \times 1}$ given by

$$\pi = [\pi_1, \pi_2, \ldots, \pi_n]^T.$$

From the expressions above it is straightforward to obtain some properties of the versatility PageRank that are derived from those of the classic PageRank. To illustrate this, in the next section, we focus on a case where we only have two layers,

4   Example with two layers

In the case of two layers, with n nodes in each one, the *multilayer adjacency tensor* becomes

$$\mathbb{M} = \begin{pmatrix} A_1 & I \\ I & A_2 \end{pmatrix},$$

where I is the identity matrix of order n, and A_i is the adjacency matrix of layer $i = 1, 2$. From this matrix we can construct an stochastic matrix P_M in

the usual way: by dividing each entry of $\mathbb{M}$ by the sum of its corresponding row entries. Once this operation is performed we obtain a Google matrix of the form

$$\mathbb{G} = \alpha P_M + (1 - \alpha)\mathbf{e}\mathbf{v}^T \qquad (2)$$

with $\mathbf{e}$ the column vector of all ones with $2n$ components, and $\mathbf{v}^T = [\mathbf{v}_1^T \ \mathbf{v}_2^T]$ where each $\mathbf{v_i}^T$ is a probability distribution vector. The analogy between equations (1) and (2) allows us to apply all the known results about the *classic* Google matrix to obtain properties of matrix $\mathbb{G}$. That is, once $\mathbb{G}$ is constructed, all the properties of the PageRank vector Π associated to $\mathbb{G}$ are the corresponding of those of the PageRank vector of G. For example, the results about the localization of the PageRank Π of $\mathbb{G}$ follow the corresponding formulas of those about the localization of the PageRank of G (see, [5]).

The open problem that remains is to relate some properties of the versatility PageRank with the classic PageRanks associated to each of the matrices A_1 and A_2. We propose this as a future work.

References

[1] F. Battiston, V. Nicosia, and V. Latora, *The new challenges of multiplex networks: Measures and models* Eur. Phys. J. Spec. Top. (2017) 226: 401. https://doi.org/10.1140/epjst/e2016-60274-8

[2] S. Boccaletti, G. Bianconi, R. Criado, C.I. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang and M. Zanin, *The structure and dynamics of multilayer networks*, Physics Reports **544**(1) (2014), 1-122.

[3] M. De Domenico, A. Solé-Ribalta, E. Cozzo, M. Kivela, Y. Moreno, M.A. Porter, S. Gómez, and A. Arenas, *Mathematical formulation of multi-layer networks*, Phys. Rev. X 3, 041022 (2013).

[4] M. De Domenico, A. Solé-Ribalta, E. Omodei, S. Gómez, and A. Arenas *Ranking in interconnected multilayer networks reveals versatile nodes*, Nature Communications 6, Article number: 6868 (2015).

[5] E. García, F. Pedroche and M. Romance, *On the localization of the Personalized PageRank of Complex Networks*, Linear Algebra and its Applications **439**, 640 (2013).

[6] R. A. Horn, and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge Univ. Press. 1991.

[7] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela, *Community Structure in Time-Dependent, Multiscale, and Multiplex Networks*, Science 328, 876 (2010). https://doi.org/10.1126/science.1184819

[8] L. Page, S. Brin, R. Motwani and T. Winograd, *The PageRank citation ranking: Bridging order to the Web*, Tech.Rep. **66**, Stanford University. 1998.

[9] F. Pedroche, M. Romance, and R. Criado, *A biplex approach to PageRank centrality: From classic to multiplex networks*, Chaos: An Interdisciplinary Journal of Nonlinear Science . 26, 065301 (2016). https://doi.org/10.1063/1.4952955

[10] F. Radicchi and A. Arenas, *Abrupt transition in the structural formation of interconnected networks*, Nature Physics **9**, 717-720 (2013). https://doi.org/10.1038/nphys2761

[11] M. Szell, R. Lambiotte and S. Thurner, *Multirelational organization of large-scale social networks in an online world*, PNAS 107 (31) 13636-13641 (2010). https://doi.org/10.1073/pnas.1004008107

[12] D. Wang, and X. Zou, *A new centrality measure of nodes in multilayer networks under the framework of tensor computation*, Applied Mathematical Modelling **54** (2018) 46-63.

Network clustering strategies for setting degree predictors based on deep learning architectures

Francisco Javier Pérez-Benito[♭] *, Esperanza Navarro-Pardo[†],
Juan M. García-Gómez[‡], and J. Alberto Conejero[♭]

(♭) Instituto Universitario de Matemática Pura y Aplicada (IUMPA),

Universitat Politècnica de València

(†) Departamento de Psicología Evolutiva y de la Educación,

Universitat de València

(‡) Biomedical Data Science Lab.

Instituto de Aplicaciones de las Tecnologías de la Información

y de las Comunicaciones Avanzadas (ITACA),

Universitat Politècnica de València

November 30th, 2018

1 Introduction

Happiness is a universal fundamental human goal. Since the emergence of Positive Psychology [8], a major focus in psychological research has been to study the role of specific factors in the prediction of happiness. Conventional methodologies are based on linear relationships, such as the commonly used Multivariate Linear Regression (MLR) [2], which may suffer from the lack of representative capacity to the varied psychological features. Using Deep Neural Networks (DNN), a *Happiness Degree Predictor* (HDP) was defined based on the answers to five standardized psychometric questionaries [7].

---

*e-mail: frapebe@doctor.upv.es

The lower-level dimensions of psychological factors are separately ensembled for being subsequently merged by higher-level dimensions until happiness is reached. The DNN that gives us this HDP was trained and tested using a cross-sectional survey targeting non-institutionalized adult population residing in Spain completed by 823 cases and recruited by different interviewers. A total of 111 survey elements were grouped by socio-demographic data, and by five psychometric scales measuring five psychological dimensions. Coping strategies factor was measured by Brief COPE Inventory [3], personality by EQPR-A [5], emotional distress by GHQ-28 [10], and social support by MOS-SSS [9]. As an outcome of *Happiness*, it was considered the result in the SDSH scale [6]. Each psychometric scale is composed of items (questions) that can be regrouped into sub-scales matching psychological subdimensions. The 28 Brief COPE Inventory items were regrouped into 14 sub-scales, the 24 EPQR-A items into 4, the 28 GHQ-28 items into 4, and the 19 MOS-SSS items into 4. Covering a great variety of psychological indicators, such as substance abuse, self-distraction, sincerity, somatic symptoms or positive social interaction.

2 Data Structure-driven Deep Neural Network as Happiness Degree predictor

We propose a hierarchical ensembling data-driven method for modeling the task in hand. The preconceived data structure has led the deep neural network layers' ensembling. The items of the psychometric scales employed for measuring the psychological factors used as predictors have been empirically proved to cluster into sub-dimensions and dimensions mentioned above.

With the 105 of the 111 elements for each participant we construct a column vector with the inputs for the deep neural network. The first element represents a numeric identifier for the interviewer. From 2nd to the 5th elements we have the socio-demographic data about the interviewee. The rest of inputs (from 6th to the 105th) are the responses to the items that make up the standardized psychometric scales. The sum of the other six elements, ranging from 0 depression to 18 happiness, formed the output becoming the gold-standard for supervised-training for the outcome of the D-SDNN.

We have mimicked this empirically-based conceptual structure in the

design of the architecture. This shapes our first contribution, the Data-Structure driven Deep Neural Network (D-SDNN) architecture, that is shown in Figure 1.
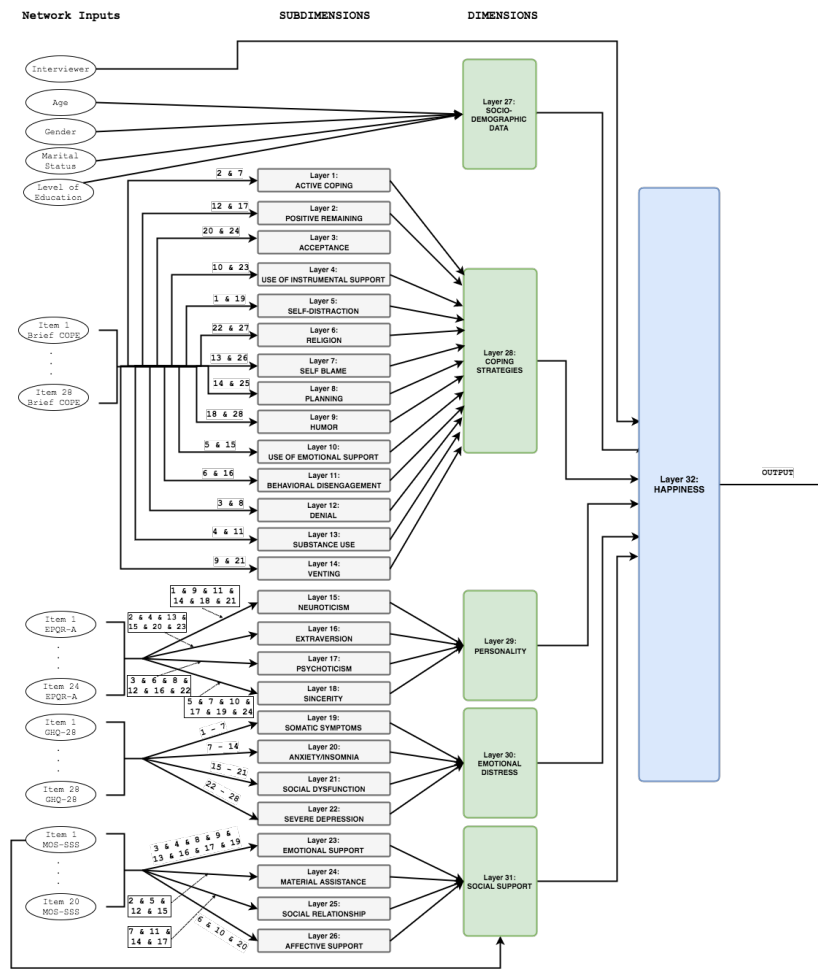
Figure 1: Data-structure architecture for our proposed neural network model. The associated numbers to each arrow are related to the number of the items enclosed into the sub-dimension. The structure is shown up to the sub-dimension level.

We have considered two metrics (1) and (2) that allow us to quantify the psychological dimensions influence in the outcome. Let n_{inp} be the number of inputs of one neuron of the layer L. In order to measure the global importance

of the inputs, we propose the following metrics regarding weights for the j_{th} neuron in the layer L

$$L_i^{(j)} = \sum_{i=1}^{n_{inp}} \frac{|w_{ij}|}{n_{inp}}, \tag{1}$$

and the positivity or negativity of the relationship is determined by

$$\mathrm{sgn}\left(L_i^{(j)}\right) = \mathrm{sgn}\left(\sum_{i=1}^{n_{inp}} w_{ij}\right). \tag{2}$$

We have used 578 instances (column vectors) of the total sample, approximately the 70%, for training the 4 tentatives D-SDNNs. Regarding the other 30%, a 15% has been used for validating and the last 15% for testing.

With performance assessment purposes, we have used the Mean Squared Error (MSE) to compare the results (just on the testing set) of our D-SDNN with the state-of-the-art model in the field, MLR, where our approach provided a better outcome (MSE: $1.46 \cdot 10^{-2}$) than MLR (MSE: $2.30 \cdot 10^{-2}$), hence improving by 37% the predictive accuracy. We have observed better performance of deep machine-learning architectures concerning traditional methodologies. These results demonstrate the success of predicting happiness degree through psychological variables assessed by standardized questionnaires. The influence metrics of the psychological dimensions are shown in Table 1.

Conceptual dimensions	$L_{32}^{(l)}$	$sgn\left(L_{32}^{(l)}\right)$	Interpretation
Interviewer	0.0311	-	Small negative influence
Socio-demographic data	0.1403	+	Small positive influence
Coping Strategies	0.4476	-	Most negatively influential
Personality	0.4186	+	Positively influential
Emotional Distress	0.3897	-	Negatively influential
Social Support	0.5025	+	Most positively influential

Table 1: Influence metric values D-SDNN.

It can be considered congruent with common sense expectations the significantly high and negative influence of Emotional Distress in the degree of happiness. The significantly high and positive influence of the Perceived Social Support in the degree of happiness is consistent with the existing literature. According to these findings, the Perceived Social Support may be seen as a buffer for the deleterious effect of Emotional Distress.

3 A new fully automated data structure construction approach

3.1 Motivation

The hidden structure of the D-SDNN provides a limited amount of information about the conceptual interpretation of the psychological factors associated with happiness. This is due to the lack of backpropagation possibility that psychological subdimension layers has some influence in the metrics computation. For this purpose, we have explored different options in order to design other DNN architectures which not only promoted a better understanding concerning the study of psychological sub-dimensions influencing happiness but also automatically built and set the DNN architecture.

On the one hand, the trending Network Science paradigm is becoming a widely used mechanism when a vast amount of data must be treated, due to its capability of data association [11] and structure creation through a natural binary relationship defined by the analyst. On the other hand, machine learning approaches suffer low popularity in applied sciences because of its lack of interpretability [4] provoked by the complexity of the models.

In this sense, we have tried to combine both technologies with the purpose of their complementation and the expectation of that the the automatic model construction would also contribute to the interpretation of the results.

3.2 Proposed methodology

With the same dataset, we have tried to measure the similarity between each pair of items, no matter to which questionnaire/dimension they belong. The more people answers in the same sense a pair of items, the more similar we consider that they are. Then a weight is assigned to the edge connecting each pair of items according to this similarity, shaping a similarity graph in which two nodes (items) are connected if and only if at least two people have answered them in the same sense. We have assigned a weight to the connection that is proportional to the number of similar pairs of answers.

We have also analyzed the modularity of this network using the work of Blondel et al. [1]. The study of different communities in which the network can be split permits us to fully-automatic design new DNN granulated-architectures that are competitive respect to the one given the HDP. The

point is that this new approach permits us also to have a better understanding about how different psychological factors are correlated and how do they influence happiness.

The challenges we are affording cover two main aspects:

1. How to fix the number of granularity levels not only to obtain a good estimation ability but also optimize the computational complexity.

2. In case the items belonging to one subdimension that fall on different dimensions, how to automatically build the minimum auxiliary layers to ensure that the conceptual data structure provided by the graph is maintained.

Nevertheless, we also point out that the proposed methodology could benefit of big data approaches, since they are radically different to the approach used in the psychology in which each validated questionnaire can only be considered as a whole or, at most, split into few dimensions.

References

[1] Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *arXiv*, Mar 2008.

[2] Daniel Campos, Ausiàs Cebolla, Soledad Quero, Juana Bretón-López, Cristina Botella, Joaquim Soler, Javier García-Campayo, Marcelo Demarzo, and Rosa María Baños. Meditation and happiness: Mindfulness and self-compassion may mediate the meditation–happiness relationship. *Personality and Individual Differences*, 93:80–85, 2016.

[3] Charles S Carver. You want to measure coping but your protocol'too long: Consider the brief cope. *International journal of behavioral medicine*, 4(1):92, 1997.

[4] Finale Doshi-Velez and Been Kim. Towards a rigorous science of interpretable machine learning. *arXiv preprint:1702.08608*, 2017.

[5] Leslie J Francis, Laurence B Brown, and Ronald Philipchalk. The development of an abbreviated form of the revised eysenck personality

questionnaire (epqr-a): Its use among students in england, canada, the usa and australia. *Personality and individual differences*, 13(4):443–449, 1992.

[6] Stephen Joseph, P Alex Linley, Jake Harwood, Christopher Alan Lewis, and Patrick McCollam. Rapid assessment of well-being: The short depression-happiness scale (sdhs). *Psychology and Psychotherapy: Theory, research and practice*, 77(4):463–478, 2004.

[7] Francisco Javier Pérez-Benito, Patricia Villacampa-Fernández, J. Alberto Conejero, Juan M. García-Gómez, and Esperanza Navarro-Pardo. A happiness degree predictor using the conceptual data structure for deep learning architectures. *Comput. Methods Programs Biomed.*, Nov 2017.

[8] Martin EP Seligman and Mihaly Csikszentmihalyi. *Positive psychology: An introduction.*, volume 55. American Psychological Association, 2000.

[9] Cathy Donald Sherbourne and Anita L Stewart. The mos social support survey. *Social science & Medicine*, 32(6):705–714, 1991.

[10] Michele Sterling. General health questionnaire–28 (ghq-28). *Journal of Physiotherapy*, 57(4):259, 2011.

[11] Gregory Tauer, Ketan Date, Rakesh Nagi, and Moises Sudit. An incremental graph-partitioning algorithm for entity resolution. *Information Fusion*, 46:171–183, 2019.

Qualitative preserving stable difference methods for solving nonlocal biological dynamic problems

M. A. Piqueras[†] [*], R. Company[†] , and L. Jódar[†]

(†) Instituto de Matemática Multidisciplinar, Universitat Politècnica de València,

Camino de Vera s/n, 46022 Valencia, Spain.

1 Introduction

In this paper we consider the nonlocal interaction biological dynamic model described by the partial integro-differential reaction-diffusion problem (PIDE), see [3]:

$$\frac{\partial U}{\partial t} = D\Delta U + \beta U(\mathbf{x}, t)\Big(1 - aU(\mathbf{x}, t) - b\int_{\Omega} \psi(\mathbf{x} - \mathbf{y})U(\mathbf{y}, t)d\mathbf{y}\Big), \ \ \mathbf{x} \in \Omega, \ \ t \in [0, T],$$

(1)

where $\Omega \subseteq \mathbb{R}^2$ is a bounded or unbounded domain, $\psi(\mathbf{x})$ is a nonnegative kernel function satisfying

$$\int_{\mathbb{R}^2} \psi(\mathbf{x})d\mathbf{x} = 1,$$

(2)

β and b are some positive constants, a is a nonnegative constant and D is a positive dispersal rate, together with the initial and boundary conditions

$$U(\mathbf{x}, 0) = f(\mathbf{x}), \ \mathbf{x} \in \Omega, \ \ U(\mathbf{x}, t) = 0, \ \mathbf{x} \in \partial\Omega,$$

(3)

where $f(\mathbf{x})$ represents an arbitrary continuous function.

---

[*]e-mail: mipigar@cam.upv.es

From the biological point of view, the first term of the right-hand side models the diffusion, the second includes the pure logistic quadratic term and the consumption of resources in some area around the average location. Note that if the kernel $\psi(\mathbf{x})$ is the Dirac delta function centered at the origin, equation (1) recovers the Fisher-KPP equation, see [2]. In ecological context, there is no real justification for assuming that the interactions are local.

In this work, we develop an explicit finite difference scheme for the numerical computation of problem (1)-(3), together with an exhaustive numerical analysis. The integral term of the PIDE is treated using Gauss quadrature rules having the versatility advantage of including both the bounded and unbounded domain cases, just adapting the quadrature rule. Positivity of the numerical solutions is crucial dealing with a population problem and needs to be guaranteed. It is also important to check that numerical solutions are bounded by the habitat carrying capacity, in agreement with the behaviour of the theoretical solution [3].

2 Discretization and numerical scheme construction

The continuous problem is discretized here, with the goal to reach an explicit finite difference scheme. Hereafter, we will work in a suitable bounded numerical domain. Let us consider the domain $[-A, A]^2 \times [0, T]$, with $A > 0$ large enough so that outside of this area the population is negligible and $T > 0$ denoting the time horizon. Let M and N be positive integers, so that the domain $[-A, A]^2 \times [0, T]$ is partitioned in $(2M+1)^2 \times (N+1)$ mesh points denoted by $(x_{1,i}, x_{2,j}, t^n)$, where $x_{1,i} = ih$, $-M \le i \le M$, $x_{2,j} = jh$, $-M \le j \le M$, and $t^n = nk$, $0 \le n \le N$. The step sizes discretizations h and k verify $hM = A$ and $kN = T$, respectively. The numerical approximation of the unknown variable at the mesh point $(x_{1,i}, x_{2,j}, t^n)$ is denoted by $u_{i,j}^n \approx U(x_{1,i}, x_{2,j}, t^n)$, while for the integral term in (1), we designate

$$g_{i,j}^n \approx G(x_{1,i}, x_{2,j}, t^n) = \int_\Omega \psi(\mathbf{x}_{i,j} - \mathbf{y})U(\mathbf{y}, t^n)d\mathbf{y}, \quad -M \le i, j \le M, \;\; n \ge 0.$$
(4)

where $\mathbf{x}_{i,j} = (x_{1,i}, x_{2,j})$.

The approximation $g_{i,j}^n$ of the integral term $G(x_{1,i}, x_{2,j}, t^n)$ is performed by means of the accurate and computationally cheap Gauss quadrature rule.

Gauss-Hermite or Gauss-Legendre quadratures are used depending on whether the support of the kernel function $\psi(x)$ is unbounded or compact, respectively. As the nodes of the quadrature rule are not necessarily mesh points of the grid, a bilinear interpolation is used for the computation of the terms $g_{i,j}^n$.

According to the expression for the Gauss-Hermite quadrature, we have

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \psi(\mathbf{x}_{i,j} - \mathbf{y})U(\mathbf{y}, t^n)d\mathbf{y}$$

$$\approx \sum_{l=1}^{L} \sum_{m=1}^{L} w_l w_m e^{x_l^2 + x_m^2} \psi(x_{1,i} - \hat{y}_{1,l}, x_{2,j} - \hat{y}_{2,m}) U(\hat{y}_{1,l}, \hat{y}_{2,m}, t^n), \qquad (5)$$

where w_l, w_m, are the weights and $\hat{y}_{1,l}$, $\hat{y}_{2,m}$, $1 \leq l, m \leq L$, are the nodes of the Gauss-Hermite quadrature, respectively.

Given a node $(\hat{y}_{1,l}, \hat{y}_{2,m})$, $1 \leq l, m \leq L$, let us consider the indexes i_l and j_m such that the grid point (y_{1,i_l}, y_{2,j_m}) verifies

$$y_{1,i_l} \leq \hat{y}_{1,l} \leq y_{1,i_l+1}, \quad y_{2,j_m} \leq \hat{y}_{2,m} \leq y_{2,j_m+1}. \qquad (6)$$

Thus, the approximation $g_{i,j}^n$ of the integral term $G(x_{1,i}, x_{2,j}, t^n)$ takes the form

$$g_{i,j}^n = \sum_{l=1}^{L} \sum_{m=1}^{L} w_l w_m e^{x_l^2 + x_m^2} \psi(x_{1,i} - \hat{y}_{1,l}, x_{2,j} - \hat{y}_{2,m}) \bar{u}(\hat{y}_{1,l}, \hat{y}_{2,m}, t^n). \qquad (7)$$

Regarding the differential part of PIDE (1), considering forward approximation for time derivatives and central approximation for spatial derivatives, the following explicit numerical scheme for (1),(3) has been constructed:

$$u_{i,j}^{n+1} = \frac{Dk}{h^2}(u_{i-1,j}^n + u_{i+1,j}^n + u_{i,j-1}^n + u_{i,j+1}^n) + \left(1 - \frac{4Dk}{h^2}\right)u_{i,j}^n$$

$$+ k\beta u_{i,j}^n(1 - au_{i,j}^n - bg_{i,j}^n), \quad -M + 1 \leq i, j \leq M - 1, \ 0 \leq n \leq N - 1, \quad (8)$$

with initial and transferred boundary conditions of our numerical domain

$$u_{i,j}^0 = f(\mathbf{x}_{i,j}), \quad u_{-M,j}^n = u_{M,j}^n = u_{i,-M}^n = u_{i,M}^n = 0, \ 1 \leq i, j \leq M. \qquad (9)$$

3   Positivity, stability and consistency

Considering the previous result regarding the theoretical solution,

$$0 \leq U(\mathbf{x}, t) \leq 1/a, \quad \mathbf{x} \in \bar{\Omega}, \ t \geq 0. \tag{10}$$

we show that under appropriate step size conditions the numerical solution $\{u_{i,j}^n\}$ is nonnegative and is bounded by the carrying capacity $1/a$ in agreement with (10). Thus the stability of the numerical solution is granted because it is bounded. Precisely, assuming that $0 \leq u_{i,j}^0 \leq 1/a$, and taking a temporal step size k such that

$$k < \frac{h^2}{4D + \beta\alpha h^2}, \quad \alpha = \max\left\{1, \frac{2b}{a}\right\}, \tag{11}$$

it is guaranteed that $0 \leq u_{i,j}^n \leq 1/a$, $1 \leq n \leq N$.

Note that stability and positivity step size condition is linked to the problem dimension. In particular, for the one dimensional case, the condition becomes

$$k < \frac{h^2}{2D + \beta\alpha h^2}, \quad \alpha = \max\left\{1, \frac{2b}{a}\right\}. \tag{12}$$

Now we study the consistency of the numerical solution, given by the scheme (8), with the problem (1)-(3). Let us consider the equation (1), in a compact form as $\mathcal{L}(U) = 0$, and the finite difference scheme (8), written as $L(u) = 0$.

Scheme $L(u)$ is said to be consistent with problem $\mathcal{L}(U)$ if local truncation error $T_{i,j}^n(U)$,

$$T_{i,j}^n(U) = L(U_{i,j}^n) - \mathcal{L}(U_{i,j}^n), \tag{13}$$

tends to zero as $k \to 0$, $h \to 0$, where $U_{i,j}^n = U(x_{1,i}, x_{2,j}, t^n)$ is the value of the exact solution of problem (1)-(3). It can be verified that the local truncation error $T_{i,j}^n(U)$ satisfies:

$$T_{i,j}^n(U) = \mathcal{O}(k) + \mathcal{O}(h^2) + \epsilon(L), \tag{14}$$

where $\epsilon(L)$ is the associated quadrature error of the two-dimensional Gauss quadrature formula. An estimation of the error bound for Gaussian quadrature rules in two dimensions can be found in [1].

4 Numerical example

Following example illustrates the stability results. Let us consider the non-local logistic diffusion model (1)-(3) in an unbounded one space dimension, with parameters values $(D, \beta, a, b) = (0.25, 5, 1, 1)$ and

$$f(x) = \begin{cases} 1/4, & -4 \leq x \leq 4, \\ 0, & \text{otherwise,} \end{cases} \tag{15}$$

$$\psi(\xi) = \begin{cases} 1/2, & -1 \leq \xi \leq 1, \\ 0, & \text{otherwise.} \end{cases} \tag{16}$$

We take $h = 0.05$, $k = 0.004$ and $L = 10$. According to the expression (12), if $k < 0.004762$, which is fulfilled in this case, the positivity and stability of the solution are guaranteed. Figure 1 shows the behaviour of the numerical solution $U(x, t)$ from $t = 0$ to the time horizon $T = 2$. If we choose a temporal step size $k = 2/398 \simeq 0.005025$, breaking the stability condition (12), it is clear from Figure 2 that the behaviour of the numerical solution $U(x, t)$ becomes unstable and it reaches negatives values.
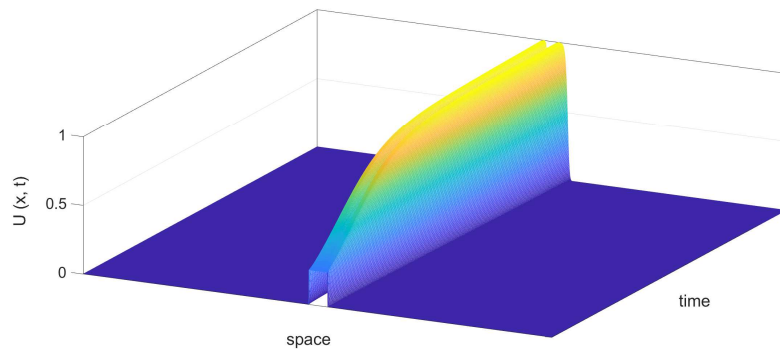
Figure 1: Numerical solution in the case of one space dimension and unbounded domain.
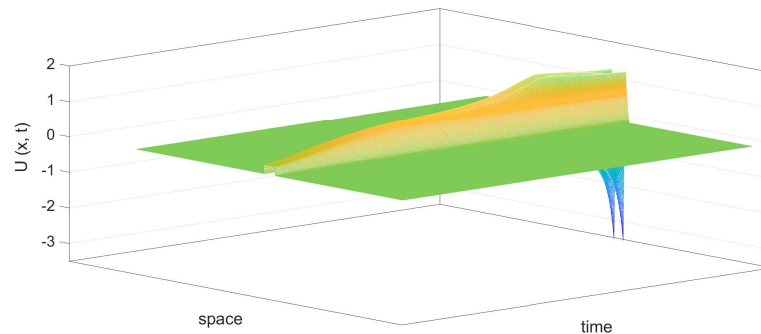
Figure 2: Numerical solution when the positivity and stability condition is broken.

References

[1] A.C. Ahlin, On error bounds for Gaussian cubature, *SIAM Review*, 4:25–39, 1962.

[2] R.A. Fisher, The wave of advance of advantageous genes, *Ann. Eugenics*, 7:335–369, 1937.

[3] V. Volpert, N. Apreutesei, N. Bessonov and V. Vougalter, Spatial structures and generalized travelling waves for an integro-differential equation, *Discrete and Continuous Dynamical Systems - Series B, American Institute of Mathematical Sciences*, 13(3):537–557, 2010.

Probabilistic solution of a random model to study the effectiveness of anti-epileptic drugs

Barrachina-Martínez, I.$^{\flat}$, Navarro-Quiles, A.†, and Ramos, M$^{\flat *}$

($\flat$) Centro de investigación de Ingeniería Económica,

Universitat Politècnica de València, València, Spain,

($\dagger$) DeustoTech, Fundación Deusto

Universidad de Deusto, Bilbao, Spain

November 30, 2018

1 Introduction and motivation

Epilepsy is one of the most ancient diseases that we have knowledge. Descriptions of epileptic seizures can be traced back to 2,000 B.C. Nowadays, even is much of this disease is still a mystery in many senses, it can be controlled giving patients much more quality of life. In Spain it is estimated that around 400,000 people are affected, with nearly 60% of patients having partial onset or focal seizures (POS) [1]. These are caused by a problem in the electrical signalling of the brain. Groups of neurons suddenly begin firing excessively, leading to involuntary responses, including strange sensations, emotions, behaviours or convulsions, muscle spasms, and possibly loss of consciousness. Anti-epileptic Drugs (AEDs) effect is centred on the greatest reduction of the number of epileptic seizures, while minimizing adverse effects and long-term toxicity as far as possible [2]

In this study we propose a method that mathematically simulates these health stage transitions, which represent a relevant epilepsy outcome. In this contribution, we consider that patients are in one of the following four

---

*e-mail: marta.a.m.ramos@gmail.com
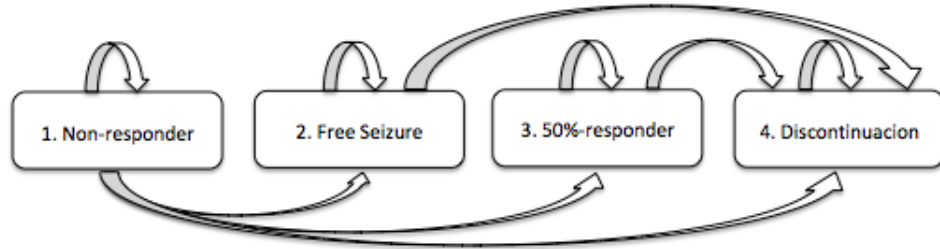
Figure 1: Health stage transitions of an epilepsy patient treated with AEDs.

states: 1. Non-responder, 2. Free seizure, 3. Partial responder (50%), 4. Discontinuation. In Figure 1 we show the health stage transition of an epilepsy patient treated with AEDs.

Let $\{x_n = (x_n^1, x_n^2, x_n^3, x_n^4)^\top, n = 0, 1, \dots\}$ be a Markov chain, where n denotes the cycle or period. Each component x_n^k lies in the interval $[0, 1]$ and denotes the percentage of population in the state k in the cycle n. Moreover, they satisfy $x_n^1 + x_n^2 + x_n^3 + x_n^4 = 1$ for every n. Then, taking into account Figure 1, given p_{ij} the probability of transition from the state i to the state j, the mathematical model is

$$x_{n+1} = p\,x_n, \quad p = \begin{pmatrix} 1 - p_{12} - p_{13} - p_{14} & 0 & 0 & 0 \\ p_{12} & 1 - p_{14} & 0 & 0 \\ p_{13} & 0 & 1 - p_{14} & 0 \\ p_{14} & p_{14} & p_{14} & 1 \end{pmatrix} \tag{1}$$

where p, usually called the transition matrix, is a matrix which entries represent the probabilities to change either from one state to another or to remain in the same state between two consecutive cycles.

These probabilities are normally obtained from experiments, then it contains a certain measurement error. This situation makes more advisable to consider these parameters as random variables (RVs) rather than deterministic constants. Most of the limitations of studies about chronic diseases, with probabilities of crisis with no explanation, is the uncertainty of results depending on each patient. Therefore, the main goal of this contribution is to solve, from a probabilistic point of view, the resulting random model. To distinguish RVs from deterministic variables, hereinafter RVs will be written

using capital letters. So the randomized model is written as

$$X_{n+1} = PX_n, \quad P = \begin{pmatrix} 1 - P_{12} - P_{13} - P_{14} & 0 & 0 & 0 \\ P_{12} & 1 - P_{14} & 0 & 0 \\ P_{13} & 0 & 1 - P_{14} & 0 \\ P_{14} & P_{14} & P_{14} & 1 \end{pmatrix}, \quad (2)$$

where P_{12}, P_{13} and P_{14} are assumed to be absolutely continuous RVs defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with joint probability density function $f_{P_{12}, P_{13}, P_{14}}(p_{12}, p_{13}, p_{14})$.

2 Solving the randomized model

As a main difference with respect to the deterministic framework, solving a randomized problem means not only to obtain its solution but also the probabilistic information associated with it such as the mean a the variance. In order to have a full probabilistic description of the solution process (SP) the main contribution of this work is the computation of its first probability density function (1-PDF), that is the 1-PDF of the effectiveness of AEDs in Epileptic patients. The computation of the 1-PDF is advantageous since it permits to compute all one-dimensional statistical moments of the solution SP

$$\mathbb{E}\left[(X_n)^k\right] = \int_0^1 x^k f_1(x; n)\, \mathrm{d}x, \quad k = 0, 1, 2, \ldots$$

As a consequence, the mean, $\mathbb{E}[X_n]$, and the variance, $\mathbb{V}[X(t)] = \mathbb{E}[(X_n)^2] - \mathbb{E}[X_n]^2$, are easily derived as particular cases.

Taking as initial condition $x_0 = (1, 0, 0, 0)^\top$, the solution SP of the randomized problem (2) is

$$X_n = P^n\, x_0 = \begin{pmatrix} (1 - P_{12} - P_{13} - P_{14})^n \\ \dfrac{P_{12}((1 - P_{14})^n - (1 - P_{12} - P_{13} - P_{14})^n)}{P_{12} + P_{13}} \\ \dfrac{P_{13}((1 - P_{14})^n - (1 - P_{12} - P_{13} - P_{14})^n)}{P_{12} + P_{13}} \\ 1 - (1 - P_{14})^n \end{pmatrix}. \quad (3)$$

Then, in order to compute the 1-PDF we apply the RVT method [3], which has been successful applied in previous randomized models [4]. For instance, applying the RVT technique, the 1-PDF of the solution SP X_n^1 of

non-responder sub-population, defined in (3) is

$$f_1(x;n) = \int_{\mathbb{R}^2} f_{P_2,P_3,P_4}\left(1 - x^{1/n} - \delta - \eta, \delta, \eta\right) \left|\frac{-x^{-1+1/n}}{n}\right| \,\mathrm{d}\delta \,\mathrm{d}\eta. \quad (4)$$

3   Graphical Example

In this example we study the effectiveness of the Brivaracetam AED as treatment of epileptic patients. Based on [5] the deterministic transition probabilities are $p_{12} = 0.02520$, $p_{13} = 0.13765$ and $p_{14} = 0.05720$. Therefore, we consider that P_{12}, P_{13} and P_{14} are independent RVs, each one with a truncated Beta Distribution in the interval T_i and parameters $(a_i; b_i)$, i.e., $P_{1i} \sim \mathrm{Be}_{T_i}(a_i; b_i)$, with

$$T_i = [p_{1i}(1 - 0.2), p_{1i}(1 + 0.2)], \quad \begin{cases} a_i &= p_{1i}\left(\dfrac{p_{1i}(1 - p_{1i})}{\sigma_i^2} - 1\right), \\ b_i &= (1 - p_{1i})\left(\dfrac{p_{1i}(1 - p_{1i})}{\sigma_i^2} - 1\right). \end{cases}$$

In Figure 3 the 1-PDF of the solution SP of non-responder sub-population is plotted for different instants time $n \in \{1, 2, \ldots 8\}$. We can observe that the number of non-responder decrease in time, going to zero when the time tends to infinity. This graphical representation is in agreement with the Figure 3 where the expectation as well as the 95% of confidence interval is represented.
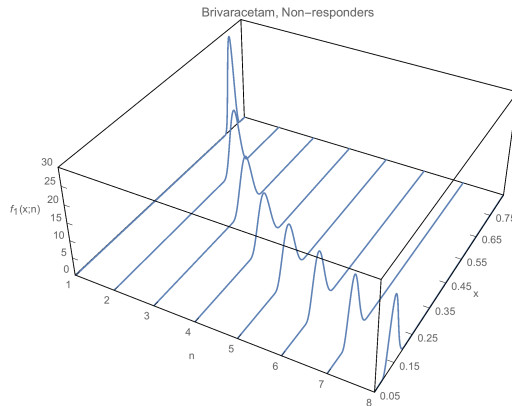
Figure 2: 1-PDF of non-responder sub-population, X_n^1, for different instants time $n \in \{1, 2, \ldots 8\}$.
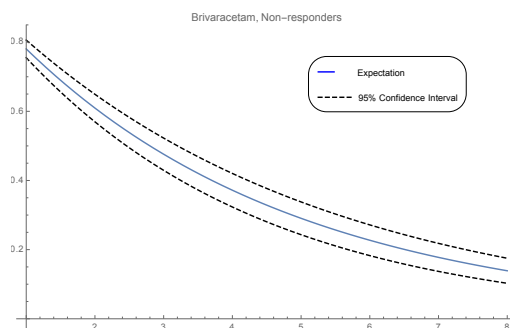
Figure 3: Expectation (line) and 95% confidence intervals (dotted line) of non-responder sub-population, X_n^1, for different instants time $n \in \{1, 2, \ldots 8\}$.

4 Conclusions

In this contribution we propose a randomized model to study the effectiveness of anti-epileptic drugs in epileptic patients. To solve this mathematical model we apply the Random Variable Transformation technique in order to compute the first probability density function of the solution stochastic process. Particularly, the distribution of the number of non-responder patients is computed. The probability density function gives us a full probabilistic description of the solution in every instant time t. Moreover, from it the mean and the variance can be easily derived, and then, confidence intervals which allow us to do predictions. Finally, a numerical example is drawn to shown the capability of the theoretical results previously established.

Acknowledgements

References

[1] I. Barrachina-Martínez, D. Vivas-Consuelo and A. Piera-Balbastre. Budget Impact Analysis of Brivaracetam Adjunctive Therapy for Partial-Onset Epileptic Seizures in Valencia Community, Spain. *Clin Drug Investig*, 38(4):353–363, 2018.

[2] Epilepsy Overview. Encyclopedia of the Neurological Sciences. Academic Press, 2003.

[3] T.T. Soong, Random Differential Equations in Science and Engineering. New York, Academic Press, 1973.

[4] J.-C. Cortés, A. Navarro-Quiles, J.-V. Romero, M.-D. Roselló. Randomizing the parameters of a Markov chain to model the stroke disease: A technical generalization of established computational methodologies towards improving real applications. *J. Comput. Appl. Math*, 324:225–240, 2017.

[5] S. Borghs. Brivaracetam is a third-generation anti-epileptic drug indicated as adjunctive therapy (16-65 years old). UCB: Brivaracetam–NMA of branded AEDs, 2015

Weighted graphs to redefine the centrality measures

M. D. López[♭] [*], J. Rodrigo[†],C. Puente[†] and J. A. Olivas[‡]

(♭) Polytechnic University of Madrid,

Applied Mathematics Dept. School of Civil Engineering,

(†) Pontificia Comillas University,

Advanced Technical Faculty of Engineering,

(‡) University of Castilla-La Mancha,

Information Technologies and Systems Dept.

November 30, 2018

1   Introduction

Causality is an important notion in every field of science. In empirical sciences, causality is a useful way to generate knowledge and provide for explanations. When a quantum physicist calculates the probability of an atom absorbing a photon, he analyses this event as the cause of the atom's jump to an excited energy level; that is, he tries to establish a cause-effect relationship [1].

Causation is a type of relationship between two entities: cause and effect. The cause provokes an effect, and the effect is a consequence of the cause. Causality can be a direct process when A causes B and B is a direct effect of A, or an indirect one when A causes C through B, and C is an indirect effect of A.

The typical form of causality is A causes B and the classic form of conditionality is If A then B. Causality and conditionality are not only restricted

---

[*]e-mail: marilo.lopez@upm.es

to these formats. Synonyms of cause or effect may indicate causality. Statements like B is due to A, A produces B, etc., are some other ways of expressing causality, as well as there are some other forms of expressing conditionality, like B if A, or A if only B. Therefore, in order to study causality, these forms need also to be taken into account.

The use of causal graphs as a way to represent information has been very present in literature, as Pearl [2], Spirtes [3], or Sobrino et al. [4] exemplifies. These representations usually have a qualitative ponderation in the edges to represent causal intensity like always, can, sometimes.... On the other hand, there are studies about causality that use a numerical degree to weight edges in a graph, which supposed and advance in the study of causal graphs. One of these studies is the one presented by López et al. [5] to obtain the causality degree of several causal paths linking two nodes. In this paper we present as novelty the application of such weighted causal graphs to the detection of new centrality measures related to causality. These measures are focused in quantifying the idea of finding the most central vertex in a graph, for example taking into account the length of the paths derived from a node. If we have a weighted graph according to causality measures, we will be able to adapt these weights to get the central vertex in a graph and predict the strongest set of effects produced by a cause for example.

In section 2 we will explain these new definitions of centrality measures depicted by a practical example.

2   Weighted Centrality Measures

In this paper, we propose change the centrality measures of causal graph's vertices defined in [5] by using the weights of the incoming and outcoming edges of such vertices. This would create the idea of dynamic graph, as edges ponderation would not be the same though it would be calculated in base to their causality degree, so centrality would be calculated from a causal point of view.

With these premises, we establish the following definitions:

Definition 1: given a causal graph of n vertices, the weighted output degree of a vertex is the addition of all weights that come out of the vertex and the input weighted degree of a vertex is the addition of all the weights of the edges linking the vertex.

Definition 2: given a causal graph of n vertices, the weighted centrality

in the output degree of a vertex is the weighted output degree of such vertex divided by $n-1$, and the weighted centrality in the input degree of a vertex is the weighted input degree of such vertex divided by $n-1$.

Remarks:

1.   To calculate the weighted centrality output degree of a vertex i, we just have to calculate the addition of the elements of the row i of the weighting matrix $A_1 + ... + A_{n-1}$ [5] divided by the order of such matrix -1.

2.   To calculate the weighted centrality input degree of a vertex i, we just have to calculate the addition of the elements of the column i of the weighting matrix $A_1 + ... + A_{n-1}$ [5] divided by the order of such matrix -1.

Definition 3: given a causal graph of n vertices, the weighted proximity centrality in the output degree of a vertex is the addition of the degrees of all causal paths going from that vertex to the rest, divided by the number of vertices.

Definition 4: given a causal graph of n vertices, the weighted proximity centrality in the input degree of a vertex is the addition of the degrees of all causal paths going from all vertices to v, divided by the number of vertices.

Remarks:

1.   To calculate the weighted proximity centrality in the output degree of a vertex i, we just have to calculate the maximum of the coefficients of the element (i, j) of the matrix $A_1 + ... + A_{n-1}$ [5] and add these maximum elements from $j = 1$ till n, dividing the result by n.

2.   To calculate the weighted proximity centrality in the input degree of a vertex i, we just have to calculate the maximum of the coefficients of the element (i, j) of the matrix $A_1 + ... + A_{n-1}$ and add these maximum from $i = 1$ till n, dividing the result by n.

In all cases we are looking for the vertex or vertices with a highest centrality measure.

Example:

With these definitions, we propose a practical example based in the graph of figure 1 and calculate the weighted centrality of its vertices. We have 9 vertices,, and the following weights on each edge:

$w(v1, v2) = 0.95, w(v1, v4) = 0.9, w(v1, v9) = 0.85, w(v2, v3) = 0.95, w(v3, v4) = 0.5, w(v4, v5) = 0.6, w(v4, v6) = 0.6, w(v5, v8) = 0.6, w(v6, v7) = 0.6, w(v7, v9) = 0.95, w(v8, v9) = 0.6$.

So, the weighting matrix of this graph is: $M =$

$$\begin{pmatrix} 0 & \frac{19}{20} & 0 & \frac{9}{10} & 0 & 0 & 0 & 0 & \frac{17}{20} \\ 0 & 0 & \frac{19}{20} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{3}{5} & \frac{3}{5} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{3}{5} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{3}{5} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{19}{20} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{3}{5} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Figure 1.Causal graph to answer the question How smoking causes death? Automatically extracted from [4], [6].

If we apply the first observation of definition 1, we will obtain the following weighted centralities in the output degree:

v_1: $\frac{1}{8}\left(\frac{19}{20} + \frac{18}{20} + \frac{17}{20}\right) = \frac{27}{80}$, v_2: $\frac{1}{8}\frac{19}{20} = \frac{19}{160}$, v_3: $\frac{1}{8}\frac{1}{2} = \frac{1}{16}$, v_4: $\frac{1}{8}\left(\frac{3}{5} + \frac{3}{5}\right) = \frac{3}{20}$, v_5: $\frac{1}{9}\frac{3}{5} = \frac{1}{15}$, v_6: $\frac{1}{8}\frac{3}{5} = \frac{3}{40}$, v_7: $\frac{1}{8}\frac{19}{20} = \frac{19}{160}$, v_8: $\frac{1}{8}\frac{3}{5} = \frac{3}{40}$, v_9: 0.

So v_1 is the vertex with highest weighted centrality in the output degree.

Applying remark 2 of definition 2, we will obtain the following weighted centralities in the input degree:

v_1: 0, v_2: $\frac{1}{8}\frac{19}{20} = \frac{19}{160}$, v_3: $\frac{1}{8}\frac{19}{20} = \frac{19}{160}$, v_4: $\frac{1}{8}\left(\frac{9}{10} + \frac{1}{2}\right) = \frac{7}{40}$, v_5: $\frac{1}{8}\frac{3}{5} = \frac{3}{40}$, v_6: $\frac{1}{8}\frac{3}{5} = \frac{3}{40}$, v_7: $\frac{1}{8}\frac{3}{5} = \frac{3}{40}$, v_8: $\frac{1}{8}\frac{3}{5} = \frac{3}{40}$, v_9: $\frac{1}{8}\left(\frac{17}{20} + \frac{19}{20} + \frac{3}{5}\right) = \frac{3}{10}$.

So v_9 is the vertex with highest weighted centrality in the input degree.

The matrix $A_1 + ... + A_{n-1}$ of this graph is:

$$\begin{pmatrix} 0 & 0.95v1v2 & 0.9025v1v2v3 & 0.9v1v4 + 0.45125v1v2v3v4 & 0.54v1v4v5 + 0.27075v1v2v3v4v5 \\ 0 & 0 & 0.95v2v3 & 0.475v2v3v4 & 0.285v2v3v4v5 \\ 0 & 0 & 0 & 0.5v3v4 & 0 \\ 0 & 0 & 0 & 0 & 0.6v4v5 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Applying observation 1 of definitions 3 and 4, we will obtain the following weighted proximity centralities in the output degree:

v_1: $1/9 \left(0.95 + 0.9025 + 0.9 + 0.54 + 0.54 + 0.324 + 0.324 + 0.85\right) = 0.5923$,

v_2: $1/9 \left(0.95 + 0.475 + 2 \times 0.285 + 2 \times 0.171\right) = 0.2597$,

v_3: $1/9 \left(0.5 + 2 \times 0.3 + 2 \times 0.18\right) = 0.1622$,

v_4: $1/9 \left(2 \times 0.6 + 2 \times 0.36\right) = 0.2133$, v_5 and

v_6: $\frac{0.6}{9} = 0.0667$, v_7, v_8 and v_9: 0.

So v_2 is the vertex with a highest weighted proximity centrality in the output degree.

Applying observation 2 of definitions 3 and 4, we will obtain the following weighted proximity centralities in the input degree:

v_1: 0, v_2: $\frac{0.95}{9} = 0.1056$, v_3: $\frac{1}{9}$ $(0.9025 + 0.95) = 0.2058$, v_4: $\frac{1}{9}$ $(0.9 + 0.475 + 0.5) = 0.2083$,

v_5: $\frac{1}{9}$ $(0.54 + 0.285 + 0.3 + 0.6) = 0.1917$,

v_6: $\frac{1}{9}$ $(0.324 + 0.171 + 0.18 + 0.36 + 0.6) = 0.1817$,

v_7: $\frac{1}{9}$ $(0.324 + 0.171 + 0.18 + 0.36 + 0.6) = 0.1817$,

v_8: $\frac{1}{9}$ $(0.324 + 0.171 + 0.18 + 0.36 + 0.6) = 0.1817$,

v_9: $\frac{1}{9}$ $(0.85 + 0.16245 + 0.171 + 0.342 + 0.36 + 0.57 + 0.95 + 0.6) = 0.4451$

So v_9 is the vertex with a highest weighted proximity centrality in the input degree.

3   Conclusions

The problem in causal graphs of selecting the most causal path linking two nodes has been largely discussed and is not trivial. In this paper we have proposed a new approach by applying weighted centrality measures to select those nodes with highest degrees and in accordance, obtain the 'best' causal path between two nodes. This approach takes into account the relationship of a node with the ones surrounding him and the input and output edges, providing better results than the ones that we used before. In addition, for future works, it will serve us for three main goals:

The first one is to select the most important nodes according to its causal weight when creating a summary. The second would be when asking a question, select the causal path that links two nodes with the highest degree of causality to include those nodes in the answer of the question. The third use would be to remove redundant nodes in a causal graph. For instance in the graph included in [6] we had "tobacco use" and "smoking". With this measurement we are able to select the node with a highest weight to work with.

4   Acknowledgements

References

[1] M. Bunge, Causality: the place of the causal principle in modern science. Cambridge, Harvard University Press, 1959.

[2] J. Pearl, Causality, models, reasoning, and inference. Cambridge, Harvard University Press, 2000.

[3] P. Spirtes, C. Glymour, R. Causation, Prediction and Search. Massachusetts, MIT Press, 2000.

[4] Sobrino A., Puente C. and Olivas J. A. Extracting answers from causal mechanisms in a medical document *Neurocomputing*, 135:53–60, 2014.

[5] López M. D., Puente C., Rodrigo J. and Olivas J. A. Weighted graphs to model causality *Proceedings of the International Conference on Artificial Intelligence*, 297–301, 2017.

[6] Puente C., Sobrino A., Olivas J. A. and Merlo, R. Extraction,analysis and representation of imperfect conditional and causal sentences by means of a semi-automatic process *Proceedings of the Fuzzy Systems (FUZZ), IEEE International Conference*, 1–8, 2010.

Numerical solution to the random heat equation with zero Cauchy-type boundary conditions

J.-C. Cortés[b] , A. Navarro-Quiles[†], J.-V. Romero[b*]
and M.-D. Roselló[b]

(♭) Instituto Universitario de Matemática Multidisciplinar,

Universitat Politècnica de València, Spain.

(†) DeustoTech, Fundación Deusto,

Universidad de Deusto, Bilbao, Spain.

November 30, 2018

Abstract

In this work we deal with a random heat equation with initial condition considering zero mixed-type boundary conditions. We propose a random finite difference scheme to solve this problem and sufficient conditions are provided in order to stablish consistency and stability in adequate norms. Finally, theoretical findings are illustrated via an example.

1 Introduction

It is well known the random nature of the parameters of the majority of ordinary differential equations and partial differential equations due to the complexity of the problem or to measurements errors. In recent years the study and use of random models has been extended. One of the classical

---

*e-mail: jvromero@imm.upv.es

equations of mathematical physics is the heat equation. The heat equation has a great deal of application in many branches of sciences [2, 1, 7]. The finite difference method is useful to solve partial differential equations in the deterministic scenario [3, 4, 6, 5]

In this work we propose a random finite difference scheme to solve the random heat equation

$$u_t(x,t) = \beta u_{xx}(x,t), \quad t > 0, \ 0 \le x \le 1, \tag{1}$$

with initial condition

$$u(x,0) = u_0(x), \tag{2}$$

and zero-mixed boundary conditions

$$u_x(0,t) = 0, \quad u(1,t) = 0, \tag{3}$$

where u_t and u_x stand for the derivatives of u respect to t and x variables, respectively, and β is a random variable defined on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Provided method is consistent and stable.

Finally, we will provide numerical examples and the obtained results will be compared using another established methods.

2 Computing the numerical solution

To compute the numerical solution, in a first step time and space are discretized in intervals equally spaced, $t_n = n\Delta t$, $n = 0, 1, 2, \ldots$, and $x_k = x_0 + k\Delta x = k\Delta x$, $k = 0, 1, \ldots, M + 1$, respectively. These discretization given us a mesh. The unknowns of these mesh are given in $k = 0, 1, \ldots, M$, $n = 1, 2, \ldots$

The following approximation is used for the temporal derivative

$$\frac{\partial u}{\partial t} = \frac{u(x, t + \Delta t) - u(x,t)}{\Delta t} + \mathcal{O}(\Delta t).$$

For the spatial derivative we use

$$\frac{\partial^2 u}{\partial x^2} = \frac{u(x - \Delta x, t) - 2u(x,t) + u(x + \Delta x, t)}{\Delta x^2} + \mathcal{O}(\Delta x^2),$$
$$\frac{\partial u}{\partial x} = \frac{u(x + \Delta x, t) - u(x - \Delta x, t)}{2\Delta x} + \mathcal{O}(\Delta x^2).$$

Introducing the notation $u(x_0 + k\Delta x, n\Delta t) = u(x_k, t_n) = u_k^n$, and using last approximations to derivatives in (1) one gets, taking $r = \beta \frac{\Delta t}{(\Delta x)^2}$,

$$u_k^{n+1} = (1 - 2r)\, u_k^n + r u_{k+1}^n + r u_{k-1}^n, \quad k = 0, 1, \ldots, M, \tag{4}$$

where

$$u_k^0 = u_0(x_k), \qquad u_{M+1}^n = u(x_{M+1}, t_n) = u(1, t_n) = 0.$$

For $k = 0$ a ghost node is introduced in equation (4)

$$u_0^{n+1} = (1 - 2r)\, u_0^n + r u_1^n + r u_{-1}^n$$

It is determined using the approximation to left-boundary condition $u_x(0, t) = 0$, obtaining $u_{-1}^n = u_1^n$, $n = 0, 1, \ldots$

Summarizing, next random difference method has been constructed in order to solve numerically problem (1)–(3)

$$
\begin{aligned}
u_0^{n+1} &= (1 - 2r)\, u_0^n + 2r u_1^n \\
u_k^{n+1} &= (1 - 2r)\, u_k^n + r u_{k+1}^n + r u_{k-1}^n, \qquad k = 1, 2, \ldots, M,
\end{aligned}
\tag{5}
$$

where

$$
\begin{aligned}
u_k^0 &= u_0(x_k), && k = 0, 1, \ldots, M, \\
u_{M+1}^n &= 0, && n = 0, 1, \ldots,
\end{aligned}
$$

$$r = \beta \frac{\Delta t}{(\Delta x)^2}, \quad \beta \text{ is a random variable.}$$

The random finite difference scheme constructed, is mean square consistent introducing adequate norms and probability spaces. Under condition

$$\Delta t \leq \frac{(\Delta x)^2}{2\beta_1}, \qquad 0 < \beta(\omega) \leq \beta_1, \ \forall \omega \in \Omega, \tag{6}$$

the random finite difference scheme is stable in a mean square sense. Also, the order of convergence can be established.

3   An illustrative example

In this section we develope a numerical example for problem (1)–(3). We have chosen the random variable as $\beta \sim \text{Be}(1; 3)$. In this case $0 < \beta(\omega) <$
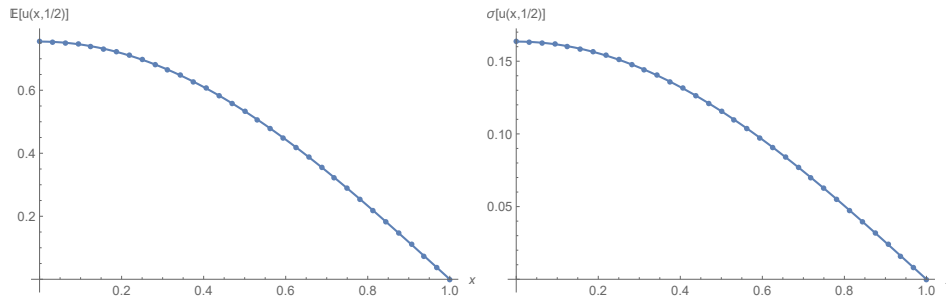
Figure 1: $\Delta x = 1/32$, $\Delta t = 1/2400$. Left: Expectation of both exact and numerical solution. Right: Variance of both exact and numerical solution.
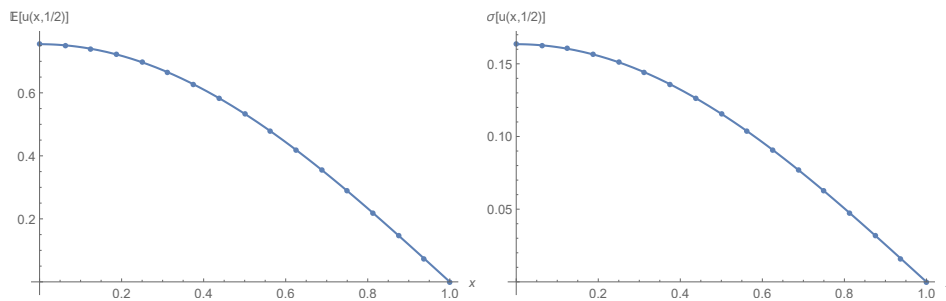
Figure 2: $\Delta x = 1/16$, $\Delta t = 1/600$. Left: Expectation of both exact and numerical solution. Right: Variance of both exact and numerical solution.
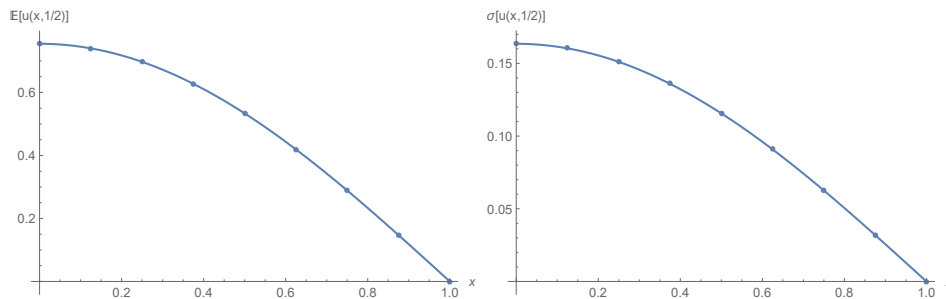
Figure 3: $\Delta x = 1/8$, $\Delta t = 1/150$. Left: Expectation of both exact and numerical solution. Right: Variance of both exact and numerical solution.

$1 = \beta_1$. Taking $u_0(x) = \cos\left(\frac{\pi x}{2}\right)$ as initial condition, the analytical solution is $u(x,t) = e^{-\frac{1}{4}\pi^2 \beta t} \cos\left(\frac{\pi x}{2}\right)$. Using the numerical scheme developed in previous section we obtain the the results displayed in Figures 1, 2 and 3 for the different meshes.

Comparing the numerical results with the analytical solution we obtain the results displayed in Table 1. This results are compatible with a numerical order of one in time and two in space. If we compare with Montecarlo we obtain Table 2. As we can observe the developed method gives better results than MonteCarlo.

M	N	Δx	Δt	mean error	standard deviation error
32	1200	1/32	1/2400	$3.78 \, 10^{-6}$	$1.51 \, 10^{-5}$
16	300	1/16	1/600	$1.51 \, 10^{-5}$	$6.05 \, 10^{-5}$
8	75	1/8	1/150	$6.06 \, 10^{-5}$	$2.42 \, 10^{-4}$

Table 1: Errors for random numerical difference scheme

simulations	mean error	standard deviation error
1000	$7.28 \, 10^{-3}$	$3.58 \, 10^{-3}$
10000	$1.60 \, 10^{-3}$	$4.78 \, 10^{-5}$
100000	$1.64 \, 10^{-4}$	$3.02 \, 10^{-4}$
1000000	$4.38 \, 10^{-5}$	$1.69 \, 10^{-4}$

Table 2: Errors for Montecarlo method

4 Conclusions

We have introduced randomness into the diffusion coefficient of the heat flow model and we have proposed a random finite difference scheme (RFDS) for solving this model with zero-mixed boundary conditions. The constructed method is mean square consistent. Sufficient conditions are provided for the mean square stability of the RFDS. The numerical experiments show that the proposed RFDS gives reliable approximations for the mean and the standard deviation of the solution stochastic process.

Acknowledgements

References

[1] F. Kreith, R. M. Manglik, and M. S. Bohn. *Principles of Heat Transfer.* Cengage Learning, Stamford, 2011.

[2] J. H. Lienhard. *A Heat Transfer Textbook.* Dover Books on Engineering. Dover Publications, 2011.

[3] G. D. Smith. *Numerical Solution of Partial Differential Equations: Finite Difference Methods.* Oxford Applied Mathematics and Computing Science Series. Clarendon Press, Oxford, 1986.

[4] J. C. Strikwerda. *Finite Difference Schemes and Partial Differential Equations.* SIAM: Society for Industrial and Applied Mathematics, New York, 2004.

[5] J. W. Thomas. *Numerical Partial Differential Equations: Finite Difference Methods*, volume 22. Springer Science & Business Media, New York, 2013.

[6] J.W. Thomas. *Numerical Partial Differential Equations: Conservation Laws and Elliptic Equations.* Texts in Applied Mathematics. Springer New York, 2013.

[7] D. V. Widder. *The Heat Equation*, volume 67. Academic Press, 1976.

A Multistate Model for Non Muscle Invasive Bladder Carcinoma

C. Santamaría[†][*], B. García-Mora[†], and G. Rubio[†]

(†) Instituto Universitario de Matemática Multidisciplinar,

Universitat Politècnica de València

November 19, 2018

1 Introduction

Bladder cancer is a challenge for urology and efforts in each step of the diagnosis and treatment of the patients should be implemented to improve oncological results. Bladder cancer is the most common cancer in the urinary tract. It is classified into two types: non-muscle-invasive bladder cancer (NMIBC) and muscle invasive tumor. In a first diagnosis, 75-85% of bladder cancers are non-muscle-invasive bladder cancer (NMIBC) and 30-80% of these NMIBC patients have a recurrence of the disease and 1-45% of these patients progress to muscle-invasive tumor.

The non-invasive bladder cancer (NMIBC) is the tumor that generates the greatest economic cost of all cancers due to its low mortality and long follow-up period. Different strategies are needed to be tackled to reduce this cost such as: to identify molecular markers that allow the use of more specific treatments protocols depending on the particular characteristics of each tumor in order to avoid unnecessary treatments; to establish cost-effective follow-up and treatment protocols and to avoid the complications generated by this tumor.

Once a diagnosis of a new non-muscle-invasive bladder cancer (NMIBC) is done in a patient, the only way to establish a successful specific follow-up and

---

[*]e-mail: crisanna@imm.upv.es

286

treatment protocol is based on a good prediction of the recurrence process and the progression process of the primary tumor. The classical Cox model has been widely used for this purpose [1], but this approach is not valid for the assessment of this kind of events because the time between recurrences in the same patient may be strongly correlated. Then new approaches for the risk estimation of recurrence and progression are needed to lead to individualized monitoring and treatment plan.

2   Successive steps towards the full model

Multi-state stochastic processes are a convenient framework for modeling the evolution of disease processes and the statistical Flowgraph approach [2] is a convenient tool to perform the task. Flowgraph technique is specifically suited for semi-Markov processes. This methodology in the context of stochastic networks was introduced by Butler and Huzurbazar in [3] and widely applied in many contexts of multistate stochastic networks and biomedical applications. See [2] for an introductory book on statistical Flowgraph models.

Our approach develops what we started in [4], with the aim of modeling the recurrrence-progression process, visualized as a multi-state process, (see Figure 1).
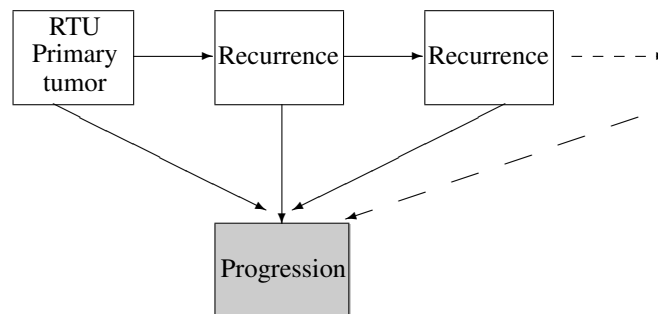
Figure 1: Recurrence-progression process

We found some difficulties within this framework. On one side, the management of covariates is not straightforward [5]. More importantly, the semi-Markov assumption implies independence among waiting time distributions. However, we need to relax this assumption, because recurrences in the same patient are not independent events, and in fact, the hypothesis of independence led us to unsatisfactory results.

We addressed the first problem in [6] and [7], using Erlang distributions, a kind of phase type distribution [8], following the approach suggested in [9].

Let us recall that the distribution $F(.)$ on $[0,\infty[$ is a phase–type distribution (PH-distribution) with representation (α, T) if it is the distribution of the time until absorption in a Markov process on the states $\{1, \ldots, m, m+1\}$ with generator

$$\begin{pmatrix} T & T^0 \\ 0 & 0 \end{pmatrix},$$

and initial probability vector (α, α_{m+1}) where α is a row m-vector.

In this way, we were able to incorporate covariates in a relatively simple manner. Computations are quite tractable combined with the flowgraph methodology.

On the other hand dependency management is not achieved with generality within the framework of Flowgraph methodology, only in a few special cases. A successful approach is required when the conditional independence assumption for waiting times does not hold. We addressed this problem in [10] and [11] using the Markovian Arrival Process (MAP), that has the relevant property of dependence between consecutive inter–arrival times in a process with multiple events. These processes are in some way a generalization of phase type distributions.

A Markovian Arrival Process (MAP) (π, D_0, D_1) is an irreducible Markov chain with a finite state space S, initial vector π and a generator matrix Q which can be represented as $Q = D_0 + D_1$ where,

- $D_1 \geq 0$, $D_1 \neq 0$

- $D_0(i, j) \geq 0$ for $i \neq j$

- (π, D_0) is a phase-type distribution.

Two useful references for both phase-type distribution and the Markovian Arrival Process are [12] and [13].

However, the introduction of MAPs entails the difficulty of not having a way to include covariates nor censoring in MAPs. Our solution for the covariates issue was discussed in [14]. The management of censored times is addressed in the current work, on the occasion of the construction of a full flowgraph model of the recurrence progression process, using a database of patients with NMIBC from La Fe University Hospital of Valencia (Spain).

3 A flowgraph model for NMIBC

The database consisted of 960 patients from University Hospital "La Fe". Valencia (Spain). Follow-up period was1995-2010 with a median follow-up of 48.1 $(3 - 160)$ months. All had a primary superficial transitional cell carcinoma of the bladder. Tumor was removed by means of TUR. 435 patients underwent a recurrence, 26 a progression, and 499 had censored times. Then, 62 patients were lost. From the remaining 373 patients, 226 patients underwent a second recurrence 19 underwent a progression and 128 were censored. 30 patients were lost. From the remaining 196 patients, 4 underwent a progression. In our model we consider up to two recurrences and progression. Main objective is to model the overall risk of progression.

For that, we have to calculate the PDF of the specific path of the overall time to progression. Taking into account the previous steps, the idea is to build the appropriate phase-type distribution for each transition. All paths between two consecutive states can be managed according to the general flowgraph methodology, using our approach with Erlangs distributions. As in path $0 \rightarrow 1 \rightarrow 2$ there is dependence, we model that path with a MAP. To do this, the matrix D_0 is made up as if all the states were independent, using the flowgraph methodology. And π and matrix D_1 are constructed using a maximum likelihood procedure that allows us to incorporate the information of censored times. It should be noted that as far as we know, the issue of censored data with MAPs has not yet been addressed.

In this way we can obtain a representation (α_k, T_k), $k = 1, \ldots, 5$ of a phase-type distribution in each branch of the graph
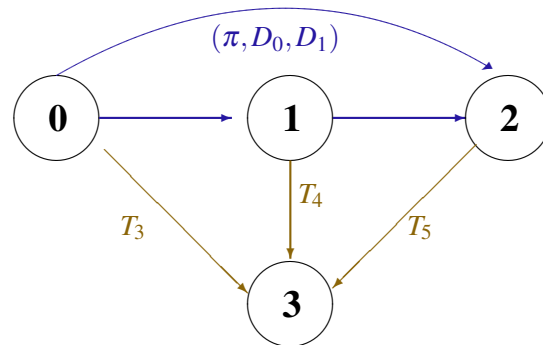
Then path $0 \rightarrow 1 \rightarrow 2$ is fitted using MAP and Flowgraph technics. T_3, T_4 and T_5 are fitted using Flowgraph technic.

In short, we have managed to extend the flowgraph methodology beyond the semi-Markovian framework, simplifying the incorporation of covariates and with-

out excluding censored times. All of which has allowed us to build a multistate model of the evolution of NMIBC.

References

[1] Sylvester, R. J., van der Meijden, A. P., Oosterlinck, W., Witjes, J. A., Bouffioux C., Denis, L., Newling, D. W., and Kurth, K. Predicting recurrence and progression in individual patients with stage ta t1 bladder cancer using EORTC risk tables: a combined analysis of 2596 patients from seven EORTC trials. *European Urology*, 49 (2006) 475–7.

[2] Huzurbazar, A. Flowgraph Models for Multistate Time-To-Event Data, Wiley, New York, 2005.

[3] Butler, R. W. and Huzurbazar, A. V. Stochastic network models for survival analysis. *Journal of the American Statistical Association* 92 (1997) 246–57.

[4] Rubio, G., García-Mora, B., Santamaría, C. and Pontones, J.L. A flowgraph model for bladder carcinoma. *Theoretical Biology and Medical Modelling 11 (2014) (Suppl 1) : S3.*

[5] Huzurbazar, A. and Williams, B. Incorporating Covariates in Flowgraph Models: Applications to Recurrent Event Data. *Technometrics*, 2010; 52(2), 198-208.

[6] Rubio, G., García-Mora, B., Santamaría, C. and Santonja, F. *Incorporating covariates in a flowgraph model for bladder carcinoma.* International Work-Conference on Bioinformatics and Biomedical Engineering (IWBBIO 2014), April 7-9, Granada, Spain.

[7] García–Mora , B., Santamaría, C., Rubio, G. and Pontones, J.L. Bayesian prediction for flowgraph models with covariates. An application to bladder carcinoma. *Journal of Computational and Applied Mathematics*, 291 (2016) 85–93.

[8] Neuts, M. F. Matrix Geometric Solutions in Stocastic Models. An Algoritmic Approach, The Johns Hopkins University Press, Baltimore, 1981.

[9] Pérez-Ocón, R. Modeling lifetimes using phase-type distributions, in: Risk, reliability and societal safety, Proceedings of the European Safety and Reliability Conference, ESREL (2007), June 25–27, Stavanger, Norway.

[10] Rubio, G., García-Mora, B., Santamaría, C. and Pontones, J.L. Incorporating multiple recurrences in a flowgraph model for bladder carcinoma. International Work-Conference on Bioinformatics and Biomedical Engineering (IWBBIO 2015), April 15-17, Granada, Spain.

[11] Rubio G., García–Mora B., Santamaría C. and Pontones J.L. Modeling dependence in multistate processes. International Work–Conference on Bioinformatics and Biomedical Engineering (IWBBIO 2016), April 20–22. Granada (Spain).

[12] Breuer, L. and Baum, D. An Introduction to Queueing Theory and Matrix-Analytic Methods, Springer, Dordrecht, Netherlands, 2005.

[13] Buchholz P., Kriege J. and Felko I. Input Modeling with Phase–Type Distributions and Markov Models. Theory and Applications, Springer Cham Heidelberg New York Dordrecht London, 2014.

[14] Santamaría C., García–Mora B., Rubio G. Introducing Covariates in Reliability Models by Markovian Arrival Processes. Modelling for Engineering & Human Behaviour 2017, Instituto Universitario de Matemática Multidisciplinar, Valencia, Spain.

Birth rate and population pyramid: A stochastic dynamical model

Joan C. Micó [a], David Soler [a], Maria T. Sanz [b*], Antonio Caselles [c], Salvador Amigó [d]

[a] *Institut Universitari de Matemàtica Multidisciplinar, Universitat Politècnica de València, Camí de Vera s/n, 46022 València, Spain.*

[b] *Departament de Didàctica de la Matemàtica. Universitat de València, Avda. Tarongers 4, 46022 València, Spain.*

[c] *IASCYS member. Departament de Matemàtica Aplicada, Universitat de València, Dr. Moliner 50, 46100 Burjassot, Spain.*

[d] *Departament de Personalitat, Avaluació i Tractaments Psicològics. Universitat de València, Av. Blasco Ibáñez 21, 46010. València, Spain.*

1. Introduction

One of the fundamental problems of some developed societies is the imbalance of their population pyramids. This imbalance can threaten the population and economic sustainability of these societies. The case of Spain and Norway, which are presented here as a case of application, is representative of this problem: its population pyramid tends to contract for working ages and to grow for retirement ages (see Fig. 1a and 1b).
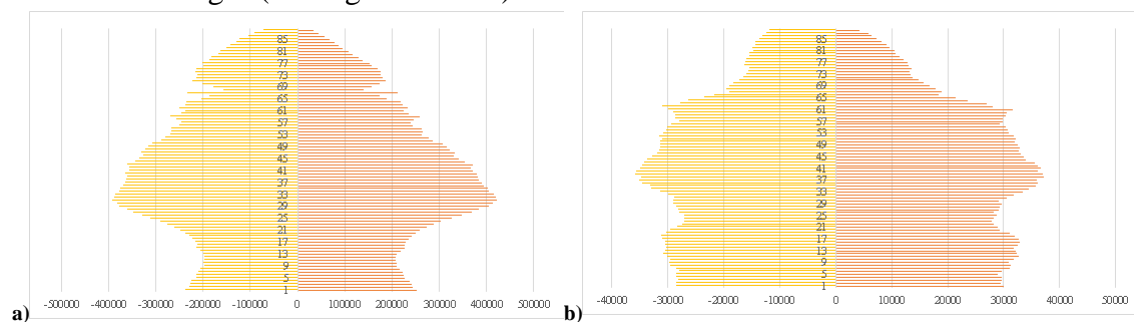
Fig. 1. Pyramid population, female (red) and male (orange) population for. a) Spain in 2017; b) Norway in 2017.

A main factor considered by demographers as the cause of the problem is the low birth rate [1, 2]. From this assumption, one could ask what would be the appropriate birth rate for a society so that, within a reasonable period, its population pyramid could reach a wished equilibrium. International agreements do not specify targets in terms of values of the dependency ratio. However, in 2005, 66 per cent of Governments were concerned about the size of their working-age population, and 52 per cent of them reported that population ageing represented an issue of major concern [3].

The aim of this work is to use the demographic model developed by the authors [4] to solve the described problem. However, in this work the model is presented slightly modified in order to include the birth rate as a control variable. In addition, the model is presented in its stochastic formulation, which is one step beyond the cited work, in which its validity in deterministic formulation is demonstrated. Once the model has been validated for the case of Spain, it is used to determine what the future evolution of the birth rate should be in order to achieve a balanced population pyramid, i.e., a convenient dependency ratio. This evolution of the birth rate, which is considered optimal, is calculated with the use of a genetic algorithm.

2. Demographic Model

---

[*] Corresponding author. Tel.: 963983285

E-mail addresses: jmico@mat.upv.es (J. C. Micó); dsoler@mat.upv.es (D. Soler); m.teresa.sanz@uv.es (M. T. Sanz); antonio.caselles@uv.es (A. Caselles); salvador.amigo@uv.es (S. Amigó)

The starting point of this demographic model is the model presented by [4]. In its continuous form it is constituted by the following equations,

$$\frac{\partial w_i(t,x)}{\partial t} + c\frac{\partial w_i(t,x)}{\partial x} = -b_i(t,x) \cdot w_i(t,x) + m_i(t,x) \tag{1}$$

$$w_i(t,0) = \int_0^\infty a_i(t,x) \cdot w_2(t,x)\, dx \tag{2}$$

$$w_i(t_0,x) = u_i(x) \tag{3}$$

Where, $i = 1$ represents men and $i = 2$ women.

Eq. 1 is a von Foerster-McKendrick equation that determines the dynamics of population density depending on time and age, $w_i(t,x)$, where $b_i(t,x)$ represents the death rate and $m_i(t,x)$ the migratory balance. Eq. 2 represents the boundary condition, that is, births at $x = 0$ ($a_i(t,x)$ represents the fertility rate). Eq. 3 is the initial condition, that is, the initial population density, $u_i(x)$, at $t=t_0$.

Based on this model, a simplification hypothesis is that the fertility rate can be calculated according to Eq. 4:

$$a_i(t,x) \approx \bar{a}_i(x) \cdot prpn_i(t) \cdot tnac(t) \cdot popt(t) \tag{4}$$

Where $prpn_i(t)$ is the proportion of men or women born (according to $i = 1$ or 2, respectively), that is, births per sex ($birs_i(t)$) divided by the total number of births ($birt(t)$) (see Eq. 5 below); $tnac(t)$ is the birth rate, i.e., total numbers of births ($birt(t)$) divided by the total population ($popt(t)$) (see Eq. 6 below); and $\bar{a}_i(x)$ is the ratio between the fertility rate and births (see Eq. 7 below).

$$prpn_i(t) = \frac{birs_i(t)}{birt(t)} \tag{5}$$

$$tnac(t) = \frac{birt(t)}{popt(t)} \tag{6}$$

$$\bar{a}_i(x) = \frac{a_i(x)}{birs_i(t_0)} \tag{7}$$

About the Eq.4, there are variables which can be calculated from other as Eq. 16 and 17 show,

$$pop(t) = \int_0^{+\infty}\big(w_1(t,x) + w_2(t,x)\big)\, dx \tag{8}$$

On the other hand, given that the migratory balance is defined by the difference between immigration and emigration, in [4] it is considered as the product of absolute migrations dependent on time and the proportions per unit of age of the population of the system (Eq. 9, 10 and 11). In our work, we consider Eq. 12 and 13.

$$m_i(t,x) = ynmi_i(t,x) - emig_i(t,x) \tag{9}$$
$$ynmi_i(t,x) = f_i(x) \cdot gryn_i(t,x) \cdot w_i(t,x) \tag{10}$$
$$emig_i(t,x) = g_i(x) \cdot grem_i(t,x) \cdot w_i(t,x) \tag{11}$$

Where,

$$grem_i(t,x) = \frac{\sum_{t=t_0}^{tn} \frac{emig_i(t,x)-emig_i(t_0,x)}{emig_i(t_0,x)}}{t_n-t_0} \tag{12}$$

$$gryn_i(t,x) = \frac{\sum_{t=t_0}^{tn} \frac{ynmi_i(t,x)-ynmi_i(t_0,x)}{ynmi_i(t_0,x)}}{t_n-t_0} \tag{13}$$

A similar simplification has been done for the death rates [4], which are defined as a function of age as $b_i(x)$ (Eq. 14). In our work, we consider Eq. 15 and 16.

$$deat_i(t,x) = b_i(x) \cdot w_i(t,x) \tag{14}$$

$$deat_i(t,x) = b_i(x) \cdot grde_i(t,x) \cdot w_i(t,x) \tag{15}$$

$$grde_i(t,x) = \frac{\sum_{t=t_0}^{tn} \frac{deat_i(t,x)-deat_i(t_0,x)}{deat_i(t_0,x)}}{t_n-t_0} \tag{16}$$

With these considerations on the initial model, the following equations are obtained:

$$\frac{\partial w_i(t,x)}{\partial t} + c\frac{\partial w_i(t,x)}{\partial x} = (-b_i(x) \cdot grde_i(t,x) + f_i(x) \cdot gryn_i(t,x) - g_i(x) \cdot grem_i(t,x)) \cdot w_i(t,x) \tag{17}$$

$$w_i(t,0) = bir(t) \cdot \frac{n_i(t)}{n_1(t)+n_2(t)} \cdot \int_0^{+\infty}(w_1(t,x)+w_2(t,x))\,dx \cdot \int_0^{+\infty}\bar{a}_i(x) \cdot w_2(t,x)\,dx \tag{18}$$

$$w_i(t_0,x) = u_i(x) \tag{19}$$

3. Model Validation

The validation of the model is done for Spain in the 2008-2016 period, since it is from these years that information is obtained, in order to fit the equations of the model, based on the Eurostat statistical database [5].

The model has been written as a set of differential and functional equations. The solutions have been calculated with the Euler Method following [6, 7], which explain that the Euler Method is more adequate to solve such equations. In the case of the integral in Eq. 9, it is calculated through the Simpson Composite Rule. This approach results in a set of finite difference equations that has been programmed in Visual Basic 6.0 using Sigem [8, 9].

The corresponding validation has been performed numerically by calculating the determination coefficients and the random residuals tests for the case of the deterministic validation. Figures 2 to 7 show some of the obtained results compared with the historical data. The validation process is considered successful for the deterministic validation because the determination coefficients, R^2, are very high, and the maximum relative error does not exceed 2.3% in any case. Due to space limitation, we do not show the stochastic validation.
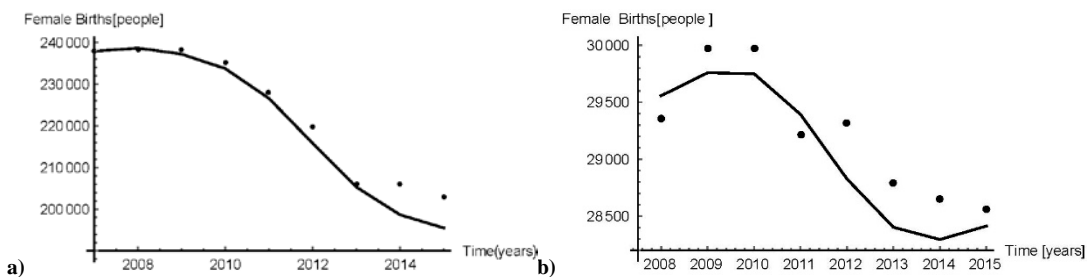
Fig. 2. Female Births in the 2008-2016 period. a) Deterministic validation for Spain, R^2=0.987666; b) Deterministic validation for Norway, R^2=0.842235.

Note that the fitting of the functions to historical data of each one of the input variables (for each country) has been made through Mathematica 11.00 [10] with the NonLinearModelFit package and with the Regint function fitter [8, 9].
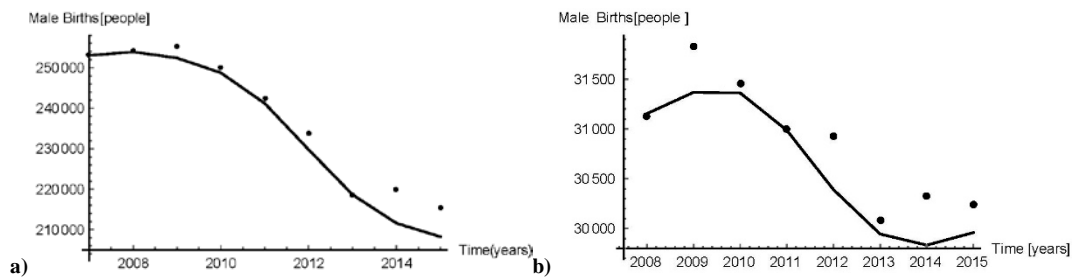
Fig. 3. Female Births in the 2008-2016 period. a) Deterministic validation for Spain, R^2=0.981417; b) Deterministic validation for Norway, R^2=0.887119.
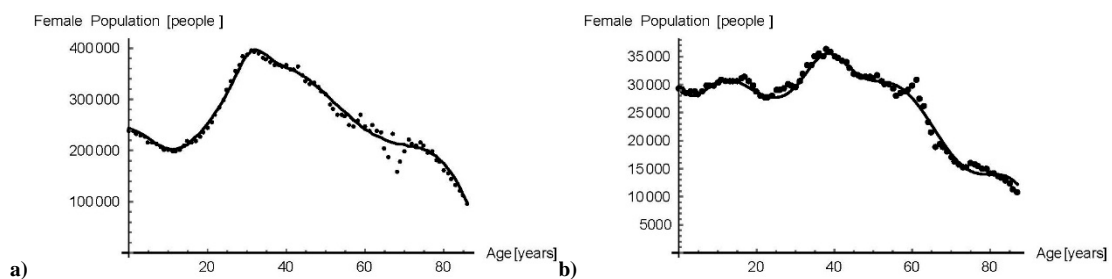
Fig. 4. Female Population by age in 2008. a) Deterministic validation for Spain, R^2=0.970624; b) Deterministic validation for Norway, R^2=0.982761.
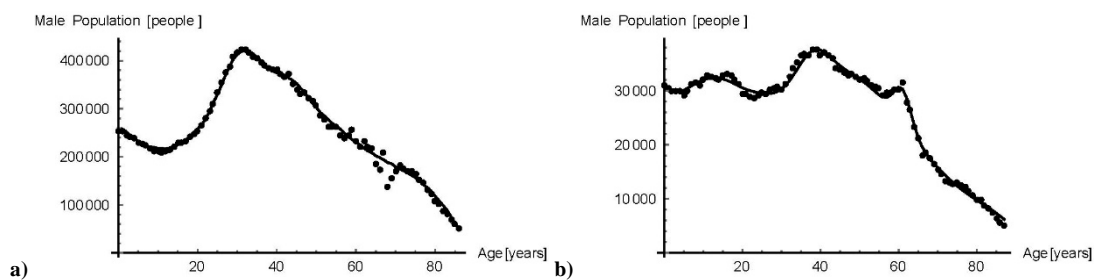
Fig. 5. Male Population by age in 2008. a) Deterministic validation for Spain, R^2=0.990115; b) Deterministic validation for Norway, R^2=0.994378.
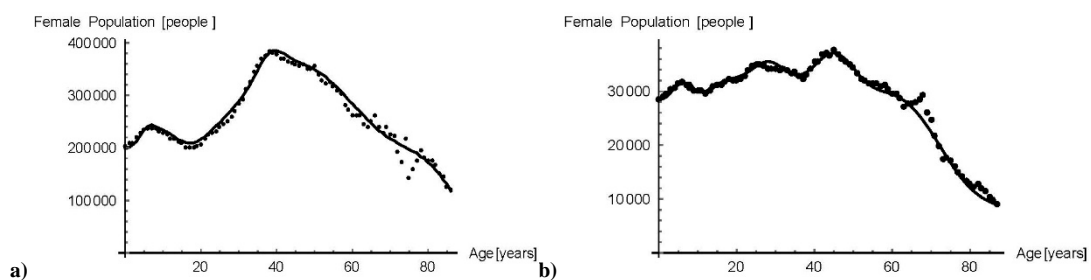
Fig. 6. Female Population by age in 2016. a) Deterministic validation for Spain, R^2=0.981417; b) Deterministic validation for Norway, R^2=0.887119.
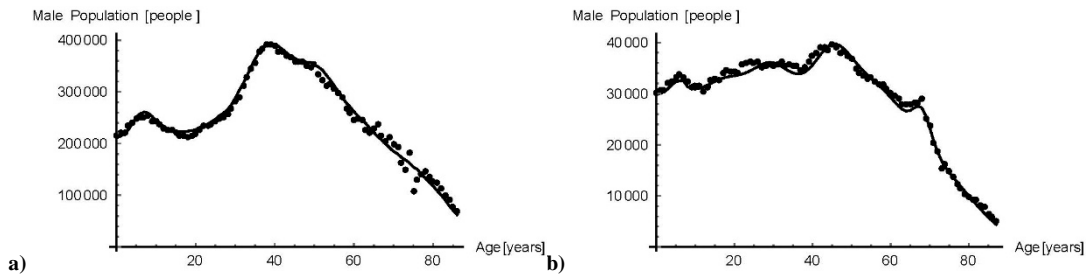
Fig. 7. Male Population by age in 2016. a) Deterministic validation for Spain, R^2=0.981417; b) Deterministic validation for Norway, R^2=0.887119.

4. Simulation and decision making: optimization with a genetic algorithm contrast to strategies and scenarios

A genetic algorithm (GA) is automatically programmed by the Sigem automatic programing tool [8, 9]. The GA allows optimizing, at each step of time, a previously defined objective variable from other variables included in the model. In this work, the goal is to maximize the proportion of the working age population over the population that does not work (Eq. (11)) in the 2017-2117 period for the case of Spain and Norway, where x_m is the age from which you can work and x_M the retirement age.

$$obje(t) = \frac{\sum_i \int_{x_m}^{x_M} w_i(t,x)\, dx}{\sum_i \left(\int_0^{x_m} w_i(t,x)\, dx + \int_{x_M}^{+\infty} w_i(t,x)\, dx \right)} \tag{20}$$

In Eq. 20 x_m is the starting age to be able to work and x_M the retirement age.

There is an important difference between the optimization with the GA and others, such as the quasi-optimization obtained with the method of strategies and scenarios (SS) [11, 12]. The difference is that the input variables that have been fitted with respect to time ($tnac(t)$) (with SS) are now calculated by the model by means of the GA, which searches for the optimal strategy to reach the goal, that is, the optimal value of $tnac(t)$ to achieve maximizing $obje(t)$. Due to the available space, only the results obtained with the model defined in its deterministic formulation are shown in Fig. 8 and 9. They reveal that the dependency rate increase with both optimization. To modify the population pyramid and to get more people of working age it is necessary to modify the trend of the birth rate. In the case of SS optimization, Fig. 8b and Fig. 9b show that the birth rate must be increased since 2017 to 2046 (Spain) and 2058 (Norway). Fig. 10 and 11 present that the absolute values of the corresponding births are increased respect to 2017 (Fig. 1) to obtain the goal proposed, in the case of Norway (Fig. 11), but the opposite case must be produced in the case of Spain (Fig. 10).
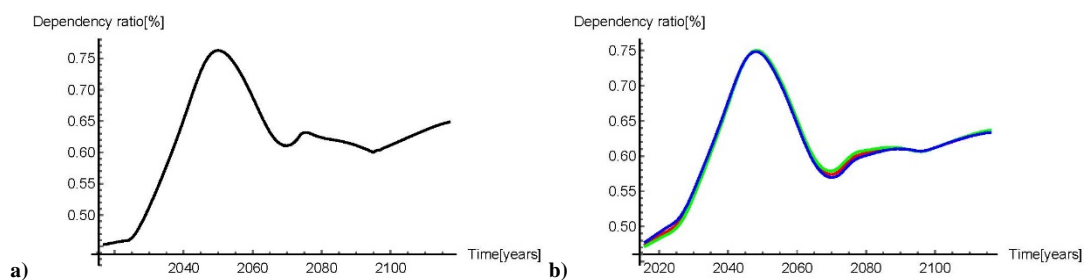
Fig. 8. Dependency ratio for Spain in the period 2017-2117. a) Optimization with GA; b) Optimization with SS, Optimist strategy: Increase the birth rate by 2% (blue), Tendency strategy: Maintain current tendency (red), Pessimistic strategy: Decrease the birth rate by 2% (green).
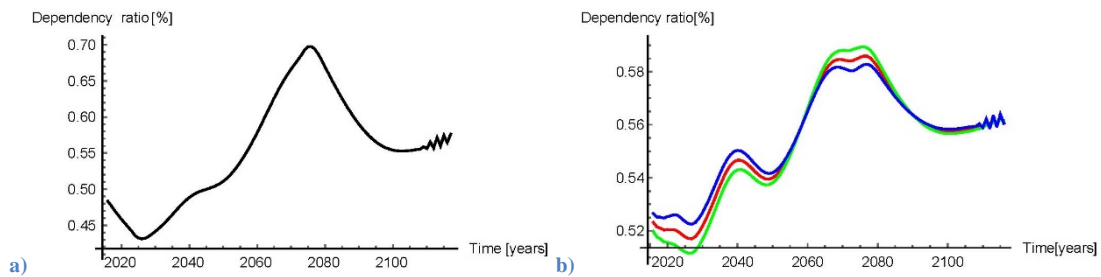
Fig. 9. Dependency ratio for Norway in the period 2017-2117. a) Optimization with GA; b) Optimization with SS, Optimist strategy: Increase the birth rate by 2% (blue), Tendency strategy: Maintain current tendency (red), Pessimistic strategy: Decrease the birth rate by 2% (green).
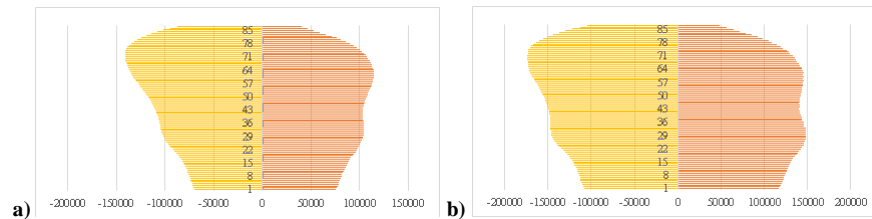
Fig. 10. Pyramid population, female (red) and male (orange) population for Spain in 2117. a) Optimization with GA; b) Optimization with SS.
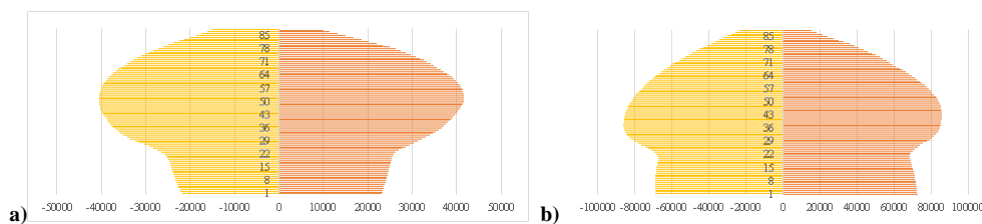
Fig. 11. Pyramid population, female (red) and male (orange) population for Norway in 2117. a) Optimization with GA; b) Optimization with SS.

References

[1] S. KC, W. Lutz, The human core of the shared socioeconomic pathways: Population scenarios by age, sex and level of education for all countries to 2100, Glob. Environ. Chang., 42 (2017), 181–192.

[2] http://ec.europa.eu/eurostat/statistics-explained/index.php/Population_structure_and_ageing/es (accessed 14.04.18).

[3] http://www.un.org/en/development/desa/population/publications/policy/world-population-policies-2005.shtml (accessed 14.04.18).

[4] J.C. Micó, A. Caselles, D. Soler, T. Sanz, E. Martínez, A Side-by-Side Single Sex Age-Structured Human Population Dynamic Model: Exact Solution and Model Validation, J. of Math. Soc., 32:4 (2008), 285-321

[5] http://ec.europa.eu/eurostat (accessed 14.04.18).

[6] K. Djidjeli, W.G. Price, P. Temarel, E.H. Twizell, Partially implicit schemes for the numerical solutions of some non-linear differential equations, Appl. Math. Comput. 96 (1998) 177–207.

[7] C. Letellier, S. Elaydi, L.A. Aguirre, A. Alaoui, Difference equations versus differential equations, a possible equivalence for the Rossler system? Phy. D: Nonlin. Phen. 195 (2004) 29–49.

[8] A. Caselles, A tool for discovery by complex function fitting, in: R. Trappl (Ed.), Cybernetics and Systems Research'98, Austrian Society for Cybernetic Studies, Vienna, 1998, pp. 787–792.

[9] A. Caselles, Modelización y Simulación de Sistemas Complejos (Modeling and Simulation of Complex Systems), Universitat de València, Valencia (Spain), 2008. Available in http://www.uv.es/caselles as well as SIGEM, (accessed 15.01.18).

[10] http://www.wolfram.com/mathematica/ (accessed 14.04.18).

[11] D. Soler, M. T. Sanz, A. Caselles, J. C. Micó, A stochastic dynamic model to evaluate the influence of economy and well-being on unemployment control, J. of Comp. and App. Math. 330 (2018) 1063–1080.

[12] M. T. Sanz, A. Caselles, J. C. Micó, D. Soler, A stochastic dynamical social model involving a human happiness index, J. of Comp. and App. Math. 340 (2018) 231–246.

Application of the finite element method in the analysis of oscillations of rotating parts of machine mechanisms

Petr Hrubý[♭] , Dana Smetanová[†] [*]

(♭) Department of Mechanical Engineering, Institute of Technology and Business,

Okružní 517/10, České Budějovice, Czech Republic,

(†) Department of Informatics and Natural Sciences, Institute of Technology and Business,

Okružní 517/10, České Budějovice, Czech Republic.

November 14, 2018

1 Introduction

The most endangered parts are rotating components in engineering constructions. Reliable shaft operations endanger two limit states in particular. In the vicinity of resonance there is an enormous increase in amplitudes of state quantities and the yield strength of the material. These conditions often occur with the coupling shafts of the cardan mechanisms. The torque is transmitted here over long distances. The shafts are long and slim and are prone to transverse bending. The gearbox shafts are compact and operating at a sufficient distance from the resonant area. In this case they are endangered by fatigue fractures. They need to be checked for safety to fatigue. A similar situation to gearboxes is with gear pump shafts. The authors have long been concerned with the design of joint shafts and gear pumps in cooperation with engineering companies.

Mathematical models pumps lead to solutions from the field of linear algebraic equations. In the case of bending oscillations, the equations of

---

[*]e-mail: smetanova@mail.vstecb.cz

motion of the basic element is a partial differential equations of the 4th order for the variables x and t.

The solving the equations is to use an analytical solution for simpler cases. For a real shaft profile, one of the most sophisticated methods is needed.

In the cases solved by us, the finite element method and transfer matrix method proved to be successful.

The Finite Element Method seems to us very good to obtain the modal and spectral properties and our eigen oscillation frequencies [1]. The continuum is discretized using this method. In solution, it does not provide infinitely many eigen frequencies because they vanish to the finite number of elements, and for higher frequencies it loses precision. In this way, frequencies can be examined not only for strength (lower frequency) and noise (higher frequencies near the operating area, which is particularly important for cars). The disadvantage of using the finite element method is the clains on the hardware because the matrix size is increased in the calculation. Our aim is therefore to develop methods that can be successfully applied within the limited resources of small and medium-sized enterprises with which we cooperate.

At present, we have a model for solving eigen oscillations in rotating parts of machine mechanisms. The next step is the generalization to solve forced oscillations. For the calculations, we also use variational analysis methods. The mathematical model is described by Euler-Lagrange equations (the equations of motions). Thus, the model formulation is provided by the Hamilton equations, which we obtain through the Legendre transformation of coordinates.

2   Mathematical model of shaft element

At present, we have a model for solving eigen oscillations in rotating parts of machine mechanisms. The next step is the generalization to solve forced oscillations. For the calculations, we also use variational analysis methods. The mathematical model is described by Euler-Lagrange equations (the equations of motions). Thus, the model formulation is provided by the Hamilton equations, which we obtain through the Legendre transformation of coordinates.
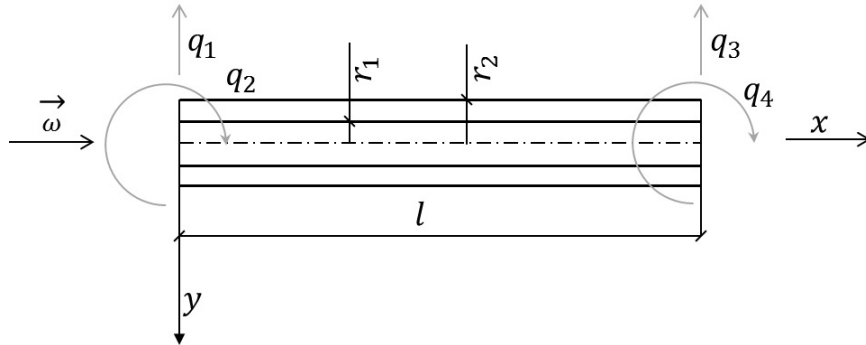
Fig. 1. The element of shaft in state of combined bending-gyratory vibration

The deflection y of the range $0 \leq x \leq l$ (see Fig. 1) is expressed as

$$y(x,t) = \sum_{i=1}^{4} q_i(t)\Phi_i(x) \tag{1}$$

where $\Phi_i(x)$ are 3rd order polynomials

$$\Phi_i(x) = a_{3i}x^3 + a_{2i}x^2 + a_{1i}x + a_{0i}$$

with coefficient $a_{3i}, a_{2i}, a_{1i}, a_{0i}$.

Boundary conditions:

$$\Phi_1(0) = 1, \ \ \Phi_1(l) = 0, \ \ \Phi_1'(0) = 0, \ \ \Phi_1'(l) = 0,$$
$$\Phi_2(0) = 0, \ \ \Phi_2(l) = 0, \ \ \Phi_2'(0) = 1, \ \ \Phi_2'(l) = 0,$$
$$\Phi_3(0) = 0, \ \ \Phi_3(l) = 1, \ \ \Phi_3'(0) = 0, \ \ \Phi_3'(l) = 0,$$
$$\Phi_4(0) = 0, \ \ \Phi_4(l) = 0, \ \ \Phi_4'(0) = 0, \ \ \Phi_4'(l) = 1.$$

Polynomials $\Phi_i(x)$:

$$\Phi_1(x) = 2\left(\frac{x}{l}\right)^3 - 3\left(\frac{x}{l}\right)^2 + 1,$$
$$\Phi_2(x) = \frac{x^3}{l^2} - 2\frac{x^2}{l} + x,$$
$$\Phi_3(x) = -2\left(\frac{x}{l}\right)^3 + 3\left(\frac{x}{l}\right)^2,$$
$$\Phi_4(x) = \left(\frac{x}{l}\right)^3 - \frac{x^2}{l}.$$

The potential energy of an element:

$$E_p = \frac{1}{2}EJ \int_0^1 \left(\frac{\partial^2 y}{(\partial x)^2}\right)^2 dx \tag{2}$$

subtituing (1) to (2) we get

$$E_p = \frac{1}{2}EJ \int_0^1 \left([\Phi''(x)] [q]\right)^2 dx, \tag{3}$$

where $[\Phi''(x)] = [\Phi_1''(x), \Phi_2''(x), \Phi_3''(x), \Phi_4''(x)]$ and $[q] = [q_1, q_2, q_3, q_4]^T$.
The kinetic energy of above element:

$$E_k = \frac{1}{2}\mu \int_0^1 \left(\left(\frac{\partial y}{\partial t}\right)^2 + (y\omega)^2\right) dx + \frac{1}{2}\bar{\mu} \int_0^1 \left(\frac{\partial^2 y}{\partial t\, \partial x}\right) dx, \tag{4}$$

where $\mu = \rho\pi\left(r_2^2 - r_1^2\right)$, $\bar{\mu} = \frac{\rho\pi}{4}\left(r_2^4 - r_1^4\right)$ and $J = \frac{\pi}{4}\left(r_2^4 - r_1^4\right)$
The Lagrange function L:

$$L = E_k - E_p, \tag{5}$$

where E_k, resp. E_p

Euler-Lagrange equations

$$\frac{\mathrm{d}}{\mathrm{dt}}\left(\frac{\partial L}{\partial \dot{q}_i}\right) - \frac{\partial L}{\partial q_i} = 0, \tag{6}$$

where $i = 1, 2, \dots 4$.
Euler–Lagrange equations of the shaft element:

$$\{[M_1] + [M_2]\}[\ddot{q}] - \{[K_1] - [K_2]\}[q] = 0, \tag{7}$$

where

$$[M_1] = \mu \int_0^1 [\Phi]^T [\Phi]\, dx = \frac{\mu l}{420}\begin{bmatrix} 156 & 22l & 54 & -13l \\ 22l & 4l^2 & 13l & -3l^2 \\ 54 & 13l & 156 & -22l \\ -13l & -3l^2 & -22l & 4l^2 \end{bmatrix}, \tag{8}$$

$$[M_2] = \bar{\mu} \int_0^1 [\Phi']^T [\Phi'] \, dx = \frac{\bar{\mu}}{30l} \begin{bmatrix} 36 & 3l & -36 & 3l \\ 3l & 4 & -3l & -l^2 \\ -36 & -3l & 36 & -3l \\ 3l & -l^2 & -3l & 4l^2 \end{bmatrix}, \qquad (9)$$

$$[K_1] = EJ \int_0^1 [\Phi'']^T [\Phi''] \, dx = \frac{EJ}{l^3} \begin{bmatrix} 12 & 6l & -12 & 6l \\ 6l & 4l^2 & -6l & 2l^2 \\ -12 & -6l & 12 & -6l \\ 6l & 2l^2 & -6l & 4l^2 \end{bmatrix}, \qquad (10)$$

$$[K_1] = \mu\omega^2 \int_0^1 [\Phi]^T [\Phi] \, dx = \mu\omega^2 \begin{bmatrix} 156 & 22l & 54 & -13l \\ 22l & 4l^2 & 13l & -3l^2 \\ 54 & 13l & 156 & -22l \\ -13l & -3l^2 & -22l & 4l^2 \end{bmatrix}. \qquad (11)$$

3   Conclusions

Possible generalizations:
    1) Legengre transformation, Hamiltonian and Hamilton equations.
    2) Generalization to connected shafts.

4   Acknowledgement

References

[1] Hrubý Petr, Náhlík Tomáš, and Smetanová Dana. Proposal Mathematical Model for Calculation of Modal and Spectral Properties. In: Post-conference proceedings of extended versions of selected papers of conference Mathematics, Information Technologies and Applied Sciences (MITAV 2017), Czech Republic, 131-140, 2017.

Using Integer Linear Programming to minimize the cost of the thermal refurbishment of a façade: An application to building 1B of the Universitat Politècnica de València, Spain

David Soler[a,*], Andrea Salandin[b], Michele Bevivino[c]

[a] Instituto Universitario de Matemática Multidisciplinar,
Universitat Politècnica de València, Camí de Vera s/n 46022, València, Spain
[b] Centro de Tecnologías Físicas,
Universitat Politècnica de València, Camí de Vera s/n 46022, València, Spain
[c] Architect and buildings energy Consultant, Via Fiume 6, 39100 Bolzano, Italy

1. Introduction

Integer Linear Programming (ILP) is increasingly applied in the field of energy and buildings to solve optimization problems [1-3]. Buildings account 40% of the total EU's energy consumption [4] and are a key potential source of energy savings. Investments in buildings refurbishment deliver energy savings with lower running costs, improved health and comfort with added value. Furthermore, around 54% of the buildings in Spain date back before 1980, when no thermal regulation was available [5]. With the aim of adapting technical standards to the new needs of buildings owners and occupants, in September 2013 the "Basic Document" for Energy Savings of the CTE was published (Documento Básico DB HE «Ahorro de energía» del Código Técnico de la Edificación) [6]. The thermal envelope of a buildings is usually the main element that needs to be refurbished and its thermal transmittance must abide by the current legislation [6] depending on the climate zone. The thermal transmittance U ($Wm^{-2}K^{-1}$) measures the rate of heat flow through the elements of the building envelope [7] and is a key magnitude to assess the energy efficiency. The transparent part represents the weakest element of the envelope from the thermal (and also acoustic) point of view for the greater thermal exchange rate. This paper extends the ILP approach presented in [3] to deal with the problem of minimizing costs for the thermal refurbishment of a façade with thickness and thermal transmittance bounds and with an intervention both on the opaque part (wall) and the transparent part (windows). Among thousands of combinations of materials, thicknesses and type of windows for the thermal refurbishment of a building envelope, the aim is to choose the one that minimizes the cost, without violating any restriction imposed to the refurbishment of the façade. Our case study will be Building 1B of the School for Building Engineering in the Vera Campus of the Universitat Politècnica de València, Spain.

2. Definition of the problem

The problem of minimizing the cost of refurbishment of a façade to comply with current energy efficiency regulations is formulated in this section as an ILP problem. To this aim, we first present some notations, the used variables and parameters, for a better understanding of the formulation.

- Let $S = S_w + S_o$ be the total surface in m^2 of the façade, where S_w is the surface corresponding to the windows *and* S_o is the surface corresponding to the opaque part of the façade. Furthermore, $S_o = S_o^l + S_o^u$, where S_o^l is the surface corresponding to the first lower meter of the opaque part of the façade, and S_o^u is the upper surface to the first lower meter of the opaque part of the façade. Those parts can show different thermal insulation materials in order to avoid rising dump. These surfaces will be also taken into account on preliminary calculations such as computing the price and the transmittance of the different types of windows to cover surface S_w, or to determine the availability or the price of certain materials for the wall.

- Let μ be the number of layers of the original façade, let σ be the sum of the quotients t_i/λ_i where λ_i ($Wm^{-1}K^{-1}$) and t_i (m) represent the thermal conductivity and the thickness respectively of each layer i of the original façade, and let τ be the sum of these thicknesses t_i, $i \in \{1, \dots, \mu\}$.

- Let n be the number of layers to be added to the external side of the original façade, which will be numbered from the inside (layer attached to the wall) to the outside. Each layer $i \in \{1, \dots, n\}$ is made of one of the m_i different materials available for this layer, and given a layer $i \in \{1, \dots, n\}$, the material $j \in \{1, \dots, m_i\}$ is available in r_{j_i} different thicknesses. For each $i \in \{1, \dots, n\}, j \in \{1, \dots, m_i\}, k \in \{1, \dots, r_{j_i}\}$, the following parameters are defined:
  - $t_{i,j,k}$ thickness corresponding to material j with type of thickness k available for layer i.
  - $c_{i,j,k}$ cost of placing in layer i $1m^2$ of material j with type of thickness k.

- Let n_{lo} be the different options for the first lower meter of the thermal insulation (first layer of S_o), for each $j \in \{1, \dots, n_{lo}\}$ there are available r_j different thicknesses of this option. For each $j \in \{1, \dots, n_{lo}\}$ and $k \in \{1, \dots, r_j\}$ the following parameters are defined:
  - $t_{j,k}^{lo}$ thickness corresponding to option j and type of thickness k for the chosen material.
  - $c_{j,k}^{lo}$ cost of placing on the first layer of the lower opaque part $1m^2$ of the chosen material with option j and type of thickness k.

- Let n_f and n_g be the number of different window's frames considered and the number of different combinations of glasses and air chambers considered for the windows respectively, all of them complying with the maximum allowed transmittance for these materials. For each $i \in \{1, \dots, n_f\}$ and $j \in \{1, \dots, n_g\}$ the following parameters are defined:
  - t_i^w thickness corresponding to window's frame of type i.
  - $c_{i,j}^w$ cost of placing $1m^2$ of window with type of frame i and type of glasses combination j. This cost includes the proportional part of removing old windows.

- Given two consecutive layers, there may exist incompatibilities between some materials and thicknesses corresponding to these layers, as described in [3].

- The total thickness of added layers is comprised between bounds t_{min} and t_{max}.

- Let U_{max} be the maximum thermal transmittance allowed for the external wall (its upper opaque part), according to the legislation for the climate zone where the building is located.

- *Let W_{min}* be the minimum recommended windowsill (difference between the final wall thickness and the new window's frame thickness).

- Let $x_{i,j,k}$ be a binary variable which value is 1 if layer i is made with material j and type of thickness k, and 0 otherwise, $i \in \{1, \dots, n\}, j \in \{1, \dots, m_i\}, k \in \{1, \dots, r_{j_i}\}$. Note that for $i=1$, this variable is associated only with the upper opaque part.

- Let $y_{j,k}$ be a binary variable which value is 1 if option j with thickness type k is chosen for the first layer of the lower opaque part, and 0 otherwise, $j \in \{1, \dots, n_{lo}\}, k \in \{1, \dots, r_j\}$.

- Let $z_{i,j}$ be a binary variable which value is 1 if window with type of frame i and type of glasses combination j is chosen for the refurbishment, and 0 otherwise, $i \in \{1, \dots, n_f\}, j \in \{1, \dots, n_g\}$.

- Given a material j, with $j \in \{1, \dots, m_i\}$ for some $i \in \{1, \dots, n\}$, and let λ_j be its thermal conductivity, according to [3] the linear constraint to comply with the thermal transmittance upper bound for the upper opaque part of the façade is:

$$\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{r_{j_i}} \frac{t_{i,j,k}}{\lambda_j} x_{i,j,k} \geq \frac{1}{U_{max}} - \frac{1}{h_{int}} - \frac{1}{h_{ext}} - \sigma \qquad (1)$$

Taking into account all the concepts, restrictions and suppositions given above, the problem of minimizing the refurbishment cost of a façade can be formulated mathematically as the following ILP problem, defined through Eqs. (2) to (13):

$$Minimize \ \ S_o \sum_{i=2}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{r_{j_i}} c_{i,j,k} x_{i,j,k} + S_o^u \sum_{j=1}^{m_1}\sum_{k=1}^{r_{j_1}} c_{1,j,k} x_{1,j,k} + S_o^l \sum_{j=1}^{n_{lo}}\sum_{k=1}^{r_j} c_{j,k}^{lo} y_{j,k} + S_w \sum_{i=1}^{n_f}\sum_{j=1}^{n_g} c_{i,j}^w z_{i,j} \qquad (2)$$

s.t.:

$$\sum_{j=1}^{m_i} \sum_{k=1}^{r_{j_i}} x_{i,j,k} = 1 \quad \forall i \in \{1, \dots, n\} \tag{3}$$

$$\sum_{j=1}^{n_{lo}} \sum_{k=1}^{r_j} y_{j,k} = 1 \tag{4}$$

$$\sum_{i=1}^{n_f} \sum_{j=1}^{n_g} z_{i,j} = 1 \tag{5}$$

$$\sum_{j=1}^{m_1} \sum_{k=1}^{r_{j_1}} t_{1,j,k} x_{1,j,k} = \sum_{j=1}^{n_{lo}} \sum_{k=1}^{r_j} t_{j,k}^{lo} y_{j,k} \tag{6}$$

$$t_{min} \leq \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{k=1}^{r_{j_i}} t_{i,j,k} x_{i,j,k} \leq t_{max} \tag{7}$$

$$\sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{k=1}^{r_{j_i}} t_{i,j,k} x_{i,j,k} + \tau - \sum_{i=1}^{n_f} \sum_{j=1}^{n_g} t_i^w z_{i,j} \geq W_{min} \tag{8}$$

$$\sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{k=1}^{r_{j_i}} \frac{t_{i,j,k}}{\lambda_j} x_{i,j,k} \geq \frac{1}{U_{max}} - \frac{1}{h_{int}} - \frac{1}{h_{ext}} - \sigma \tag{9}$$

$$x_{i,j,k} + x_{(i+1),j',k'} \leq 1 \quad \forall (i,j,k-(i+1),j',k') - incompatible \tag{10}$$

$$x_{i,j,k} \in \{0,1\} \quad \forall i \in \{1, \dots, n\}, j \in \{1, \dots, m_i\}, k \in \{1, \dots, r_{j_i}\} \tag{11}$$

$$y_{j,k} \in \{0,1\} \quad \forall j \in \{1, \dots, n_{lo}\}, k \in \{1, \dots, r_j\} \tag{12}$$

$$z_{i,j} \in \{0,1\} \quad \forall i \in \{1, \dots, n_f\}, j \in \{1, \dots, n_g\} \tag{13}$$

Where:
- Eq. (2) is the objective function, that is, the total cost of the refurbishment.
- Eqs. (3) and (4) ensure that each layer is made exactly of one material with a given thickness. Note that layer 1 has one material for its lower part and another one for its upper part.
- Eq. (5) guarantees that only one type of window is chosen for the whole transparent part.
- Eq. (6) ensures that both the lower part of layer 1 and its upper part have the same thickness.
- Eq. (7) restricts the total thickness of added layers within the established bounds.
- Eq. (8) guaranties that the width of the windowsill is at least W_{min}.
- Eq. (9) is the key restriction with respect to energy efficiency. It ensures that the upper opaque part of the wall does not exceed the maximal allowed thermal transmittance.
- Eq. (10) forbids to place a material j' with thickness k' in the next layer to the one (layer i) containing the material j with thickness k (we denote this fact $(i,j,k-(i+1),j',k')$-incompatibility). At most one of the two materials will appear in the corresponding layer.
- Finally, Eqs. (11) to (13) define the variables of the problem as binary.

Note that the above formulation contains the most usual constrains given in the refurbishment of a façade, but it could include other types of linear constraints to fit as much as possible the real problem.

3. Case study

Our case study will be building 1B of the School for Building Engineering in the Vera Campus of the Universitat Politècnica de València, Spain. Built in the late '60ties as first building ever of 83, its classification is F for energy consumptions, with 350 kW·h·m^{-2}·year^{-1}, and E for CO_2 emissions, with 64 kgCO$_2$·m^{-2}·year^{-1}. The opaque part of the façade is a 3-layer wall: two external layers of concrete (20 mm each) and an interior layer made by simple wood chips mixed with mortar (60 mm). Its transmittance is $U_{opaque}= 1.2 \ Wm^{-2}K^{-1}$, which is over the limit for the correspondent B3 climate zone ($U_{max,opaque}=0.82 \ Wm^{-2}K^{-1}$). The transparent part shows windows with metallic frame and simple glass with a transmittance $U_{frame} = U_{glass} = 5.7 \ Wm^{-2}K^{-1}$ also over the maximum values ($U_{max,windows}=3.6$ (east)

- 3.8 (north) $Wm^{-2}K^{-1}$). Both values are estimations as there are no available project data. We will study two representative façades of the building: the façade with the highest amount of transparent part (54%) on the east side and the largest façade (307 m^2) on the north side, as shown in Figure 1.
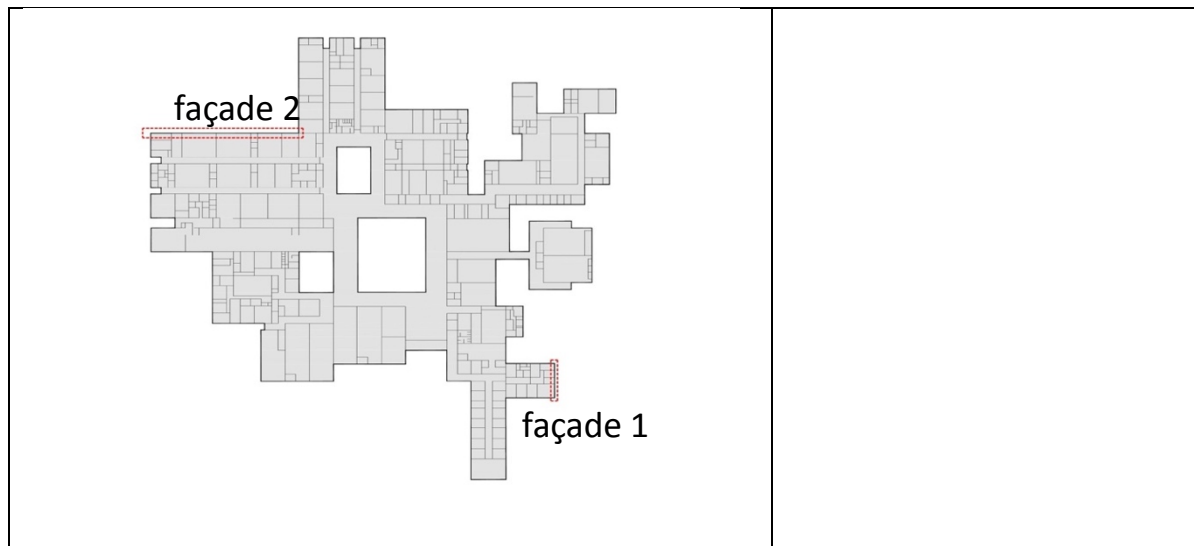
Figure 1. Plan of the ETSEE with position of façade 1 and 2, and detail of the existing façade.

Table 1 summarizes the main data of the two façades, where $U_{transparent} = 0.86 \cdot U_{glass} + 0.14 \cdot U_{frame}$.

Table 1. Data of the chosen façades.

	Façade 1 (EAST)	Façade 2 (NORTH)
Surface [m^2]	*72.80*	*307.00*
Opaque part [m^2]	*33.6*	*267.8*
U_{opaque} [$Wm^{-2}K^{-1}$]	*1.2*	*1.2*
MAX admissible U_{opaque} in zone B3	*0.82*	*0.82*
Transparent part [m^2]	*39.2*	*39.2*
$U_{transparent}$ [$Wm^{-2}K^{-1}$]	*5.7*	*5.7*
MAX admissible $U_{transparent}$ in zone B3	*3.6*	*3.8*
Frame [m^2] (units)	*5.6 (14)*	*5.6 (14)*
U_{frame} [$Wm^{-2}K^{-1}$]	*5.7*	*5.7*
Glass [m^2]	*33.6*	*33.6*
U_{glass}[$Wm^{-2}K^{-1}$]	*5.7*	*5.7*

The suggested refurbishment solution includes the removal of the old windows, the preparation of the "holes" (as a fixed cost), new windows with double glass (standard or low emissive) and 2 different options for the frame (PVC, aluminum) as well as an added multilayer "coat" with flexible configuration (thermal insulation, air chamber, new panel and external finish).

We have selected 50 options for the thermal insulation of the upper part of the façade, 12 options for the thermal insulation of the lower part, 9 options for the air chamber, 7 options for the new panel, 11 options for the external finish, 2 options for the windows's frame and 40 options for the combinations of glasses and internal air chamber of the window. Main data are included in Table 2 and 3.

Table 2. Opaque part.

Layer	Function	Material	Thickness [mm]
Layer 0	Existing façade	Concrete plate with wood chips and mortar	100
		Projected Polyurethane	30 up to 60
Layer 1		Extruded polystyrene	30 up to 60

	Thermal insulation	Mineral wool	30 up to 60
		Expanded polystyrene	30 up to 60
		Expanded cork	30 up to 60
Layer 2	Air cavity	Light ventilated	30, 50, 80, 100
		Not ventilated	30, 50, 80, 100
		Absence	0
Layer 3	New panel	Facing brick	115
		Pressed facing brick	120
		Composite panel	5
		Extruded ceramic panel	16
		Absence	0
Layer 4	External finish	Regular plaster	10 up to 20
		Thermal plaster	10 up to 20
		Absence	0

Table 3. Transparent part.

Element	Material	Composition [mm]
Frame	PVC	82 glass packages - max 52mm thickness
	Aluminum	53 glass packages - max 30mm thickness
Double glass	Standard	External glass: 4, 5, 6, 8
		Air chamber: 6, 8, 10, 12, 14, 16
		Air chamber with argon: 10, 12, 16
		Internal glass: 4, 5, 6, 8
	Low emissivity	External glass: 4, 6, 8
		Air chamber: 6, 8, 10, 12, 14, 16
		Air chamber with argon: 10, 12, 16
		Internal glass: *6,8*

The costs of the different constructive solutions include materials, staff and site facilities, and have been consulted in the cost generator website of CYPE Ingenieros [8] during March 2018.

4. Results and conclusions

After removing all the incompatible combinations (e.g. adding facing brick plus external finish) we have a total of 792,000 possible solutions for the refurbishment of this façade, which shows the complexity of choosing the adequate solution.

To solve both problems (two façades) we have run *Mathematica* 11.0 [9] on a PC Intel® Core™ I7-6700 with 4 processors, 3.46GHz and 8GB RAM. The CPU time to obtain each one of the optimal solutions was little than 0.05s, which is insignificant given the number of possible solutions.

According to the obtained results, three main constructive solutions have been selected as shown in Figure 2. The first constructive solution is a direct application over the existing wall of 30 mm of expanded polystyrene with 10 mm of regular plaster achieves the lowest total thickness with an U-value of 0.597 $Wm^{-2}K^{-1}$. The lowest U= 0.399 $Wm^{-2}K^{-1}$ is obtained with 60 mm of expanded polystyrene. The first meter will be always made by extruded polystyrene for avoiding rising dump.

The second constructive solution increases up to three the number of layers (expanded polystyrene, light ventilated air chamber, extruded ceramic panel) with a lower final U= 0.36 $Wm^{-2}K^{-1}$. The third solution includes a waterproof face brick of 24x11.5x5 mm. With a non ventilated air gap and the inclusion of 30 mm of expanded polystyrene as thermal insultation we reach an U=0.44 $Wm^{-2}K^{-1}$.
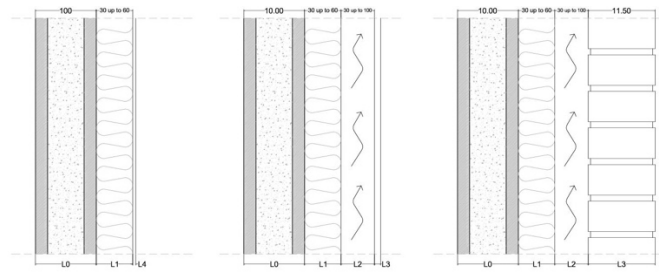
Figure 2. Suggested constructive solutions for thermal refurbishment.

Cost per m^2 are quite different depending on the chosen constructive solutions as well as on the transparency of the façade (54% for façade 1 and 13% for façade 2). For solution 1 we have a cost of 76.86€/m^2 for façade 1 and 34.40€/m^2 for façade 2 with a great difference. Solution 2 shows almost the same price for façade 1 (111.57€/m^2) and façade 2 (104.16€/m^2) due the more expensive material used for layer 3. Lastly costs for solution 3 are intermediate but closer to the ones for solution 2 (95.12 €/m^2 and 84.98€/m^2 respectively for façade 1 and façade 2).

In addition we identified that a small variation in the thickness of the wall can increase considerably the final cost with a small improvement on the thermal transmittance of the wall. The thickness change if we change the constructive solution. For constructive solution 1 we have thickensses between 40 and 80 mm for example. Solution 2 starts with 76 mm and solution 3 with 175 mm. Furthermore we found out that an improved windows with PVC frame and a 4/6/4 glass package is the commoun and cheapest option for all constructive solutions.

Finally the ILP approach seams to be convenient for finding out the best constructive solution under budget, transmittance and thickness constraints taking into account the improvement of the thermal behaviour of a façade that needs to fulfil the new regulation in force and create a better user's comfort.

References

[1] A.S.O. Ogunjuyigbe, T.R. Ayodele, O.E. Oladimeji, Management of loads in residential buildings installed with PV system under intermittent solar irradiation using mixed integer linear programming, Energy Build. 130 (2016) 253-271.

[2] D. Soler, A. Salandin, J.C. Micó, Lowest thermal transmittance of an external wall under budget, material and thickness restrictions: An Integer Linear Programming approach, Energy Build. 158 (2018) 222–233.

[3] A. Salandin, D. Soler, Computing the minimum construction cost of a building's external wall taking into account its energy efficiency, J. Comput. Appl. Math. 338 (2018) 199–211.

[4] Boosting building renovation. What potential and value for Europe? Study for the ITRE Committee. http://www.europarl.europa.eu/RegData/etudes/
STUD/2016/587326/IPOL_STU2016587326_EN.pdf (accessed 22.02.18).

[5] Boletín Especial Censo 2011 Parque edificatorio, Publicaciones del Ministerio de Fomento, http://www.fomento.gob.es/MFOM.CP.Web/handlers/pdfhandler.ashx?idpub=BAW021(accessed 08.03.18).

[6] CTE. Código Técnico de la Edificación (Spanish Technical Building Act). Documento Básico de Ahorro de Energía (Basic Document for Energy Saving). Version of 2013 with comments of 2016. http://www.codigotecnico.org/images/stories/pdf/ahorroEnergia/DccHE.pdf (accessed 20.02.2018).

[7] R. McMullan, Environmental Science in Building, Palgrave Macmillan, Basingtoke, 2012.

[8] Generador de Precios de Elementos de la Construcción, CYPE Ingenieros, S.A., España, 2017 http://www.generadordeprecios.info (accessed 23.02.2018).

[9] Wolfram, Mathematica, http://www.wolfram.com/mathematica (accessed 23.02.2018).

Modeling the Effects of the Immune System on Bone Fracture Healing

Imelda Trejo[♭] [*], Hristo Kojouharov[♭], and Benito Chen-Charpentier[♭]

(♭) Department of Mathematics, The University of Texas at Arlington

P.O. Box 19408, Arlington, TX 76019-0408, USA,

November 30, 2018

1   Introduction

Bone fracture healing is an efficient regenerative process that results in a complete reconstruction of the bone. However, in immune-compromised individuals the healing process can fail or take longer to heal [2]. In addition, surgical complications, disabilities, and high morbidity rates often occurs in individuals with osteoporotic fractures and severe traumas [2]. Therefore, it is important to have a better understanding of the bone fracture healing process and develop new strategies for fracture treatments under a variety of pathological conditions.

Recent experimental results have indicated that the modulation of inflammation via macrophages and mesenchymal stem cells (MSCs) provides new opportunities to optimize bone healing [1, 2]. In [3], a mathematical model based on the interactions among the macrophages, MSCs, and osteoblasts was developed to study the regulatory effects of pro- and anti-inflammatory cytokines during bone fracture healing. It was found that high concentrations of pro-inflammatory cytokines negatively affect the healing time of a fracture and that the administration of anti-inflammatory cytokines can accelerate the healing process in a dose-dependent manner. However, the model assumed that the only source of anti-inflammatory cytokines is given by the

*e-mail: imelda.trejo@mavs.uta.edu

309

MSCs, which may not be enough to promote and correctly represent the complex pattern of bone fracture healing formation. Therefore, it is important to account for the capabilities of macrophages as additional source of anti-inflammatory cytokines during the bone fracture healing process [1, 2].

In this paper, a new mathematical model is developed to better understand of the regulatory effect of cytokines in the bone fracture healing process. It represents an extension of the model in [3], as it separately incorporates the two different phenotypes of macrophages: classically and alternatively activated macrophages. Both phenotypes have the capabilities to modulate the inflammation, however that is achieved though their different phagocytic rates and cytokines productions. Classically activated macrophages release high levels of pro-inflammatory cytokines, including the TNF-α which exhibits inhibitory and destructive properties in high concentrations [2]. In contrast, alternatively activated macrophages are characterized by the secretion of the anti-inflammatory cytokines, such as the IL-10, which increase their phagocytic activities and promote growth of the tissue cells [1, 2].

2   Modeling Assumptions

The most important interactions between macrophages and tissue cells during the bone fracture healing process are observed during the inflammatory and repair phases [3]. In the inflammatory phase, macrophages together with the MSCs modulate and resolve the inflammation, while during the repair phase macrophages provide an optimal environment for the cellular proliferation, differentiation, and tissue production. The primary cells during the inflammatory and repair phases of the bone fracture healing process are debris (D), classically activated macrophages (M_1), alternatively activated macrophages (M_2), MSCs (C_m), and osteoblasts (C_b). Their cellular dynamics are regulated by two generic pro- and anti-inflammatory cytokines: c_1 and c_2, respectively. It is assumed that the regenerative process is given by the production of two extracellular matrices: the fibrocartilage (m_c), and the woven bone (m_b). The variables represent homogeneous quantities in a given volume. Their dynamics are depicted in Figure 1, where the cellular dynamics are represented by the circular shapes and solid arrows. The molecular concentrations and their production/decay are represented by the octagonal shapes and dashed arrows. The pro- and anti-inflammatory cytokines activation/inhibition effects on the cellular functions are represented
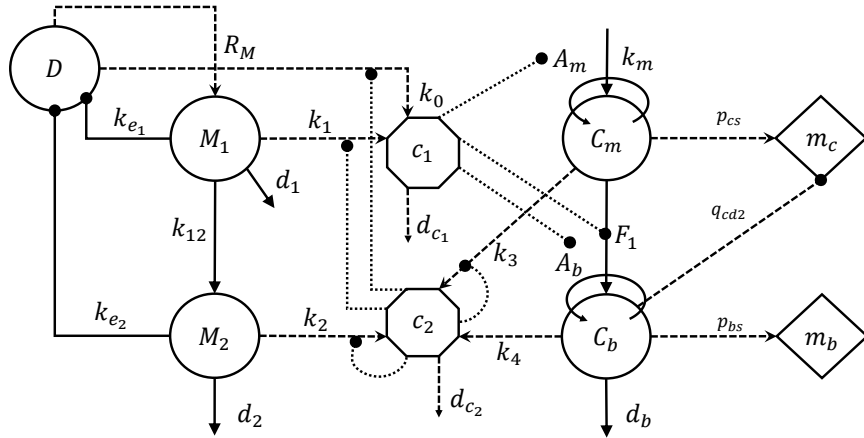
Figure 1: Flow diagram of the cellular and molecular dynamics during the inflammatory and repair phases of the bone fracture healing process.

by the dotted arrows. Removal of debris and the negative effect among the variables are represented by the dot-ending dotted arrows.

For modeling the inflammatory and repair phases of the healing process, the same assumptions are used as in [3]. In addition, it is assumed that the migrating macrophages are classically activated and that they switch to their alternative phenotype at a constant rate. The migration rate of macrophages increases proportionally to the density of debris. The maximal density of the macrophages is denoted by M_{max}. It is assumed that M_1 deliver the c_1 while M_2 and MSCs release the c_2. Both M_1 and M_2 engulf debris at different rates.

3   Model Formulation

The process of bone fracture healing is modeled with a mass-action system of nonlinear ordinary differential equations. Following the flow diagram given in Figure 1 and the above biological assumptions yields the resulting system of equations:

$$\frac{dD}{dt} = -R_D(k_{e_1}M_1 + k_{e_2}M_2) \tag{1}$$

$$\frac{dM_1}{dt} = R_M - k_{12}M_1 - d_1M_1 \tag{2}$$

$$\frac{dM_2}{dt} = k_{12}M_1 - d_2M_2 \tag{3}$$

$$\frac{dc_1}{dt} = H_1(k_0D + k_1M_1) - d_{c_1}c_1 \tag{4}$$

$$\frac{dc_2}{dt} = H_2(k_2M_2 + k_3C_m) - d_{c_2}c_2 \tag{5}$$

$$\frac{dC_m}{dt} = A_mC_m\left(1 - \frac{C_m}{K_{lm}}\right) - F_1C_m \tag{6}$$

$$\frac{dC_b}{dt} = A_bC_b\left(1 - \frac{C_b}{K_{lb}}\right) + F_1C_m - d_bC_b \tag{7}$$

$$\frac{dm_c}{dt} = (p_{cs} - q_{cd1}m_c)C_m - q_{cd2}m_cC_b \tag{8}$$

$$\frac{dm_b}{dt} = (p_{bs} - q_{bd}m_b)C_b. \tag{9}$$

Here, the engulfing rate R_D and migration rate R_M of macrophages are modeled as below:

$$R_D = \frac{D}{a_{ed} + D}, \quad R_M = k_{max}\left(1 - \frac{M_1 + M_2}{M_{max}}\right)D,$$

the inhibitory effects of c_2 are modeled by the following functions:

$$H_1 = \frac{a_{12}}{a_{12} + c_2}, \quad H_2 = \frac{a_{22}}{a_{22} + c_2},$$

and the proliferation and differentiation rates of C_m and C_b are modeled by:

$$A_m = k_{pm} \times \frac{a_{pm}^2 + a_{pm_1}c_1}{a_{pm}^2 + c_1^2}, \quad A_b = k_{pb} \times \frac{a_{pb}}{a_{pb} + c_1}, \quad F_1 = d_m \times \frac{a_{mb_1}}{a_{mb_1} + c_1}.$$

4   Discussion and Conclusion

A complete qualitative analysis of the model was performed. It was determined that there are three biologically meaningful equilibria: two nonunions

and one successful outcome. Their corresponding stability properties are defined in terms of the tissue cells' proliferation and differentiation rates (k_{pm}, k_{pb}, d_m and d_b). From the stability conditions of the successful outcome, the parameter values of the model were selected from [3, 5] and fixed to run different sets of numerical simulations. First, simulations were performed to numerically monitor the evolution of a broken bone for different debris concentration, i.e., different types of fractures. Next, a set of numerical simulations was run to explore possible therapeutic treatments through the administration of anti-inflammatory cytokines, both by increasing the initial concentration of c_2 and also the c_2 production rate by M_2, which is given by the parameter k_2. Figure 2 displays the simulation curves of the evolution of a moderate fracture for different values of the parameter k_2. For
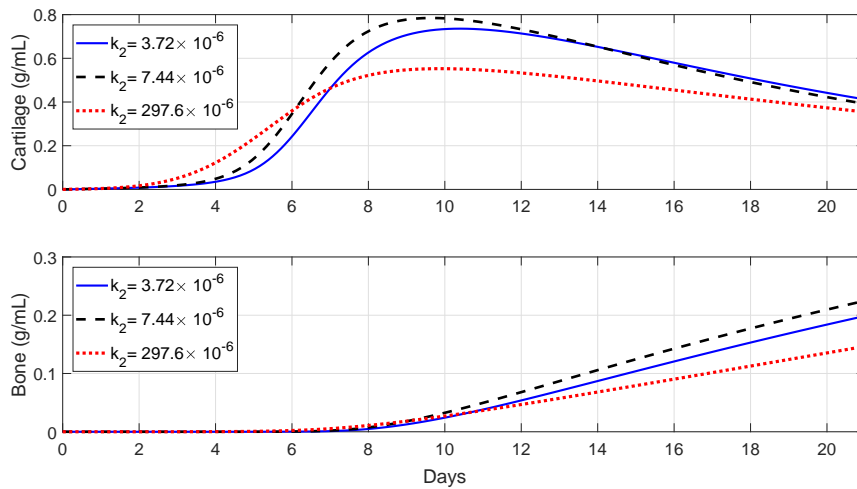
Figure 2: Tissues evolution in a moderate fracture under different anti-inflammatory cytokines treatments.

the base value of $k_2 = 3.72 \times 10^{-6} \ ng/cells/day$ (solid line), the simulation shows that the constant administration of anti-inflammatory cytokines in the moderate fractures improves the tissues evolution but in a dose-dependent manner. On one hand, when k_2 is doubled (dashed line) it enhances the early production of cartilage and increases the bone synthesis, while a drastic increase of k_2 by eighty times (dotted lines) leads to less cartilage and bone formation. Similar results were obtained when $c_1(0)$ was increased to 10 and

100 ng/mL, respectively. Therefore, the presented mathematical model and corresponding numerical simulations suggest that increasing the production of anti-inflammatory cytokines does not always improve the bone fracture healing process, as it has already been observed experimentally and reported in the scientific literature [4].

The flexibility and strength of the new mathematical model allows for a variety of different types of numerical simulations to be performed quickly and cost-effectively to replicate the bone fracture healing process under different conditions. For example, the parameter values of the model can be easily adjusted to simulate the healing of a broken bone in edged and senile osteoporosis individuals. In addition, the model can be used to simulate different medical interventions in the bone fracture healing process, such as MSCs injection and transplantation. The model can also be easily adapted to a variety of other therapeutic approaches and be used to guide clinical experiments and bone tissue engineering strategies.

References

[1] Chung E. and Son Y. Crosstalk between mesenchymal stem cells and macrophages in tissue repair. *Tissue engineering and regenerative medicine*, Volume(11):431-438, 2014.

[2] Loi F., Córdova L.A., Pajarinen J., Lin T.h., Yao Z., and Goodman S.B. Inflammation, fracture and bone repair. *Bone*, Volume(86):119–130, 2016.

[3] Kojouharov H.V., Trejo I., and Chen-Charpentier B.M. Modeling the effects of inflammation in bone fracture healing. *AIP Conference Proceedings*, Volume(1895):020005, 2017.

[4] Mountziaris P.M. and Mikos A.G. Modulation of the inflammatory response for enhanced bone tissue regeneration. *Tissue Engineering Part B: Reviews*, Volume(14):179-186, 2008.

[5] Wang Y., Yang T., Ma Y., Halade G.V., Zhang J., Lindsey M.L. and Jin Y.F. Mathematical modeling and stability analysis of macrophage activation in left ventricular remodeling post-myocardial infarction, *BMC genomics*, Volume(13):S21, 2012.

Metamaterial Acoustics
on the Einstein Cylinder

Michael M. Tung[*]

Instituto de Matemática Multidisciplinar,

Universitat Politècnica de València,

Camino de Vera, s/n, 46022 Valencia, Spain.

November 19, 2018

1 Introduction

Shortly after concluding with the formulation of the general theory of relativity in 1916, Einstein moved on to devise relativistic models of the universe, applying his new theory to the realm of physical cosmology. Assuming uniformity and isotropy for a universe on a very large scale, he produced a simple cosmological model of a finite, static universe with constant spherical curvature, nowadays called the *Einstein cylinder* [1, 2].

Such spaces of constant curvature represent maximally symmetric geometries. This property explains its fundamental importance in many physics and engineering applications, *e.g.* in the description of uncharged, perfect relativistic fluids [3] and other standard cosmological models [4, p. 59]. Moreover, in the past years, quantum mechanical phenomena in spaces of constant curvature have attracted the focus of intense investigation [5], raising critical questions beyond their possible experimental verification. Nonetheless, the simulation of acoustic phenomena [6] in such spaces has so far been vastly neglected.

In this work, we explore the possibilities to simulate acoustics on the Einstein cylinder with the help of acoustic metamaterials—materials which

---

[*]e-mail: mtung@imm.upv.es

enable researchers and engineers to contrive extraordinary devices with exceptional properties, exceeding the limits established by nature. These meta-materials offer researchers and engineers unique opportunities to design and build novel artificial devices with exceptional characteristics, see *e.g.* Refs. [6–9].

Modelling acoustic wave propagation with this particular geometry can be shown to result from a simple variational principle for the acoustic potential, a framework developed in Ref. [10] and extended to various other spacetime geometries [11–14]. This approach yields a wave equation in the form of a partial differential equation for the potential, connected to a harmonic time dependence and to a Sturm-Liouville problem for the radial isotropic coordinates, which can be treated analytically.

The same framework also permits to determine the acoustic parameters corresponding to the postulated spacetime via the so-called constitutive equations [10]. Exactly this fine-tuning implements the acoustic wave propagation for the curved background spacetime under consideration.

Analytical results and numerical estimates conclude this discussion for interesting test cases, which might motivate future laboratory experiments.

2 Spacetime geometry of the Einstein cylinder

Here, we consider the special case of constant positive curvature, $a > 0$, for two dimensions. It is the 2-sphere S^2, which we then embedded into $(2 + 1)$D spacetime with Lorentzian signature. This is the prototype model of a homogeneous and isotropic spacetime. The $(2+1)$-dimensional *Einstein cylinder*, $M = \mathbb{R}_+ \times S^2$, is defined by the following metric [2]

$$g = - (cdt) \otimes (cdt) + (ad\Omega_2) \otimes (ad\Omega_2) \,, \tag{1}$$

where Ω_2 denotes the familiar solid angle for S^2, apart from the time coordinate with the constant speed $c > 0$. The constant $a > 0$ represents a natural scale factor for length.

In isotropic radial coordinates (see Fig. 1 for further explanation) the
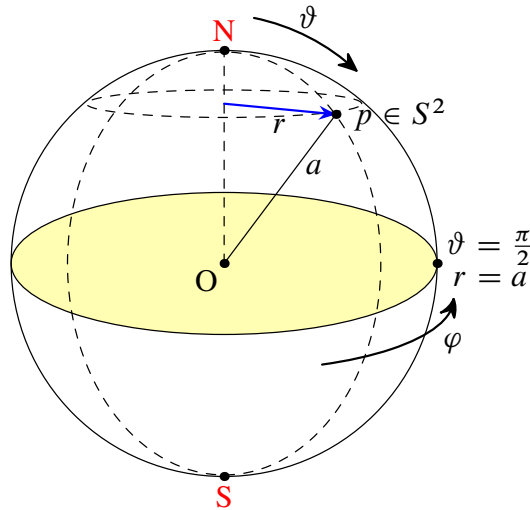
Figure 1: Isotropic radial coordinates are defined by $r = a \sin \vartheta$ and are quite commonly used in the description of spherically symmetric spacetimes. The graphical representation shows that $0 \leq r < a$, or alternatively $0 \leq \vartheta < \pi/2$. As usual for the azimuthal angle holds $0 \leq \varphi < 2\pi$.

original metric of Eq. (1) is recast into the more convenient form

$$\boldsymbol{g} = -(\boldsymbol{cdt}) \otimes (\boldsymbol{cdt}) + \underbrace{\frac{\boldsymbol{dr}}{\sqrt{1 - r^2/a^2}}}_{} \otimes \underbrace{\frac{\boldsymbol{dr}}{\sqrt{1 - r^2/a^2}}}_{\boldsymbol{\theta^1}} + \underbrace{(\boldsymbol{r}\, \boldsymbol{d\varphi}) \otimes (\boldsymbol{r}\, \boldsymbol{d\varphi})}_{\boldsymbol{\theta^2}},$$
$$\underbrace{\phantom{-(\boldsymbol{cdt}) \otimes (\boldsymbol{cdt})}}_{\boldsymbol{\theta^0}}$$

(2)

where we have introduced the local coframe with dual base θ^μ ($\mu = 0, 1, 2$).

As the name implies, this frame $(\theta^0, \theta^1, \theta^2)$ possesses local flatness. Apart from this, orthogonality holds, such that $\eta = -\theta^0 \otimes \theta^0 + \theta^1 \otimes \theta^1 + \theta^2 \otimes \theta^2$, where η is the Minkowski metric. Then Cartan's structure equations will allow us to compute the curvature 2-form Ω. The final result of this calculation yields that for $M = \mathbb{R}_+ \times S^2$ the only non-zero and independent component of the Riemann curvature tensor in the coframe is

$$\hat{R}^1_{\ 212} = \frac{1}{a^2} \quad \Rightarrow \quad G_{00} = \hat{G}_{00} = \frac{1}{a^2}, \tag{3}$$

where the 00-component of the Einstein tensor is the same in the local frame and coordinate frame.

This immediately implies for the energy-matter density ρ_0:

$$\underbrace{G_{00}}_{1/a^2} = \frac{8\pi G}{c^4} \underbrace{T_{00}}_{\rho_0 c^2} \quad \Rightarrow \quad \rho_0 = \frac{c^2}{8\pi G a^2} > 0, \tag{4}$$

where G is the gravitational constant. Thus, a universe filled uniformly with energy-matter will implement the static model of an Einstein cylinder.

3   Acoustic space and metamaterial tuning

The prescription for a constant energy-matter density throughout the Einstein cylinder, *viz.* Eq. (4), in the general theory of relativity has its counterpart in acoustic theory. More precisely, there exists a 1-to-1 correspondence between the metric of a given spacetime metric and the acoustic parameters of the metamaterial which will display analogous properties. This relationship was derived in Ref. [10].

Accordingly, the acoustic engineer who wishes to implement a spacetime $(M, \mathbf{g})$ in the laboratory environment (*physical space*) has to fine-tune the mass-density tensor ϱ and bulk modulus κ in such a manner that it will produce the desired acoustic wave propagation in the corresponding acoustic space (*virtual space*).

In explicit form, both spaces—physical and virtual space—are linked by the *constitutive relations* [10], and in the case for the Einstein cylinder a straightforward calculation shows that the bulk modulus κ and the density tensor ρ have to be

$$\kappa = \frac{1}{\sqrt{1 - r^2/a^2}}, \quad \rho_0 \rho^{ij} = \sqrt{1 - r^2/a^2} \begin{pmatrix} 1 & 0 \\ 0 & 1/r^2 \end{pmatrix}, \tag{5}$$

where $0 < r < a$ and $i, j = r, \varphi$. Recall that $\sqrt{1 - r^2/a^2} = \cos\vartheta$, and the variables are restricted to a domain without coordinate singularities.

4   Wave propagation and its simulation

The fundamental law which governs acoustic wave propagation within a curved spacetime background is dictated by a variational principle, similar to Fermat's principle of least time in theoretical optics.

According to this principle, for a given spacetime M with metric g, the action is stationary with respect to variations of the acoustic potential $\phi :$ $M \to \mathbb{R}$ such that [10]:

$$\frac{\delta}{\delta\phi} \int\limits_{\Omega \subseteq M} d\mathrm{vol}_g \, g(\nabla\phi, \nabla\phi) = 0. \tag{6}$$

Here $\Omega \subseteq M$ is a bounded spacetime domain and $d\mathrm{vol}_g$ its volume element. As a consequence, the corresponding physical propagation law will have its equivalent in equations of motion with self-adjoint differential operators acting on the related field variables [15, p. 351]. This produces separable partial differential equations which are Sturm-Liouville problems for one of the field variables with analytical or at least semi-analytical solutions.

Eq. (6) corresponds to the following simple Euler-Lagrange equation which contains the Laplace-Beltrami operator Δ_M on manifold M:

$$\Delta_M \phi = \frac{1}{\sqrt{-g}} \left(\sqrt{-g} \, g^{\mu\nu} \phi_{,\mu} \right)_{,\nu} = 0 \tag{7}$$

where $g = \det g < 0$. As usual, spacetime indices $\mu, \nu = 0, 1, 2$ preceded by a comma stand for partial derivatives with respect to x^μ and x^ν, respectively.

For acoustic wave propagation on the Einstein cylinder, Eq. (7) converts to

$$-\frac{1}{c^2} \frac{\partial^2 \phi}{\partial t^2} + \left[\left(1 - \frac{r^2}{a^2} \right) \frac{\partial^2}{\partial r^2} + \frac{1}{r} \left(1 - \frac{2r^2}{a^2} \right) \frac{\partial}{\partial r} \right] \phi + \frac{\partial^2 \phi}{\partial \varphi^2} = 0. \tag{8}$$

This is exactly the acoustic wave equation with background metric Eq. (2), engineered and implemented by the acoustic parameters provided in Eqs. (5).

To probe the spacetime properties of the Einstein cylinder, we will choose concentric radial waves leaving the origin:

$$\phi(t, r, \varphi) = A^+ e^{i\omega t} \phi_1(r). \tag{9}$$

The amplitude is $A^+ > 0$ and the frequency is $\omega > 0$. Function $\phi_1(r)$ represents the radial dependency in the full expression of the acoustic potential

$$\phi(t, r, \varphi) = \phi_0(t) \, \phi_1(r) \, \phi_2(\varphi). \tag{10}$$

Applying the separation of variables method as the standard procedure, it is straightforward to recognize that the time dependence in Eq. (10) displays a

simple harmonic behaviour, *i.e.* $\phi_0(t) = e^{i\omega t}$. Because of the radial symmetry of the concentric prototype waves, the angular factor $\phi_2(\varphi)$ in Eq. (10) is just a constant which can be absorbed into the amplitude A^+ in Eq. (9).

So ultimately all of the non-trivial behaviour for the wave propagation will be contained in the radial contribution $\phi_1(r)$. An explicit analysis reveals that the second-order linear differential equation determining $\phi_1(r)$ has regular singular points at $r = 0, a, \infty$. The canonical forms for such differential equations are either Gauss's differential equation or the generalized hypergeometric equation [16, 17]. Then, a lengthy computation yields the following general solution for the remaining radial part of the concentric prototype waves in Eq. (9):

$$
\phi_1(r) = C_1 \; _2F_1 \left(\begin{array}{cc} \frac{1}{4}\left(1 + \sqrt{1 + 4a^2\frac{\omega^2}{c^2}}\right) & \frac{1}{4}\left(1 - \sqrt{1 + 4a^2\frac{\omega^2}{c^2}}\right) \\ & 1 \end{array} \middle| \; \frac{r^2}{a^2} \right)
$$
$$
+ \; C_2 \; G_{2,2}^{2,0} \left(\begin{array}{cc} \frac{1}{4}\left(3 + \sqrt{1 + 4a^2\frac{\omega^2}{c^2}}\right) & \frac{1}{4}\left(3 - \sqrt{1 + 4a^2\frac{\omega^2}{c^2}}\right) \\ 0 & 0 \end{array} \middle| \; \frac{r^2}{a^2} \right),
$$
$$(11)$$

where $_2F_1$ are hypergeometric functions [17] and $G_{2,2}^{2,0}$ Meijer G-functions [18]. These are well defined functions, implemented with high precision on many modern computer systems.

For the numerical wave simulation, we normalize the amplitude $A^+ = 1$ paired with frequency $\omega = 1/2\pi$, and put length scale $a = 100$. In the graphical representations, Fig. 2, we pick several, distinct boundary conditions to illustrate the non-trivial propagation behaviour for the concentric prototype waves travelling on the Einstein cylinder. In cases (a)–(c), we observe a characteristic damping of the wave, which is due to the stretching of θ^1, see Eq. (2), for the isotropic radius approaching the equator at $r = a$, *viz.* Fig. 1. Case (d) shows a wave travelling inwards—from close to the equator to the origin—displaying significant amplification.

5　Conclusions

We presented a detailed analysis of acoustic wave propagation on the Einstein cylinder, motivated by the simplicity of its underlying spacetime geometry, being maximally symmetric and having constant positive curvature.

Using concentric waves for probing characteristic features of this spacetime, a detailed study led us to fully analytic results expressed in terms of hypergeometric functions and Meijer G-functions. We also provide the corresponding acoustic metamaterial parameters for future engineering of such an analogue spacetime in the laboratory setting and for making it subject to challenging experiments.

Acknowledgements
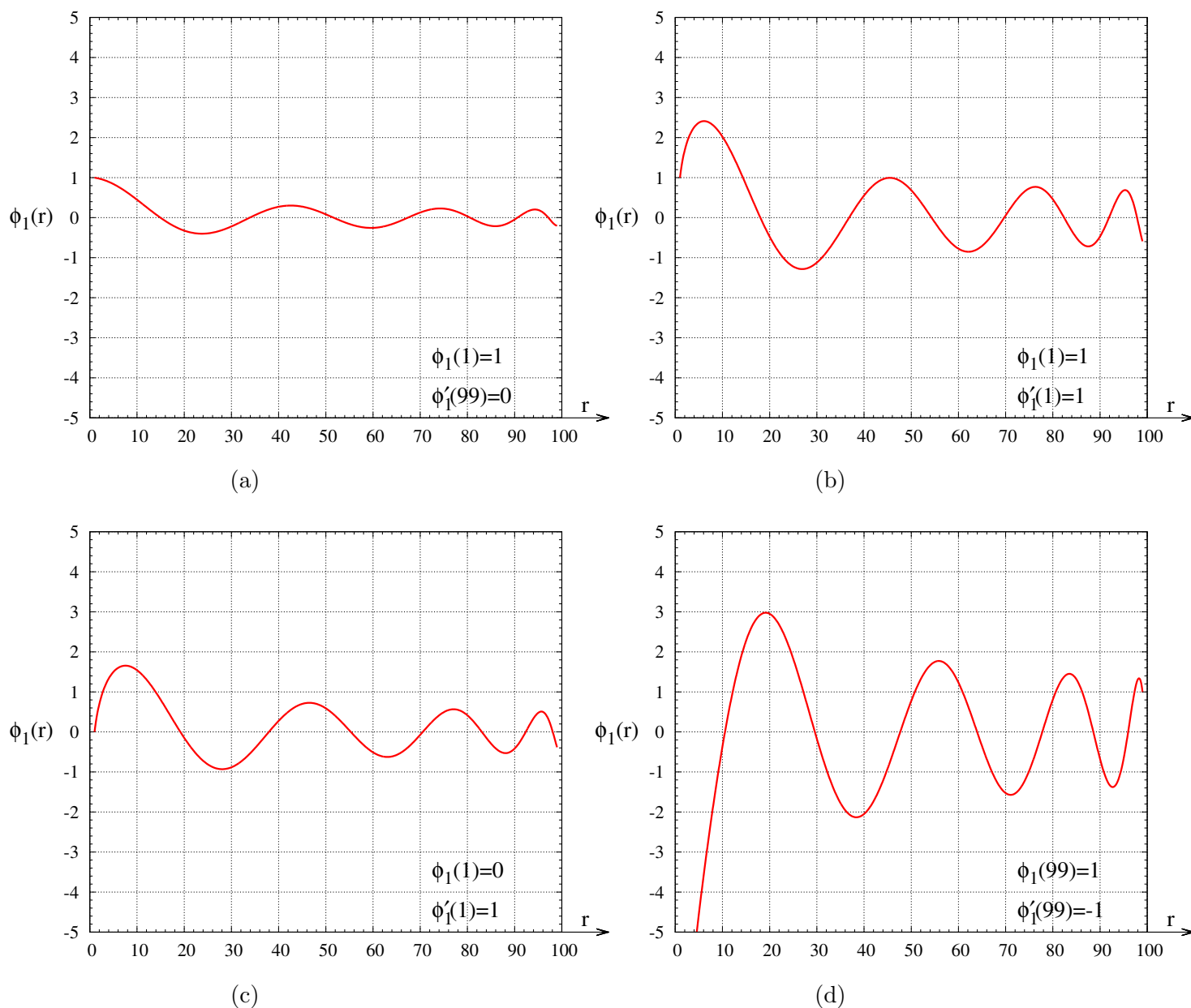
Figure 2: Representation of the non-trivial radial behaviour $\phi_1(r)$ for concentric wave propagation on the Einstein cylinder with length scale $a = 100$. The amplitude A^+ and the speed c are both normalized to unity. The frequency is always $\omega = 1/2\pi$. For illustrative purposes, different boundary conditions are chosen, and they are individually given in the legend of all subfigures. In cases (a)–(c) a characteristic wave damping is observed. Case (d) shows a wave travelling inwards from the equator to the origin.

References

[1] A. Einstein, Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie, *Sitzungsb. König. Preuss. Akad.*, 142–152, 1917.

[2] Y. Choquet-Bruhat, *Introduction to General Relativity, Black Holes, and Cosmology.* Oxford, Oxford University Press, 2015.

[3] B. Kuchowicz, Conformally flat space-time of spherical symmetry in isotropic coordinates, *International Journ. Theor. Phys.*, **7**: 259–262, 1972.

[4] J. N. Islam, *An Introduction to Mathematical Cosmology.* Cambridge. Cambridge University Press, 2001.

[5] V. M. Redkov and E. M. Ovsiyuk, *Quantum Mechanics in Spaces of Constant Curvature.* Hauppauge, NY, Nova Science Publishers Inc., 2012.

[6] S. A. Cummer, Transformation Acoustics. *Acoustic Metamaterials*, R. V. Craster, S. Guenneau (eds.), Springer Series in Materials Science **166**:197–218, Berlin, Springer, 2013.

[7] S. A. Cummer. A sound future for acoustic metamaterials, *J. Acoust. Soc. Am.*, **141**(5):3451, 2017.

[8] M. Haberman and M. Guild. Acoustic metamaterials, *Physics Today*, **69**(6):42–48, 2016.

[9] G. Ma and P. Sheng. Acoustic metamaterials: From local resonances to broad horizons, *Sci. Adv.*, **2**:e1501595, 16 pp., 2016.

[10] M. M. Tung, A fundamental Lagrangian approach to transformation acoustics and spherical spacetime cloaking, *Europhys. Lett.*, **98**:34002–34006, 2012.

[11] M. M. Tung, E. B. Weinmüller, Gravitational frequency shifts in transformation acoustics, *Europhys. Lett.*, **101**:54006–54011, 2013.

[12] M. M. Tung, J. Peinado, A covariant spacetime approach to transformation acoustics. *Progress in Industrial Mathematics at ECMI 2012*, M.

Fontes, M. Günther, N. Marheineke, (eds.), Mathematics in Industry **19**, pp. 335–340, Berlin, Springer, 2014.

[13] M. M. Tung, Modelling acoustics on the Poincaré half-plane, *J. Comput. Appl. Math.*, **337**:336–372, 2018.

[14] M. M. Tung, E. B. Weinmüller, Acoustic metamaterial models on the (2+1)D Schwarzschild plane, *J. Comput. Appl. Math.*, **346**:162–170, 2019.

[15] C. Lanczos, *The Variational Principles of Mechanics.* Mineola, NY, Dover Publications, 1970.

[16] G. E. Andrews, R. Askey, R. Roy, *Special Functions.* Cambridge, Cambridge University Press, 1999.

[17] A. M. Mathai, R. K. Saxena, *Generalized Hypergeometric Functions with Applications in Statistics and Physical Sciences.* Berlin, Springer-Verlag, 1973.

[18] R. Beals, J. Szmigielski, Meijer *G*-Functions: A Gentle Introduction, *Notices Am. Math. Soc.*, **60**(7):866–872, 2013.

Extrapolated Stabilized Explicit Runge–Kutta methods

J. Martín-Vaquero[†], A. Kleefeld [*]

University of Salamanca. Salamanca. E37008 Spain[†]

Forschungszentrum Jülich GmbH, Jülich Supercomputing Centre,
52425 Jülich, Germany[*]

September 20, 2018

1 Introduction

Traditionally classical explicit methods have not been used for stiff ordinary differential equations due to their stability limitations. However, very often, the dimension is high and the eigenvalues of the Jacobian matrix are known to be in a long narrow strip along the negative real axis. This situation typically arises when discretizing spatially parabolic equations or hyperbolic-parabolic equations such as advection-diffusion-reaction equations (with dominating diffusion or reaction). In this case, stabilized explicit Runge-Kutta methods were demonstrated to be very efficient (see [1, 3, 4, 7] and references therein).

Extrapolated Stabilized Explicit Runge-Kutta methods (ESERK) are proposed, in this work, to solve multi-dimensional non-linear partial differential equations (PDEs). For such methods it is necessary to evaluate the function n_t times per step, but the stability region is $O(n_t^2)$. Hence, the computational cost is $O(n_t)$ times lower than for a traditional explicit algorithm. In that way stiff problems can be integrated by the use of simple explicit evaluations in which case implicit methods usually had to be used.

We first calculate the first-order SERK method. Later, we compute the numerical results of the initial value problem (IVP) by performing n_i steps with step size Δt_i to obtain $y_{\Delta t_i}(x_0 + h) := RE_{i,1}$ from $y(x_0)$. We do these calculations with this method for various length-step values $\Delta t_1 > \Delta t_2 > \Delta t_3 > \ldots$ (taking $\Delta t_i = \Delta t/n_i$, n_i being a positive integer). This idea has been considered to develop fourth- (ESERK4, [6]) and fifth-order (ESERK5, [5]) ESERK method.

During the next years, we are working in two different issues:

---

[*]e-mail: jesmarva@usal.es, a.kleefeld@fz-juelich.de

i) We are planning to develop from second- to sixth-order codes based on ESERK methods, and combine all of them in one code. In several papers, including [6], the authors showed how depending on the prescribed tolerance, but also the stiffness of the problem (and also some functions or intervals considered), lower-order methods have better results sometimes, and in others it was necessary to obtain good approximations faster.

ii) Since we need s stages to obtain the first-order stabilized explicit Runge-Kutta approximation ($RE_{1,1}$), the total number of function evaluations of the fourth-order method is $n_t = 10s$, and for the fifth-order scheme is $n_t = 15s$. To increase the speed of these higher-order methods, we are working on parallelized versions of the codes described in [5, 6].

In this work, we briefly explain how we derived sixth-order ESERK (ESERK6) methods and how we are parallelizing all these codes: ESERK4, ESERK5, and ESERK6 algorithms.

2   Stabilized Explicit Runge-Kutta methods

For the construction of these kinds of algorithms two problems need to be solved:

i) Finding stability functions with extended stability domains along the negative real axis;

ii) Finding explicit Runge-Kutta methods with those polynomials as stability functions.

2.1   Stability functions with extended stability domains

The main ingredient for these methods is a Chebyshev polynomial of the first kind:
$$T_s(x) = \cos(s \arccos(x)).$$
If we now consider
$$R_{s,p}(z) = \frac{T_s(w_{0,s,p} + w_{1,s,p}z)}{T_s(w_{0,s,p})}, \quad w_{0,s,p} = 1 + \frac{\mu_p}{s^2}, \quad w_{0,s,p} = \frac{T_s(w_{0,s,p})}{T_s'(w_{0,s,p})}, \quad (1)$$

(s being the number of stages and p the order of the final extrapolated scheme) we obtain polynomials oscillating between $-\lambda_p$ and λ_p in a region which is $O(s^2)$, and $R_{s,p}(z) = 1 + z + O(z^2)$.

The parameter λ_p is always a value smaller than 1, which depends on the order of convergence, p, of the final extrapolated method, see [6]. For example, we calculated $\lambda_3 \leq 0.368008$, $\lambda_4 \leq 0.311688$, $\lambda_5 \leq 0.277923$, and $\lambda_6 \leq 0.25658$ numerically. We took $\mu_4 = 27/16$, $\mu_5 = 192/100$ and $\mu_6 = 208/100$ in Equation (1) to develop fourth-, fifth- and sixth-order schemes, respectively.

As for the new sixth-order schemes, the number of stages of the built Runge-Kutta methods were: $s = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19,$

$20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600,$
$700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400, 2600, 2800, 3000, 3200,$
$3400, 3600, 3800,$ and $4000.$

2.2 Derivation of the explicit Runge-Kutta method with those stability functions

We first construct the first-order methods for $s = qm$ using the following theorem:

Theorem 1 *Let the stabilized explicit Runge-Kutta method be:*

$$g_0 = y_0,$$

$$g_1 = g_0 + \alpha \cdot \Delta t \cdot f(g_0),$$

$$g_j = 2g_{j-1} - g_{j-2} + 2\alpha \cdot \Delta t \cdot f(g_{j-1}) \quad j = 2, \ldots, m,$$

$$g_{m+1} = g_m + \alpha \cdot \Delta t \cdot f(g_m),$$

$$g_j = 2g_{j-1} - g_{j-2} + 2\alpha \cdot \Delta t \cdot f(g_{j-1}) \quad j = m+2, \ldots, 2m, \tag{2}$$

$$\ldots$$

$$g_{(q-1)m+1} = g_{(q-1)m} + \alpha \cdot \Delta t \cdot f(g_{(q-1)m}),$$

$$g_j = 2g_{j-1} - g_{j-2} + 2\alpha \cdot \Delta t \cdot f(g_{j-1}) \quad j = (q-1)m+2, \ldots, s,$$

and $y_1 = \sum_{j=0}^{s} b_j g_j$ where $\alpha = \alpha_4 = 2/s^2$ for the fourth-order method, $\alpha = \alpha_5 = 100/(49s^2)$ for the fifth-order method, $\alpha_6 = 100/(47s^2)$; and b_j are the solutions of the linear system

$$R_{s,p}(z) = b_0 T_0 + \sum_{j=1}^{q} \sum_{i=1}^{m} \left(b_{i+m(j-1)} T_i T_m^{j-1} \right), \tag{3}$$

where $T_i = T_i(1 + \alpha_p z)$ are the shifted Chebyshev polynomials.

This algorithm has $R_{s,p}(z)$ as its stability function and, therefore, it is first-order accurate.

Later extrapolated techniques are employed to increase the order of convergence of the methods, in our case we utilized the Aitken-Neville algorithm with the "harmonic sequence". Thus, for example, the previous fourth- (given as $RE_{4,4} = y(x_{n+1}) + O(h^4)$), and fifth-order ($RE_{5,5} = y(x_{n+1}) + O(\Delta t^5)$) schemes were calculated as:
- Fourth-order methods:

$$RE_{4,4} = \frac{64 y_{\Delta t/4}(x_{n+1}) - 81 y_{\Delta t/3}(x_{n+1}) + 24 y_{\Delta t/2}(x_{n+1}) - y_{\Delta t}(x_{n+1})}{6}.$$

- Fifth-order methods:

$$RE_{5,5} = \frac{625 y_{\Delta t/5}(x_{n+1}) - 1024 y_{\Delta t/4}(x_{n+1}) + 486 y_{\Delta t/3}(x_{n+1}) - 64 y_{\Delta t/2}(x_{n+1}) + y_{\Delta t}(x_{n+1})}{24}.$$

In the case of the new sixth-order schemes the Aitken-Neville algorithm provides us the following formula:

$$RE_{6,6} = 324/5 y_{\Delta t/6}(x_{n+1}) - 3125/24 y_{\Delta t/5}(x_{n+1}) +$$

$$+256/3 y_{\Delta t/4}(x_{n+1}) - 81/4 y_{\Delta t/3}(x_{n+1}) + 4/3 y_{\Delta t/2}(x_{n+1}) - 1/120 y_{\Delta t}(x_{n+1}). \quad (4)$$

Example: Let us consider the case $p = 6$ (order), $s = 4$ (number of stages), $m = 2$, $q = 2$, with $\alpha_6 = 100/(47 s^2)$ for all the sixth-order methods.

(1.) We first calculate $R_4(z)$ taking $\mu_6 = 208/100$ in (1). Precisely, we obtain $w_{0,4,6} = 1.13$, and

$$w_{1,4,6} = \frac{T_4(1.13)}{T_4'(1.13)} = 0.136284,$$

and hence

$$R_4(z) = 1 + z + 0.25853 z^2 + 0.02391 z^3 + 0.00072083 z^4.$$

(2.) Now, we can write $R_4(z)$ as a combination of the modified Chebyshev polynomials for $x = 1 + 100z/(47 s^2)$:

$$R_4(z) = -0.20992\, T_0(x) + 0.03262\, T_1(x) + 0.12802\, T_2(x) +$$

$$+0.47299\, T_3(x) + 0.57629\, T_4(x).$$

(3.) The stabilized explicit Runge-Kutta first-order method (with $R_4(z)$ as stability function) is derived applying equation (2).

(4.) We utilize Richardson extrapolation to obtain the higher-order scheme. Let us suppose that $y_0 \approx y(x_0)$ is the solution previously obtained, and Δt is the length step for the following iteration. Using the latter step, a first-order approximation is obtained, $S_{1,1} \approx y(x_0 + \Delta t)$. When we utilize y_0 and two steps of the first-order SERK scheme given in (3.), but with $\Delta t/2$, then we obtain $RE_{2,1}$, and so on. Finally we employ (4) to obtain the sixth-order approximation $RE_{6,6}$.

2.3 Parallelization of the ESERK schemes

The idea of the parallelization for ESERK schemes is very simple: it is possible to calculate at the same time $RE_{i,1} = y_{\Delta t/i}(x_{n+1})$ separately from $RE_{j,1} = y_{\Delta t/j}(x_{n+1})$ for different values of i, j. At the same time, we know that the computational cost of calculating $RE_{i,1}$ is proportional to the number of function evaluations necessary to calculate it: $i \times s$. Hence, when the final order of the ESERK method is even $p = 4$ or 6, we calculate $RE_{p,1}$ in one processor, $RE_{p-1,1}$ and $RE_{1,1}$ in another one, etc. And finally $RE_{p/2,1}$ in the last one.

If the final order of the ESERK method is odd, as with $p = 5$, we calculate $RE_{p,1}$ in one processor, $RE_{p-1,1}$ and $RE_{1,1}$ in another one, etc. For example, for $p = 5$, we calculate $RE_{5,1}$ in one processor, $RE_{4,1}$ and $RE_{1,1}$ in another processor, and $RE_{3,1}$ and $RE_{2,1}$ in a third one. In this way, the computational cost is proportional to $5s$ and not to $15s$ as in the sequential code.

2.4   Variable-step and number of stages algorithm

The step size estimation and stage number selection are similar to the ones given for other similar schemes, but it is necessary to change the formulae according to the order of the methods derived, see [5, 6] and references therein. First, we select the step size in order to control the local error and then, later we choose the minimum number of stages such that the stability properties are satisfied. The best results (for these extrapolated schemes) were obtained using techniques described in [2] for (traditional) extrapolated methods.

3   Numerical example and conclusions

Let us consider the following two-dimensional non-linear problem from combustion theory, see [8].

$$u_t = d\triangle u + \frac{R}{\alpha\delta}(1 + \alpha - u)\mathrm{e}^{\delta(1-1/u)}, \tag{5}$$

defined on the unit square for $t \geq 0$. The problem is subjected to $u(x, y, 0) = 1$. For $t > 0$ we consider Dirichlet boundary condition $u = 1$ at $x = 1$, $y = 1$, but Neumann boundary condition at $x = 0$, $y = 0$. We used second-order schemes to approximate the Neumann condition. The parameter values in this problem are $d = 2.5$, $\alpha = 1$, $\delta = 20$, and $R = 5$. We used $N = 600$ equispaced nodes in each variable and solved first in the interval $[0, 1.45]$ and later from 1.45 to 1.48.

First, in Table 1, numerical results at $t_{end} = 1.45$ (L_∞ errors) are shown in the upper part of the Table. In this interval the solution is very smooth as it is explained in [8]. We can clearly see how, for tolerances 10^{-6} and 10^{-8}, and similar errors $\sim 10^{-5}, 10^{-6}$, the number of steps given by ESERK5 are smaller than with ESERK4, but the number of function evaluations and CPU times are bigger in this interval. It is more difficult to compare RKC with this table in this interval, however for moderate tolerances is faster than the other two codes.

However, we also solved this problem at $t_{end} = 1.48$, and calculated times, Nfe and number of steps in the interval $[1.45, 1.48]$. We show these differences in Table 1 (bottom part).

We show how the number of steps (with the three methods) is higher in the small interval $[1.45, 1.48]$ than previously in $[0, 1.45]$. It is caused because of the stiff solution in this interval (see [8]), and this motivates the use of variable-step algorithms with very stiff PDEs. However, the number of function evaluations, and CPU times do not grow in the same proportion; obviously, the reason is all these codes vary the number of stages to optimize the CPU times. Finally, in

Interval	Tolerance	Method	max. err.	Time (s)	NFE	Steps
$[0, 1.45]$	10^{-6}	RKC	0.5151_{-2}	393.24	40264	123
		ESERK4	0.2598_{-4}	2368.47	243728	51
		ESERK5	0.2208_{-4}	2562.71	285750	34
$[0, 1.45]$	10^{-8}	RKC	0.2471_{-3}	841.13	87300	545
		ESERK4	0.7085_{-6}	3476.68	358539	122
		ESERK5	0.9726_{-7}	5709.96	418466	75
$[1.45, 1.48]$	10^{-6}	RKC	0.1797_{0}	57.75	9211	262
		ESERK4	0.9094_{-3}	420.30	64712	112
		ESERK5	0.7583_{-3}	707.97	74171	71
$[1.45, 1.48]$	10^{-8}	RKC	0.8426_{-2}	212.01	21040	1259
		ESERK4	0.2288_{-4}	668.13	105674	314
		ESERK5	0.3217_{-5}	745.04	103212	177

Table 1: Maximal absolute error, CPU times, number of function evaluations, steps, and maximal steps for the methods RKC, ESERK4, and ESERK5 for $[0, 1.45]$ (up), and $[1.45, 1.48]$ (bottom).

this interval, for moderate errors $< 10^{-5}, 10^{-6}$ higher-order codes provide faster more accurate solutions. This motivates that we want to combine ESERK codes with different convergence orders into one final code.

Acknowledgements

References

[1] A. Abdulle. Fourth order Chebyshev methods with recurrence relation. *SIAM J. Sci. Comput.*, 23(6):2041–2054, 2001.

[2] E. Hairer, S. P. Nørsett, and G. Wanner. *Solving Ordinary Differential Equations I (2nd Revised. Ed.): Nonstiff Problems.* Springer-Verlag, New York, 1993.

[3] W. H. Hundsdorfer and J. G. Verwer. *Numerical solution of time-dependent advection-diffusion-reaction equations.* Springer, Berlin, 2007.

[4] D. I. Ketcheson and A. J. Ahmadia. Optimal stability polynomials for numerical integration of initial value problems. *Commun. Appl. Math. Comput. Sci.*, 7(2):247–271, 2012.

[5] J. Martín-Vaquero and A. Kleefeld. ESERK5: a fifth-order extrapolated stabilized explicit Runge-Kutta method. *Journal of Computational Physics (submitted)*, 2018.

[6] J. Martín-Vaquero and B. Kleefeld. Extrapolated stabilized explicit Runge–Kutta methods. *Journal of Computational Physics*, 326:141–155, 2016.

[7] B. Sommeijer, L. Shampine, and J. Verwer. RKC: An explicit solver for parabolic PDEs. *Journal of Computational and Applied Mathematics*, 88(2):315–326, 1997.

[8] J. G. Verwer. Explicit Runge-Kutta methods for parabolic partial differential equations. *Appl. Numer. Math.*, 22(1–3):359–379, 1996.

Modelling and simulation of biological pest control in broccoli production

Luz Vianey Vela-Arevalo[♭] [*], Roberto Alejandro Ku-Carrillo[†],
and Sandra Elizabeth Delgadillo-Aleman[†]

(♭) École Polytechnique Fédérale de Lausanne CH-1015 Lausanne, Switzerland

(†) Universidad Autónoma de Aguascalientes, 20131, Aguascalientes, México

November 19, 2018

1   Introduction

Pest control in crops is of utmost importance for agricultural companies. Pesticides have proven to be a reliable option to control pest invasions, but there is a potential hazard for consumers and also to our environment. Biological pest control is widespread because it results in organic and environmentally sustainable products; it involves the introduction of natural predators or parasitoids and the use of biological pesticides.

Mathematical models of biological pest control involving ordinary differential equations have been shown to be successful [1]; for instance, in [2], the impact of parasites to control larva pests using Lotka-Volterra equations is described. Another approach is the use of difference equations [3], where both successes and remaining problems were present, signaling the need for more work in this field. More complex models include also the effect of the temperature [4] or rainfall [5] in the growing rate of pests and all the species involved. Other efforts include stochastic models [6], models with stage-structure [7], etc. Several works in this topic which raise interesting questions from the mathematical and modeling point of view such as in [8].

---

[*]e-mail:luz.velaarevalo@epfl.ch

Our work responded to a call by a local company which grows, freezes and packs vegetables in central Mexico. This work is focused on the production of broccoli, which involves transplanting the baby plants from a greenhouse to an open location where they grow for 90-120 days until harvested. During their growth stage, the plants are attacked by several larvae, among them the Plutella Xylostella (diamond back moth or DBM). The DBM larvae is a major issue because of its quick growing rate and the fact that it kills the plant after feeding from the leaves. The biological control consisted of the release of wasp parasitoids of the genera Diadegma and Trichogramma. The company also periodically spread toxins such as Bacillus Thuringiensis to restrain the outbursts of DBM.

This work proposes simple mathematical models for the pest population dynamics and control. We use the exponential growth model with migration to model the pest population dynamics, and the effect of biological control is thought in two ways: one that abruptly decreases the pest population modelled with infinite impulses; and one where the effect of the control lasts for an interval of time, modelled as a square wave. The main advantage of these models is that it is possible to compute analytic solutions. Also, discrete models for the dynamics are easily obtained and conditions for stability can be found. Furthermore, we were able to use experimental time series provided by the company to show that the model fits adequately the data, showing the potential benefits of mathematical modeling in pest management.

2   Modelling pest dynamics and control

We propose a model for the population dynamics of the pest as an exponential growth model with positive migration, given by

$$\dot{x}(t) = \alpha x(t) + \beta, \qquad x(0) = x_0, \tag{1}$$

where $x(t)$ is the size of the pest population, α is the growth rate, β is the migration rate and x_0 is the initial pest population size.

2.1   Pest control as infinite impulses

We first assume that the control has an instantaneous effect on the pest population. This is modelled with a control term consisting of infinite impulses

at times $t = T_1, T_2, \ldots$, given by

$$\dot{x} = \alpha x + \beta - \sum_{i=1}^{\infty} \gamma\, x \delta(t - T_i), \quad x(0) = x_0, \tag{2}$$

where $\delta(t)$ is the Dirac delta function and γ is the effectiveness of the control. The analytical solution $x(t)$ of equation (2) is obtained by Laplace transform methods and involves $H(t)$, the step or Heaviside function. The values of the solution $x(T_i)$ at the times of the control applications are calculated to satisfy continuity, and a recursive formula for them can be obtained. A graph of a solution of (2) is shown in Figure 1(a).

For the case of regularly spaced applications $T_n = nT$, the solution of (2) can be expressed explicitly and we obtained a sequence $x(nT)$ of the points where the population hits a minimum as an immediate response to the control application. It can be shown that this sequence converges when $e^{\alpha T}/(1 + \gamma) < 1$ and diverges otherwise. The local maxima correspond to the limit $x(nT^-)$ as $x \to nT$ with $t < nT$, and are given by $x(nT^-) = (1 + \gamma)x(nT)$. This provides a simple criteria to optimize the response.

2.2 Pest control with finite effect rate

We assume that the application of the biological control is periodic with period T, and that its effect is active for a duration τ, with $\tau < T$. The application occurs at times $t = nT - \tau$ and the end of its effect is at $t = nT$. The model proposed is

$$\begin{aligned} \dot{x} &= \alpha x + \beta - f(t)x, \\ f(t) &= \begin{cases} 0, & nT \le t \le (n+1)T - \tau, \\ \gamma, & (n+1)T - \tau < t < (n+1)T, \end{cases} \quad n = 0, 1, 2, \ldots \end{aligned} \tag{3}$$

We calculated the analytical solution of model (3), an example of this solution is presented in Figure 1(b).

Let us remark, the times at which the pest population changes abruptly are when: a) the pesticide is applied ($t = nT - \tau$), that corresponds to a local maximum; and b) the pesticide effect ends ($t = nT$), that corresponds to a local minimum. From the analytic solution, it is possible to conclude that any solution converges to a periodic solution when $\gamma > (T/\tau)\alpha$), and diverges otherwise.
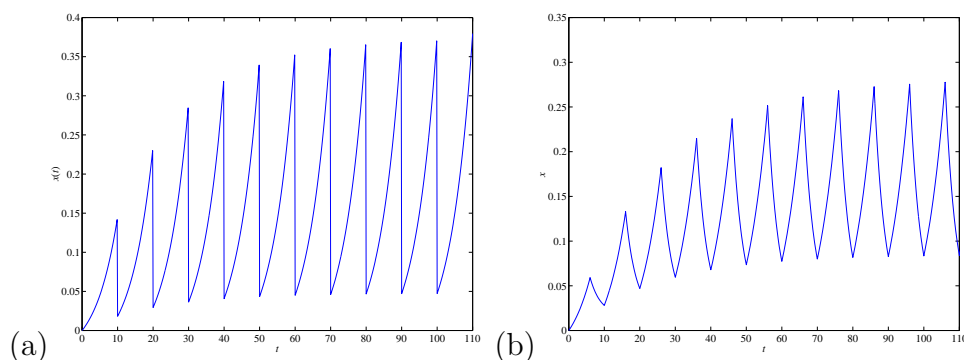
Figure 1: (a) A solution for model (2). (b) A solution for model (3).

3 Broccoli production with biological control

Our aim is to illustrate the application of the studied models in the real world, but we are aware that this is first approximation to a very complex problem.

3.1 Experimental data and parameter identification

We were provided with experimental data by a company that produces, freezes and packs broccoli and uses biological pest control methods. Our first step was to approximate the parameters that correspond to the models studied in this work.

The parameter estimation was obtained minimizing a least-square function error $E(p) = \sum_k^N (x_k^* - x_k(p, x_0))^2$, where x_k^* is the observed DBM larva population per plant at time t_k. $x_k(p, x_0)$ is the model solution for (3) and $p = (\alpha, \beta, \gamma, T_i)$, where T_i, $i = 1, 2, 3, ..., T_N$ are the application times of the control. It is important to mention that model (3) considers equally space time intervals cycles of length T, but we adapted the model for different lengths of time cycles.

3.2 Simulations

The results of the simulations based on the estimated parameters are presented in Figure 2 for two land plots. We can observe that the solutions fit well the experimental data in the beginning of the simulation. Generally, for
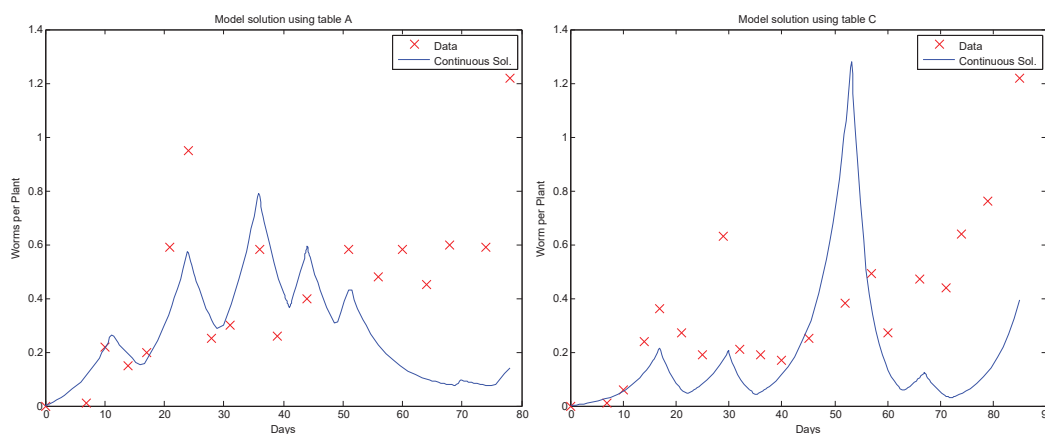
Figure 2: Comparison of simulations for the model (3) with the experimental data (X) of the population of DBM in broccoli production for two land plots, using parameter appoximations.

the longer times (closer to 90 days), the model solution does not approximate the experimental data appropriately. This could be a consequence of a number of factors that were not consider in our model such as the age-structure of all involved populations, the delays or the stochasticity of different parameters such as the rain, temperature, etc. However, our results show that a simple model can be useful to qualitative and quantitative describe the DBM population dynamics.

4   Conclusions

We have presented and studied models for the population dynamics of a pest under biological controls. Particularly, two basic models were analyzed: one with infinite impulses and another one with square wave function to model biological control applications. We provided analytical solutions in each case and described criteria for the convergence of solutions. These results are of theoretical and practical interest. On one side they describe analytically the solution of the model and the asymptotic behavior of the solutions. On the other side, we have shown that these models can be applied to real world problems. Let us say, that our models are based on the effectiveness of the control and can be used to simulate different scenarios. An important part of the study of pest control is the estimation of the parameters, which we did

based on experimental data. The parameters were used in the simulations, obtaining promising results.

References

[1] S Tang, Y Xiao, et al. Integrated pest management models and their dynamical behaviour. *Bulletin of Mathematical Biology*, 67(1):115–135, 2005.

[2] HEZ Tonnang, LV Nedorezov, et al. Assessing the impact of biological control of plutella xylostella through the application of Lotka–Volterra model. *Ecological Modelling*, 220(1):60–70, 2009.

[3] J Liang, S Tang, et al. Beverton–holt discrete pest management models with pulsed chemical control and evolution of pesticide resistance. *Communications in Nonlinear Science and Numerical Simulation*, 36:327–341, 2016.

[4] CA Marchioro and LA Foerster. Development and survival of the diamondback moth, plutella xylostella (l.)(lepidoptera: Yponomeutidae) as a function of temperature: effect on the number of generations in tropical and subtropical regions. *Neotropical entomology*, 40(5):533–541, 2011.

[5] Y Kobori and H Amano. Effect of rainfall on a population of the diamondback moth, plutella xylostella (lepidoptera: Plutellidae). *Applied entomology and zoology*, 38(2):249–253, 2003.

[6] C Zhu and G Yin. On competitive Lotka–Volterra model in random environments. *Journal of Mathematical Analysis and Applications*, 357(1):154–170, 2009.

[7] R Shi and L Chen. Staged-structured Lotka–Volterra predator–prey models for pest management. *Applied Mathematics and Computation*, 203(1):258–265, 2008.

[8] S Nundloll, L Mailleret, et al. Two models of interfering predators in impulsive biological control. *Journal of Biological Dynamics*, 4(1):102–114, 2010.

Preliminary study of fuel assembly vibrations in a nuclear reactor

A. Vidal-Ferràndiz[♭], D. Ginestar[‡][*] A. Carreño[♭], G. Verdú[♭],

(♭) Instituto de Seguridad Industrial: Radiofísica y Medioambiental,

Universitat Politècnica de València, Camino de Vera, s/n, 46022, València,

(‡) Instituto Universitario de Matemática Multidisciplinar,

Universitat Politècnica de València, Camino de Vera, s/n, 46022, València.

November 30, 2018

1   Introduction

Being able to monitor the state of nuclear reactors while they are running at nominal conditions is a safety requirement. The early detection of anomalies gives the possibility to take proper actions before such problems lead to safety concerns or impact plant availability. The CORTEX project [1], funded by the European Commission in the Euratom 2016-2017 work program, aims at developing an innovative core monitoring technique that allows detecting anomalies in nuclear reactors, such as excessive vibrations of core internals, flow blockage, coolant inlet perturbations, etc. The technique is based on primarily using the inherent fluctuations in neutron flux recorded by in-core and ex-core instrumentation, often referred to as neutron noise, from which the anomalies will be detected.

In this work, we aim to simulate the neutron field behaviour of nuclear reactor when one fuel assembly is vibrating. These vibrations cause neutron flux and power oscillations, also known as neutron noise [3]. Similar studies have been performed in the time domain [5] and in the frequency domain [6].

[*]e-mail:dginesta@mat.upv.es

To study the neutron flux fluctuations due to bundle vibrations it is necessary to simulate them accurately. The simulation is performed with a continuous Galerkin finite element method and a semi-implicit numerical time integration [4]. The small amplitude of the fuel assembly vibrations makes an accurate simulation of the neutron power evolution a challenging problem. One dimensional numerical examples are studied in this work.

2   Results

In order to test the numerical tools developed a simple one dimensional benchmark is defined. The benchmark is composed of 11 assemblies of 25 cm where the vibrating assembly is placed in the middle of the reactor as Figure 1 shows. The cross sections are defined in Table 1 and zero flux boundary conditions are imposed. The problem is made critical before starting the time dependent calculation.

The oscillation of the central assembly is defined as

$$x_i(t) = x_{i0} + A\sin(2\pi ft), \tag{1}$$

where $x_i(t)$ is each position of the vibrating assembly along time, originally placed in x_{i0}. A is the oscillation amplitude and f is the oscillation frequency.

Figure 2 shows the total power evolution for an oscillation of 1 mm of amplitude and a frequency of 1 Hz along 10 periods. It can be seen a sinusoidal change in the total power with a really small amplitude, about 7.87e-8, with a constant increment along time. This increment is caused because the reactor is supercritical when the central assembly moves from its starting position. Figure 3 displays the static k_{eff} through the positions travelled during one period. It can be seen that the change in the k_{eff} is less that 1.2e-9. The behaviour of the total power was solved analytically in a point kinetic reactor in [2].

In these Figures, 2 non-equidistant meshes are compared. One mesh with 47 cells and a second mesh with the double of cells, 94. Also a uniform mesh with 17600 cell is compared. All computations are calculated with 5th degree polynomials in the finite element method. These meshes display almost equal results. Then, the results with the local 47 cells mesh are converged.
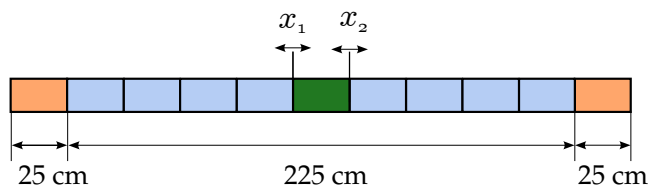
Figure 1: Geometry of the one dimensional benchmark.

Table 1: Cross sections of the materials of the one dimensional benchmark.

Material	g	$\mathbf{D_g}$ (cm)	$\mathbf{\Sigma_{ag}}$ (1/cm)	$\mathbf{\nu\Sigma_{fg}}$ (1/cm)	$\mathbf{\Sigma_{fg}}$ (1/cm)	$\mathbf{\Sigma_{12}}$ (1/cm)
Fuel	1	1.40343	1.17659e–2	5.62285e–3	2.20503e–3	1.60795e–2
	2	0.32886	1.07186e–1	1.45865e–1	5.90546e–2	
Vibrating	1	1.40343	1.17659e–2	5.60285e–3	2.19720e–3	1.60795e–2
Assembly	2	0.32886	1.07186e–1	1.45403e–1	5.88676e–2	
Reflector	1	0.93344	2.81676e–3	0.00000e+0	0.00000e+0	1.08805e–2
	2	0.95793	8.87200e–2	0.00000e+0	0.00000e+0	

Figure 4 shows the neutron power evolution for different oscillation amplitudes from 0.3 mm to 3 mm while the frequency is fixed to 1 Hz. Obviously as the oscillation amplitude increases its effect in the total power increases. Figure 5 displays the spatial resolution of the neutron flux in 4 different time stamps.
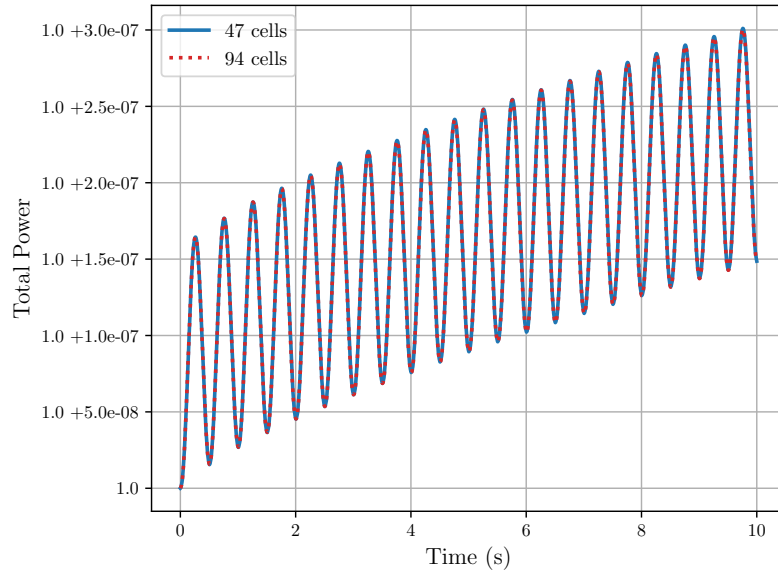
Figure 2: Total neutron power along 10 periods.
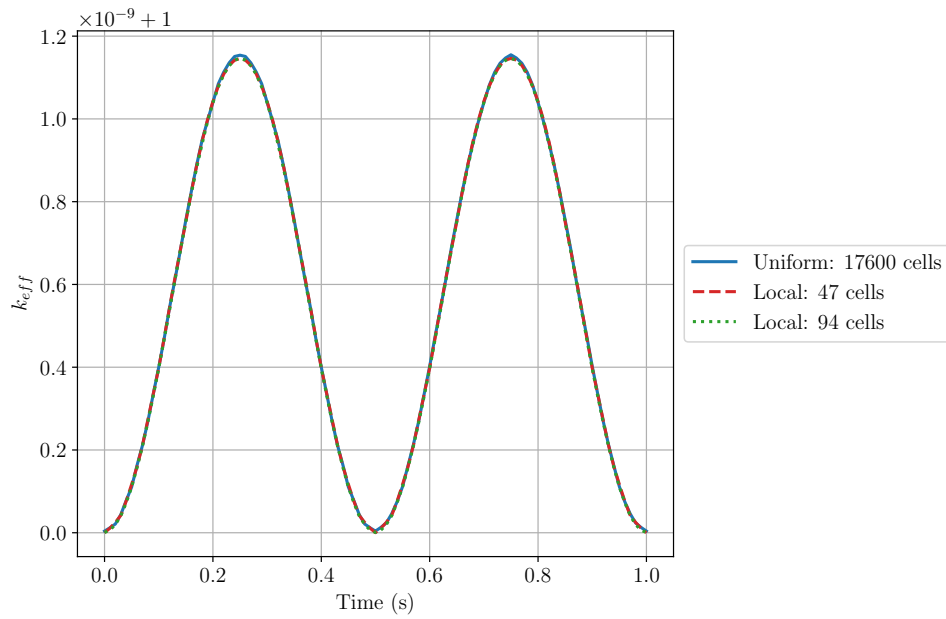
Figure 3: Multiplicative factor along one period.
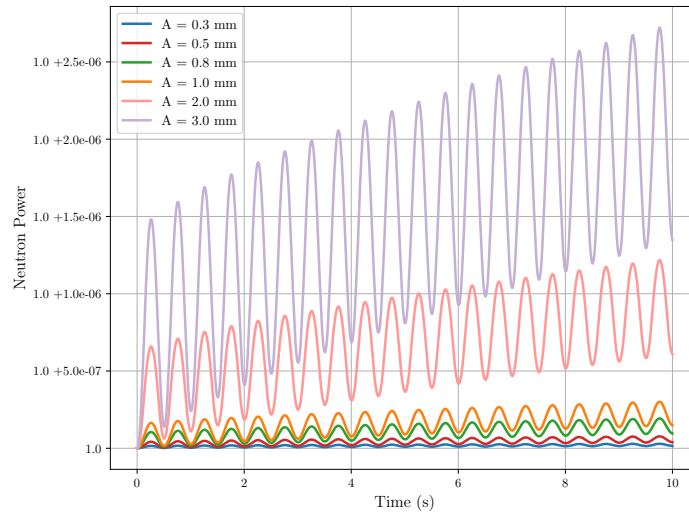
Figure 4: Total neutron power evolution for different oscillation amplitudes.
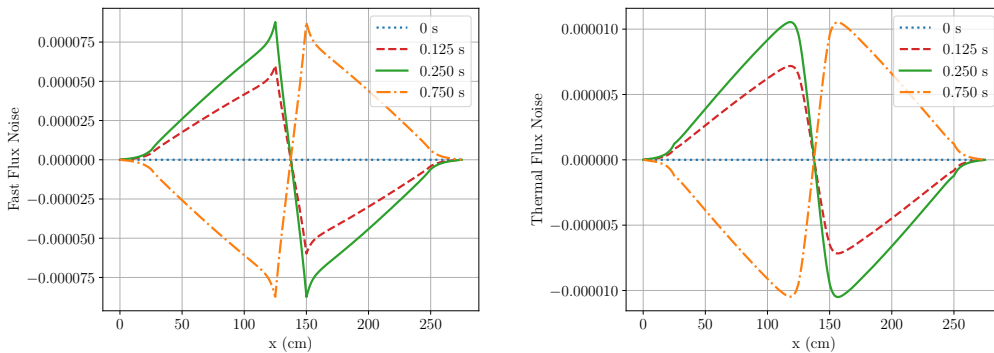
(a) Fast neutron flux noise

(b) Thermal neutron flux noise

Figure 5: Evolution of the neutron noise.

3   Conclusions

A time-domain FEM kinetic code is being developed to solve the neutron distribution inside a nuclear reactor with vibrating assemblies. The results show that the variation in the k_{eff} is about 10^{-8} in the transient and the variation in the total power is around 10^{-9} for an oscillation with an amplitude of 1 mm. This implies that we need to work with a very high precision. These initial results show that fuel assembly vibration cannot cause large noise instabilities in normal conditions without a coupling with thermal-hydraulics system or several fuel assembly vibrations.

Acknowledgements

References

[1] C. Demazière, P. Vinai, M. Hursin, S. Kollias, and J. Herb. Noise-Based Core Monitoring and Diagnostics - Overview of the project. In *Advances in Reactor Physics (ARP-2017)*, pages 1–4, Mumbai, India, 2017.

[2] P. Ravetto. Reactivity oscillations in a point reactor. *Ann. Nucl. Energy*, 24(4):303–314, 1997.

[3] U. Rohde, M. Seidl, S. Kliem, and Y. Bilodid. Neutron noise observations in german KWU built PWRs and analyses with the reactor dynamics code DYN3d. *Ann. Nucl. Energy*, 112:715–734, 2018.

[4] A. Vidal-Ferràndiz, R. Fayez, D. Ginestar, and G. Verdú. Moving meshes to solve the time-dependent neutron diffusion equation in hexagonal geometry. *J. Comput. Appl. Math.*, 291:197–208, 2016.

[5] M. Viebach, N. Bernt, C. Lange, D. Hennig, and A. Hurtado. On the influence of dynamical fuel assembly deflections on the neutron noise level. *Prog. Nucl. Energy*, 104:32–46, 2018.

[6] T. Yamamoto. Implementation of a frequency-domain neutron noise analysis method in a production-level continuous energy monte carlo code: Verification and application in a BWR. *Ann. Nucl. Energy*, 115:494–501, 2018.

Evolution and prediction with uncertainty of the bladder cancer of a patient using a dynamic model

C. Burgos[♭], N. García-Medina[♭],
D. Martínez-Rodríguez[♭] *and R.-J. Villanueva[♭]

(♭) Instituto de Matemática Multidisciplinar,

Universitat Politècnica de València, camino de Vera s/n, Valencia, Spain.

November 30, 2018

1 Introduction

Bladder cancer has become one of the most common and dangerous neoplasms in the urinary system [1], [2]. In a high number of cases, a recurrence is expected. However, there is not information about the mechanism that makes this recurrence happens. The aim of this work is to provide a useful tool to urologists and anatomical pathologists which is able to predict the dates of the recurrences in order to improve the success of the treatment applied. For this reason, the evolution of the disease and the treatment is studied, a mathematical model is proposed and then the mathematical model is calibrated in order to predict the evolution of a specific patient.

2 Available data

The procedure followed during the treatment of the disease is the following. First, the patient goes to the family doctor because a hematuria (blood in

---

*e-mail: damarro3@upv.es

the urine). The doctor address the patient to the urologist, who diagnoses a tumour in the bladder. A trans-urethral resection (TUR) is made and the biological sample is sent to the anatomical pathologist, who determines if the tumour is malignant. If it is, intra intravesical instillations of bacillus Calmette-Guerin (BCG) are administrated in order to stimulate the immune system. If there is recurrence, this procedure is repeated until the patient is cured or the cancer evolves into a higher stage.

   With the aim to know what information is available during the medical daily practice, the archive of "Hospital Universitario y Politécnico La Fe" [3] in Valencia was studied, obtaining the data of a patient called Patient X.

Date	Activity	Size (mm)	Inflammatory cells/Field
01/03/2012	Ultrasound	3-5	—
14/06/2012	TUR	25	260
15/02/2015	Cystoscopy	1-2	—
28/04/2015	TUR	5	515
30/01/2017	Cystoscopy	20	—
14/03/2017	TUR	30-35	508

Table 1: Evolution of patient X. Data obtained from the archive of "Hospital Universitario y Politécnico La Fe" [3].

3   Model building

As far as our research has found in literature, only the model developed in [4] is able to explain the interactions of bladder cancer elements including BCG instillations. With the purpose of using the data available during the daily medical practice shown in Table 1 we have made some modifications in order to achieve this objective. The modified model that shows the interaction between all the macroscopic elements that form bladder cancer in shown in Equations 1, 2, 3 and 4.

$$E(t + 1) = E(t) - \mu_2 E(t) + \alpha T_u(t) + p_4 E(t)B(t) - p_5 E(t)T_u(t) \quad (1)$$

$$T_u(t + 1) = T_u(t) - p_2 B(t)T_u(t) - p_6 E(t)T_u(t) + k(T_u(t)) \quad (2)$$

$$T_i(t + 1) = T_i(t + 1) - p_3 E(t) T_i(t) + p_2 B(t) T_u(t) \qquad (3)$$

$$B(t + 1) = -\mu_1 B(t) - p_1 E(t) B(t) - p_2 B(t) T_u(t) + b(t) \qquad (4)$$

The model is a system of difference equations with four different elements:

- $E(t)$: Inflammatory cells in the tumour microenvironment. It is measured as the mean of five inflammatory cells microenvironment counting made with a 40x microscope magnifications.

- $T_u(t)$: Size of uninfected with BCG tumour cells measured as mm of diameter.

- $T_i(t)$: Size of infected with BCG tumour cells measured as mm of diameter.

- $B(t)$: Amount of BCG cells measured as ml of instillations.

The different parameters are the interactions between the elements of the system. For further information, see [4].

4   Model calibration

Once the model has been proposed in order to be used with data of Table 1, we are using an optimization algorithm in order to know if the model is able to represent the reality of our patient.

Objective function F:

- Substitute the values of the model parameters.

- Run the model and obtain the outputs in the same time instants as in Table 1

- Obtaining the root mean square difference between the output of the model and the values of Table 1.

To optimize the value of the output of the objective function F, Random Particle Swarm Optimization algorithm is used [5]. The result of the model calibration is shown in Figure 1.
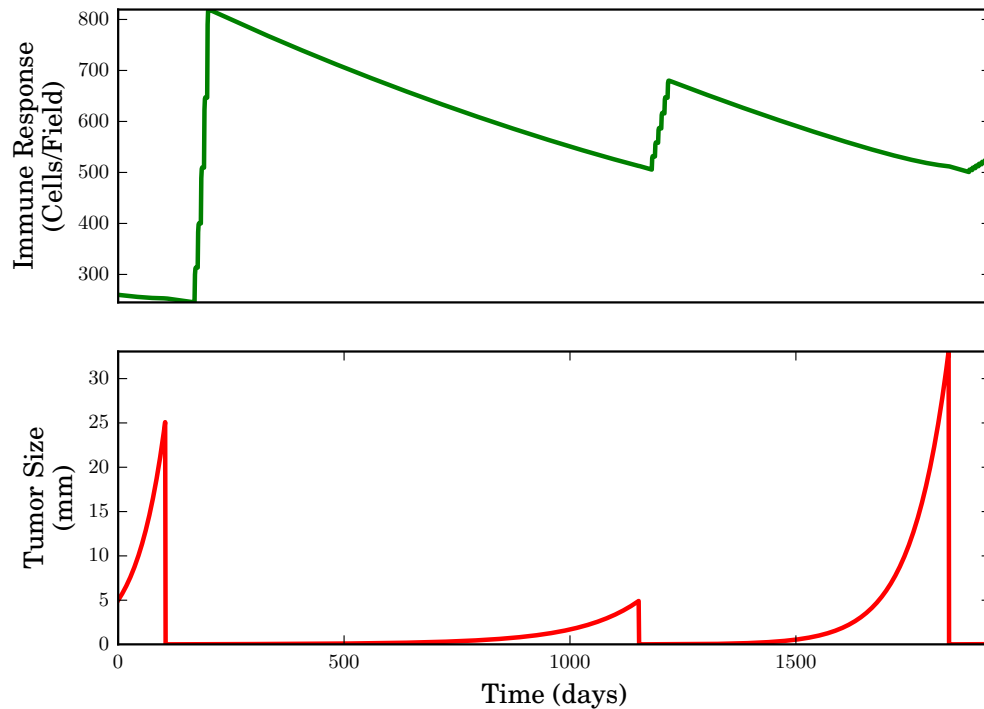
Figure 1: Evolution of the bladder cancer tumour. The first day is settle the 01/03/2012.

References

[1] Official Site for Spanish Medic Oncology Society, *https://www.seom.org*, accessed: 08/05/2018

[2] R. T. Greenlee, M. B. Hill-Harmon, T. Murray and M. Thun. Cancer Statistics, 2001 *A Cancer Journal for Clinicians*, 51:15–36, 2001.

[3] Official Site for Hospital La Fe, *http://www.hospital-lafe.com*, accessed 07/05/2018

[4] S. Bunimovich-Mendrazitsky, E. Shochat and L. Stone Mathematical Model of BCG Immunotherapy in Superficial Bladder Cancer, *Bulletin of Mathematical Biology*, 69:1847–1870, 2007.

[5] C. Jacob and N Khemka. Particle Swarm Optimization in *Mathematica. An exploration kit for evolutionary optimization, IMS'04, Proc. Sixth International Mathematica Symposium*, 2004.

Dynamics of a family of Ermakov-Kalitlin type methods*

Alicia Cordero[♭] [†] Juan R. Torregrosa[♭], and Pura Vindel[‡]

(♭) Instituto de Matemática Multidisciplinar,
Universitat Politècnica de València, Spain.
(‡) Instituto de Matemáticas y Aplicaciones de Castellón,
Universitat Jaume I, Spain.

November 30, 2018

1 Introduction

The application of iterative methods for solving nonlinear equations $f(z) = 0$, with $f : \mathbb{C} \to \mathbb{C}$, gives rise to rational functions whose dynamical properties are not well-known. The simplest model is obtained when $f(z)$ is a quadratic polynomial and the iterative process is Newton's method. The study on the dynamics of Newton's scheme has been extended to other point-to-point iterative methods used for solving nonlinear equations, with convergence order up to three (see, for example [1], [2], [3], among others).

In [4], one third-order family of Ermakov-Kalitkin type was described. It showed to be convergent, as well in case of equations as in the multidimensional case, when Newton's method even did not converge. The iterative expression corresponding to

the class described in [4], main aim of this work, is:

$$y_k = x_k - \alpha \frac{f(x_k)}{f'(x_k)},$$

$$x_{k+1} = x_k - \frac{f(x_k)^2}{bf(x_k)^2 + cf(y_k)^2} \frac{f(x_k)}{f'(x_k)}$$

where $b = \frac{1-\alpha+2\alpha^2}{2\alpha^2}$ and $c = \frac{1}{2\alpha^2(\alpha-1)}$. This family is denoted by PM class.

On the other hand, it is known (see, for example, [5, 6]) that the roots of a polynomial can be transformed by an Möbius map $h(z)$ with no qualitative changes on the dynamics of the family of polynomials, where

$$h(z) = \frac{z-a}{z-b}.$$

By applying this conjugacy map, the operator of PM class is conjugated to the rational function

$$LG(z,\alpha) = \frac{z^3\left(\alpha^2 + 2(1+z)^2(\alpha-1)\right)}{\alpha^2 z^2 + 2(1+z)^2(\alpha-1)}. \tag{1}$$

2 Fixed and critical points

We study now the dynamics of operator $LG(z,\alpha)$ as a function of parameter α. Firstly, we calculate the fixed points of $LG(z,\alpha)$ and then, its critical points. As we will see, the number and the stability of the fixed and critical points depends on the parameter α.

The fixed points satisfy

$$LG(z,\alpha) = z.$$

The roots of this equation are $z \in \{0, \infty, 1, -1\}$. To study the stability of these fixed points, we need the derivative of the operator $LG(z,\alpha)$,

$$LG'(z,\alpha) = z^2 \frac{\alpha^4 z^2 + 12(1+z)^4(1-2\alpha) + 2\alpha^2(1+z)^2\left(\alpha\left(3 - 2z + 3z^2\right) + \left(3 + 14z + 3z^2\right)\right)}{\left(\alpha^2 z^2 + 2(1+z)^2(-1+\alpha)\right)^2}.$$

Proposition 2.1 *Given $\alpha = a + ib$, $a, b \in \mathbb{R}$, the stability of the fixed point $z = 1$ satisfies the following statements:*

1. *Fixed point $z = 1$ is an attractor, that is $|LG'(1, \alpha)| < 1$ if,*

$$a \leq -17 \quad ,$$

$$-17 < a < 4\left(-2 - \sqrt{5}\right) \quad , \quad -\sqrt{\frac{-16 + 32a - 15a^2 - a^3}{17 + a}} < b < \sqrt{\frac{-16 + 32a - 15a^2 - a^3}{17 + a}}$$

$$4\left(-2 + \sqrt{5}\right) < a < 1 \quad , \quad -\sqrt{\frac{-16 + 32a - 15a^2 - a^3}{17 + a}} < b < \sqrt{\frac{-16 + 32a - 15a^2 - a^3}{17 + a}}.$$

2. *$z = 1$ is indifferent, that is $|LG'(1, \alpha)| = 1$ if*

$$16 + 15a^2 + a^3 + 17b^2 + a(-32 + b^2) = 0.$$

3. *In any other case, $z = 1$ is a repulsive fixed point.*

This result comes from the analysis of the curve

$$16 + 15a^2 + a^3 + 17b^2 + a(-32 + b^2) = 0,$$

that can be separated in two different parts, as can be observed in Figure 1.
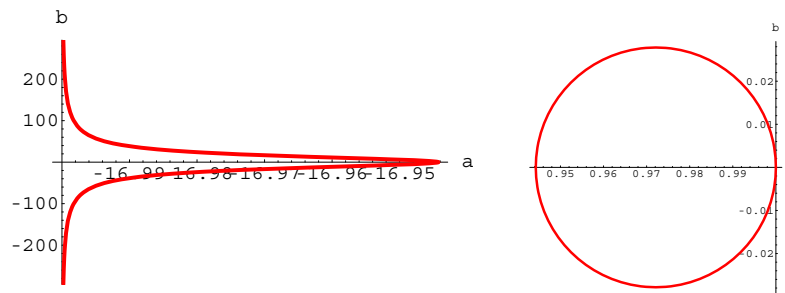
Figure 1: Boundary of the loci of stability of $z = 1$

Moreover, it is easy to check that $|LG'(-1, \alpha)| = 1$, as it is stated in the following resut.

Proposition 2.2 *The fixed point $z = -1$ is an indifferent point for every value of the parameter.*

In addition, from the expression of $LG'(z, \alpha)$ (or from the order of convergence of the members of the family), we obtain that the fixed points $z = 0$ and $z = \infty$ are also critical. In addition, we also have four free critical points

$$c_1(\alpha) = \frac{C_1 + \sqrt{C_1^2 - 4}}{2}, \quad c_2(\alpha) = \frac{C_1 - \sqrt{C_1^2 - 4}}{2},$$

$$c_3(\alpha) = \frac{C_2 + \sqrt{C_2^2 - 4}}{2}, \quad c_4(\alpha) = \frac{C_2 - \sqrt{C_2^2 - 4}}{2},$$

(a) Critical points c_1 and c_2 (b) A detail
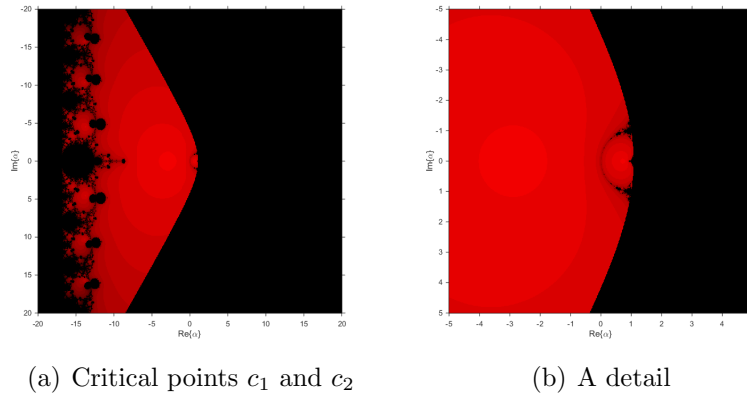
Figure 2: Parameter plane of $LG(z, \alpha)$ for c_1 and c_2

where

$$C_1 = \frac{-4\left(-1+\alpha\right)\left(-6+6\alpha+\alpha^2\right) - \alpha^2\sqrt{2\left(1-\alpha\right)\left(26-26\alpha+3\alpha^2\right)}}{6\left(-1+\alpha\right)\left(-2+2\alpha+\alpha^2\right)},$$

$$C_2 = \frac{-4\left(-1+\alpha\right)\left(-6+6\alpha+\alpha^2\right) + \alpha^2\sqrt{2\left(1-\alpha\right)\left(26-26\alpha+3\alpha^2\right)}}{6\left(-1+\alpha\right)\left(-2+2\alpha+\alpha^2\right)}.$$

Nevertheless, it is easy to check that $c_1(\alpha)c_2(\alpha) = 1$ and $c_3(\alpha)c_4(\alpha) = 1$ for every value of the parameter and their parameter planes are equivalent, so only two of them are independent and give rise to different parameter planes. Moreover, the number of critical points decreases for some values of the parameter.

(a) Critical points c_3 and c_4 (b) A detail
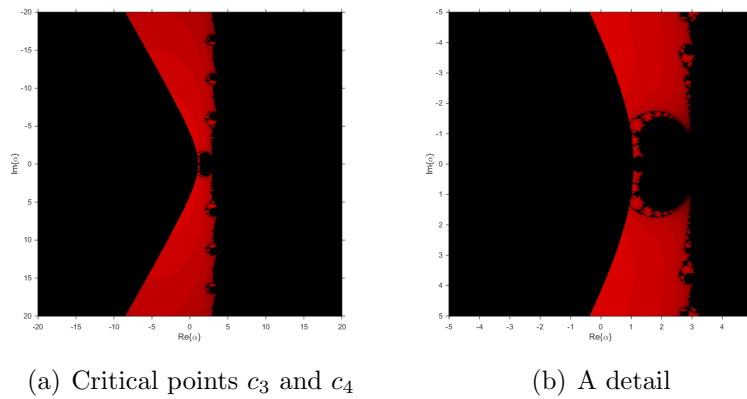
Figure 3: Parameter plane of $LG(z, \alpha)$ for c_3 and c_4

3   Parameter planes

As we have said, the dynamical behavior of operator $LG(z, \alpha)$ depends on the values of the parameter α and it is obtained by following the orbit of a free critical point. In this family there are four free critical points, but only two of them are independent. The fact that they are inverses two-by-two implies that if one critical orbit converges to $z = 0$ then the other one converges to $z = \infty$; therefore, it is enough to analyze the asymptotic behavior of one of the critical orbits, i.e. the orbits of the critical points, to study the existence of any other attractor than the roots. In fact, we only have two different parameter planes (Figures 2 and 3).

These parameter planes show us the reason for the bad general behavior of this family, since for all α values there are at least two free critics that will be in basins of attraction different from those of the roots. We must continue the analysis in order to detect why, when Newton fails, these methods behave better than it.

References

[1] S. Amat, S. Busquier, C. Bermúdez and S. Plaza. On two families of high order Newton type methods. *Appl. Math. Lett.*, 25:2209–2217, 2012.

[2] S. Amat, S. Busquier and S. Plaza. On the dynamics of a family of third-order iterative functions. *ANZIAM*, 48:343-359, 2007.

[3] J.M. Gutiérrez, M.A. Hernández and N. Romero. Dynamics of a new family of iterative processes for quadratic polynomials. *Comput. Appl. Math.*, 233:2688–2695, 2010.

[4] D.A. Budzko, A. Cordero and J.R. Torregrosa. A new family of iterative methods widening areas of convergence. *Appl. Math. Comput.*, 252:405-417, 2015.

[5] A.F. Beardon. *Iteration of rational functions*, Graduate Texts in Mathematics. Springer-Verlag New York, 1991.

[6] P. Blanchard. Complex Analytic Dynamics on the Riemann Sphere. *Bull. AMS*, 11(1):85–141, 1984.

A Family of Optimal Fourth Order Methods for Multiple Roots of Non-linear Equations *

Fiza Zafar^{a,b}[†], A. Cordero^{a}[‡] and Juan R. Torregrosa^{a}[§]

^{a} Instituto Universitario de Matematica Multidisciplinar,

Universitat Politècnica de Valencia, València 46022, Spain

^{b} Centre for Advanced Studies in Pure and Applied Mathematics,

Bahauddin Zakariya University, Multan 60800, Pakistan

1   Introduction

Construction of stable and optimal iterative methods for multiple roots having prior knowledge of multiplicity ($m > 1$) is one of the most important and challenging tasks in computational mathematics. Some optimal and non-optimal fourth-order methods have been developed in the recent past, however, it is indeed the need of time to design iterative methods for multiple roots not only in a general, optimal and efficient context but also in terms of deep analysis of their stable regions of convergent of initial estimations. Most recently, Behl et al. [1] in (2015), Behl et al. [2] in (2016), Behl et al. [3] (2017) and Lee et al. [5] (2017) have constructed and analyzed such families of methods. Moreover, most of these schemes are either the modification or extension of Newton's method or Newton-like methods by involving additional functional evaluations and increasing the amount of substeps of the original methods.

In this work, we propose an iterative family that has the flexibility of choice at both steps. The development of the scheme is based on using weight functions. The first step can not only recapture Newton's method for multiple roots as special case but is also capable of defining new choices of first substep and hence different iterative schemes in terms of both substeps. We compare our

[†]e-mail: fizazafar@gmail.com
[‡]e-mail: acordero@mat.upv.es
[§]e-mail: jrtorre@mat.upv.es

methods with the existing ones of the same order for standard test problems. From the numerical results, we find that our methods can be considered as a better alternative for the exiting methods of the same order.

2 Construction of Optimal Fourth-Order Scheme

Let α be a multiple zero with integer multiplicity $m > 1$, of $f : \mathbb{C} \to \mathbb{C}$ an analytic function in the neighborhood of α. Then, for a given initial guess x_0, we define the following iterative scheme in order to find an approximate zero of f:

$$
\begin{aligned}
y_n &= x_n - h(x_n)\frac{f(x_n)}{f'(x_n)}, \\
x_{n+1} &= x_n - Q_f(v_n)\frac{f(x_n)}{f'(x_n)},
\end{aligned}
\tag{1}
$$

where weight functions $h : \mathbb{C} \to \mathbb{C}$ and $Q_f : \mathbb{C} \to \mathbb{C}$ are analytic in the neighborhoods of α and 0, respectively with $v_n = \left[\frac{f'(y_n)}{f'(x_n)}\right]^{\frac{1}{m-1}}$.

The investigation on the convergence analysis of the proposed family (1) and the conditions on weight functions $h(x_n)$ and $Q_f(v_n)$ are apparent from the following result.

Theorem 1 *Let $f : \mathbb{C} \to \mathbb{C}$ be an analytic function in the neighborhood of the required multiple zero α of multiplicity $m \in \mathbb{N} - \{1\}$. In addition, we also consider that $Q : \mathbb{C} \to \mathbb{C}$ and $h : \mathbb{C} \to \mathbb{C}$ are an analytic functions in the neighborhood of origin and multiple zero α, respectively. Then, for an initial guess x_0 sufficiently close to α the family of iteration functions (1) has fourth-order convergence when the following conditions hold:*

$$
\begin{aligned}
h(\alpha) &= m, \quad h'(\alpha) = 0, \quad h''(\alpha) = 0, \\
Q_f(0) &= m, \quad Q'_f(0) = m, \quad Q''_f(0) = \frac{4m^2}{m-1}
\end{aligned}
\tag{2}
$$

and also $\left|h'''(\alpha)\right| < \infty, \left|Q'''_f(0)\right| < \infty$.

Remark 1 *Proposed family (1) has an advantage of making selection at both steps. It is also clear that the first step recaptures Newton's method as special case and it is capable of obtaining first step different from the traditional choice of Newton's scheme.*

From Theorem 1, we can obtain several new multiple root finding two-point methods by using different cases for $h(x_n)$ and $Q_f(v_n)$ in the proposed scheme (1). Some particular cases of the proposed scheme are given as follows:

NM1: We take $h(x_n) = m$ and $Q_f(v_n) = m + mv_n + \frac{2m^2}{m-1}v_n^2 + Q_3v_n^3$, for $Q_3 = 32.6$ in (1).

NM2: Also as another special case let $h(x_n) = \frac{m+w_n^3}{1+a_2w_n^3}$ with $w_n = f(x_n)$, $a_2 = -16.3$ and $Q_f(v_n)$ is same as that of NM1.

NM3: Lastly, we take $h(x_n) = m + a_3 f(x_n)^m$ for $a_3 = 50$ with $Q_f(v_n)$ is same as that of NM1.

It is noteworthy that the selection of specific values of parameters Q_3, a_2 and a_3 can be made under the point of view of an improvement of the stability and a widening of the set of converging initial estimations. These aspects will be analyzed in future works.

3   Numerical Results

We investigate the performance and convergence behavior of our proposed fourth order methods namely denoted by NM1, NM2 and NM3, respectively, by carrying out some test functions involving standard nonlinear functions. We compare the methods with the recent optimal fourth order method given by Lee [5] (LKM).

For numerical tests, all computations have been performed in computer algebra software Maple 16 using 1000 significant digits of precision. Table 1 shows the per step numerical errors of approximating real root $|x_n - x_{n-1}|$ and the absolute residual error of the test function for the first three iterations , where $E(-i)$ denotes $E \times 10^{-i}$ in all the tables. The initial approximation x_0, the computational order of convergence (COC, see [6])

$$r_c \approx \frac{\log |f(x_{n+1})/f(x_n)|}{\log |f(x_n)/f(x_{n-1})|}$$

and asymptotic error constant

$$a_c = \frac{|x_{n+1} - x_n|}{|x_n - x_{n-1}|^4},$$

are also included in this table. We have taken into consideration the following standard test problem.

Example 1 *Let us consider the following standard non-linear test function:*

$$f_1(x) = \left(\sin\left(\frac{1}{x}\right) - x^3 + 1 \right)^3. \tag{3}$$

The above function has a multiple zero at $\alpha \approx 1.20253919024135112296187908278$ of multiplicity $m = 3$; in our tests, we use as with initial guess $x_0 = 1.25$.

Example 2 *Assume another non-linear test problem which is given below:*

$$f_2(x) = (x-1)\left(x\ln(x) - \sqrt{x} + x^4\right)^2. \tag{4}$$

Function f_2 has a multiple zero at $\alpha = 1$ with multiplicity $m = 3$, and initial guess $x_0 = 1.1$.

$f_1(x) = \left(\sin\left(\frac{1}{x}\right) - x^3 + 1\right)^3, x_0 = 1.25$					
Methods	n	$|x_n - x_{n-1}|$	$|f(x_n)|$	r_c	a_c
LKM	1	4.745384266(-2)	3.749911402(-14)		
	2	6.967096919(-6)	6.184240718(-60)	3.966126232	1.373935383
	3	3.820695713(-21)	4.575159589(-243)	3.999998725	1.621567542
NM1	1	4.745890969(-2)	7.606120898(-16)		
	2	1.900060722(-6)	3.226081940(-74)	4.410031791	3.745384886(-1)
	3	6.626315657(-26)	1.031678958(-307)	4.000088731	5.083960864(-3)
NM2	1	4.746080422(-2)	1.875993439(-23)		
	2	5.530835722(-9)	1.179732013(-104)	3.895632250	1.090059958(-3)
	3	4.738506100(-36)	1.844915398(-429)	4.000000186	5.063821720(-3)
NM3	1	4.746080809(-2)	5.109598493(-25)		
	2	1.664086562(-9)	6.492241143(-111)	3.833089962	3.279709665(-4)
	3	3.883099658(-38)	1.692082630(-454)	4.000000052	5.063780615(-3)

Table 1: Comparison of multiple root finding methods for $f_1(x)$

Example 3 *Continuous Stirred Tank Reactor (CSTR)*

Consider the isothermal continuous stirred tank reactor (CSTR). Components A & R are fed to the reactor at rates of Q and q-Q respectively. The following reaction scheme develops in the reactor:

$$\begin{aligned} A + R &\rightarrow B \\ B + R &\rightarrow C \\ C + R &\rightarrow D \\ D + R &\rightarrow E. \end{aligned}$$
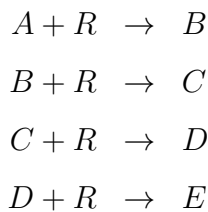
The problem was analysed by Douglas [4] in order to design simple feedback control systems. In the analysis, he gave the following equation for the transfer function of the reactor:

$$K_C \frac{2.98\,(x + 2.25)}{x^4 + 11.50x^3 + 47.49x^2 + 86.0325x + 51.23266875} = -1,$$

$f_2(x) = (x-1)\left(x\ln(x) - \sqrt{x} + x^4\right)^2$, $x_0 = 1.1$					
Methods	n	$\lvert x_n - x_{n-1}\rvert$	$\lvert f(x_n)\rvert$	r_c	a_c
LKM	1	9.973228391(−2)	3.888577634(−10)		
	2	2.677160895(−4)	2.352566232(−40)	3.853170778	2.706022566
	3	2.264847390(−14)	3.175439431(−161)	3.999892196	4.409009687
NM1	1	9.983089756(−2)	9.796949329(−11)		
	2	1.691024370(−4)	2.880566614(−47)	4.327807202	1.702511144
	3	1.124650632(−16)	1.859782372(−193)	4.001739973	1.375365624(−1)
NM2	1	9.979306935(−2)	1.795411740(−10)		
	2	2.069306440(−4)	3.353782987(−46)	4.368837238	2.086523533
	3	2.549000828(−16)	3.417356322(−189)	4.002164322	1.390179047(−1)
NM3	1	9.979300583(−2)	1.797065853(−10)		
	2	2.069941600(−4)	3.366337536(−46)	4.368901372	2.0871692921
	3	2.552177510(−16)	3.468814483(−189)	4.002165039	1.390203910(−1)

Table 2: Comparison of multiple root finding methods for $f_2(x)$

where K_C is the gain of the proportional controller. The control system is stable for values of K_C that yields roots of the transfer function having negative real part. If we choose $K_C = 0$ we get the poles of the open-loop transfer function as roots of the nonlinear equation:

$$f_3(x) = x^4 + 11.50x^3 + 47.49x^2 + 86.0325x + 51.23266875 = 0 \tag{5}$$

given as:

$$x = -1.45, -2.85, -2.85, -4.35.$$

So, we see that there is one multiple root with multiplicity 2. We take $m = 2$ and $x_0 = -3$.

It is apparent from the construction and numerical results that our proposed family is optimal, efficient in terms of small residual errors and flexible in terms of choice of first substep different from Newton's method.

$f_3(x) = x^4 + 11.50x^3 + 47.49x^2 + 86.0325x + 51.23266875$, $x_0 = -3.0$					
Methods	n	$\|x_n - x_{n-1}\|$	$\|f(x_n)\|$	r_c	a_c
LKM	1	1.521916174(−1)	1.008561681(−5)		
	2	2.191856237(−3)	1.197664817(−13)	2.160003501	4.085536881
	3	2.388130175(−7)	7.327948015(−58)	5.578713509	1.034688766(4)
NM1	1	1.500116811(−1)	2.865442731(−10)		
	2	1.168116988(−5)	9.489712144(−44)	4.075138694	2.306672973(−2)
	3	2.125772928(−22)	1.141014571(−177)	4.000006184	1.141751293(−2)
NM2	1	1502676260(−1)	1.504078150(−7)		
	2	2.676260156(−4)	7.279363956(−33)	4.606463395	5.248879797(−1)
	3	5.887583365(−17)	3.950506405(−134)	4.000186954	1.147687774(−2)
NM3	1	1.432777532(−1)	9.492439797(−5)		
	2	6.722245222(−3)	4.948225872(−18)	4.927870715	1.595141817(1)
	3	1.535023789(−9)	8.435094413(−75)	4.273783471	7.517227707(−1)

Table 3: Comparison of multiple root finding methods for $f_3(x)$

References

[1] R. Behl, A. Cordero, S.S. Motsa, J.R. Torregrosa, On developing fourth-order optimal families of methods for multiple roots and their dynamics, *Appl. Math. Comput.* 265(15) (2015) 520–532.

[2] R. Behl, A. Cordero, S.S. Motsa, J.R. Torregrosa, V. Kanwar, An optimal fourth-order family of methods for multiple roots and its dynamics, *Numer. Algor.* 71(4) (2016) 775–796.

[3] R. Behl, A. Cordero, S.S. Motsa, J.R. Torregrosa, Multiplicity anomalies of an optimal fourth-order class of iterative methods for solving nonlinear equations, *Nonlinear Dyn.* 91 (2018) 81–112.

[4] J. M. Douglas, Process Dynamics and Control, Vol. 2, Prentice Hall, Englewood Cliffs, NJ, 1972.

[5] M. Lee, Y.I. Kim, Á.A. Magreñán, On the dynamics of a triparametric family of optimal fourth-order multiple-zero finders with a weight function of the principal mth root of a function-to function ratio, *Appl. Math. Comput.* 315 (2017) 564–590.

[6] L.O. Jay, A note on Q-order of convergence, *BIT Numer. Math.* 41 (2001) 422–429.

Randomizing the von Bertalanffy growth model: Theoretical analysis and computing

J. Calatayud[♭], J.-C. Cortés[♭], M. Jornet[♭],[*]
R.J. Villanueva[♭]

(♭) Instituto Universitario de Matemática Multidisciplinar,

Universitat Politècnica de València, Spain.

November 30, 2018

1 Bertalanffy model

Bertalanffy model [1] is a biological ordinary differential equation model that describes the relationship between the metabolism and the growth of an organism. The metabolism is divided into anabolism (synthesis) and catabolism (destruction). The model assumes that the body weight $W(t)$ of an animal is the result of the counteraction of the processes of anabolism and catabolism:

$$W'(t) = \eta W^m(t) - \kappa W^n(t),$$

where η and κ are the constants of anabolism and catabolism, proportional to some power of the body weight (law of allometry).

The surface rule states that the dependence of anabolism on body weight takes the power $m = 2/3$ [2, 3]. Bertalanffy justified that the rate of catabolism should have the power $n = 1$. Bertalanffy model thus becomes

$$W'(t) = \eta W^{\frac{2}{3}}(t) - \kappa W(t).$$

---

[*]e-mail: marjorsa@doctor.upv.es

2   Random non-autonomous Bertalanffy model

The random non-autonomous Bertalanffy model is defined as

$$
\begin{cases}
x'(t,\omega) = a(t,\omega)x(t,\omega) + b(t,\omega)x(t,\omega)^{\frac{2}{3}}, \ t \in [t_0, T], \ \omega \in \Omega \\
x(t_0, \omega) = x_0(\omega), \ \omega \in \Omega.
\end{cases}
\tag{1}
$$

We work on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $a(t,\omega)$ and $b(t,\omega)$ are stochastic processes and $x_0(\omega)$ is a random variable. The response

$$
x(t,\omega) = \left(x_0(\omega)^{\frac{1}{3}} e^{\frac{1}{3}\int_{t_0}^t a(s,\omega)\, \mathrm{d}s} + \frac{1}{3} \int_{t_0}^t b(s,\omega) e^{\frac{1}{3}\int_s^t a(r,\omega)\, \mathrm{d}r}\, \mathrm{d}s \right)^3
\tag{2}
$$

solves (1) in some probabilistic sense.

Our goal is to understand the probabilistic behavior of $x(t,\omega)$ and to compute/approximate its PDF, $f_{x(t)}(x)$. The following results are proved from extant theorems on the random non-autonomous linear differential equation, see [4].

3   Solution process in the sample path and mean square senses

In this section, we show two results on the existence of a sample path and mean square solution to (1).

Theorem 3.1 (Sample path solution) *Suppose $a(\cdot, \omega), b(\cdot, \omega) \in \mathrm{L}^1([t_0, T])$, for a.e. $\omega \in \Omega$. Then the stochastic process $x(t, \omega)$ given by (2) satisfies that, for a.e. $\omega \in \Omega$, $x(\cdot, \omega)$ is absolutely continuous on $[t_0, T]$ and satisfies (1) for a.e. $t \in [t_0, T]$.*

If $a(\cdot, \omega)$ and $b(\cdot, \omega)$ are continuous on $[t_0, T]$, then $x(\cdot, \omega)$ is in $C^1([t_0, T])$ and satisfies (1) for all $t \in [t_0, T]$.

Theorem 3.2 (Mean square solution) *If $a(t, \omega)$ and $b(t, \omega)$ are continuous in the $\mathrm{L}^{48}(\Omega)$ and $\mathrm{L}^{24}(\Omega)$ setting, respectively,*

$$
S := \sup_{t \in [t_0, T]} \left\| e^{\pm \int_{t_0}^t a(s,\omega)\, \mathrm{d}s} \right\|_{\mathrm{L}^{48}(\Omega)} < \infty
\tag{3}
$$

and $x_0 \in \mathrm{L}^4(\Omega)$, then $x(t, \omega)$ defined by (2) is differentiable in the mean square sense and satisfies the random Bertalanffy model (1).

4   Obtaining the PDF of the solution stochastic process

Let $a(t,\omega)$ and $b(t,\omega)$ be stochastic processes in $\mathrm{L}^2([t_0, T] \times \Omega)$. We can expand both $a(t,\omega)$ and $b(t,\omega)$ via a Karhunen-Loève expansion:

$$a(t,\omega) = \mu_a(t) + \sum_{j=1}^{\infty} \sqrt{\nu_j}\, \phi_j(t)\xi_j(\omega), \quad b(t,\omega) = \mu_b(t) + \sum_{j=1}^{\infty} \sqrt{\gamma_j}\, \psi_j(t)\eta_j(\omega),$$

(4)

respectively.

From $a, b \in \mathrm{L}^2([t_0, T] \times \Omega)$, we have $a(\cdot,\omega), b(\cdot,\omega) \in \mathrm{L}^1([t_0, T])$, therefore the process $x(t,\omega)$ has absolutely continuous sample paths and solves the random Bertalanffy model (1). Under stricter assumptions, the process $x(t,\omega)$ will be a mean square solution too.

Consider the following truncations:

$$a_N(t,\omega) = \mu_a(t) + \sum_{j=1}^{N} \sqrt{\nu_j}\, \phi_j(t)\xi_j(\omega), \quad b_N(t,\omega) = \mu_b(t) + \sum_{j=1}^{N} \sqrt{\gamma_j}\, \psi_j(t)\eta_j(\omega),$$

$$x_N(t,\omega) = \left(x_0(\omega)^{\frac{1}{3}} e^{\frac{1}{3}\int_{t_0}^{t} a_N(s,\omega)\,\mathrm{d}s} + \frac{1}{3}\int_{t_0}^{t} b_N(s,\omega)e^{\frac{1}{3}\int_{s}^{t} a_N(r,\omega)\,\mathrm{d}r}\,\mathrm{d}s \right)^3. \quad (5)$$

Denote $\boldsymbol{\xi}_N = (\xi_1, \ldots, \xi_N)$, $\boldsymbol{\eta}_M = (\eta_1, \ldots, \eta_M)$ and

$$K_a(t, \boldsymbol{\xi}_N) = \int_{t_0}^{t} \left(\mu_a(s) + \sum_{j=1}^{N} \sqrt{\nu_j}\, \phi_j(s)\xi_j \right)\,\mathrm{d}s,$$

$$S_b(s, \boldsymbol{\eta}_N) = \mu_b(s) + \sum_{i=1}^{N} \sqrt{\gamma_i}\, \psi_i(s)\eta_i.$$

Suppose that x_0 and $(\xi_1, \ldots, \xi_N, \eta_1, \ldots, \eta_N)$ are absolutely continuous (AC) and independent, for each $N \geq 1$. By the Random Variable Transformation (RVT) technique, for $0 \neq x \in \mathbb{R}$,

$$f_{x_N(t)}(x) = \frac{1}{x^{\frac{2}{3}}}\mathbb{E}\left[f_{x_0}\left(\left\{ x^{\frac{1}{3}}\, e^{-\frac{1}{3}K_a(t,\boldsymbol{\xi}_N)} - \frac{1}{3}\int_{t_0}^{t} S_b(s, \boldsymbol{\eta}_N)e^{-\frac{1}{3}K_a(s,\boldsymbol{\xi}_N)}\,\mathrm{d}s \right\}^3 \right) \right.$$

$$\left. \cdot \left\{ x^{\frac{1}{3}}\, e^{-\frac{1}{3}K_a(t,\boldsymbol{\xi}_N)} - \frac{1}{3}\int_{t_0}^{t} S_b(s, \boldsymbol{\eta}_N)e^{-\frac{1}{3}K_a(s,\boldsymbol{\xi}_N)}\,\mathrm{d}s \right\}^2 e^{-\frac{1}{3}K_a(t,\boldsymbol{\xi}_N)} \right] \quad (6)$$

(this expectation may be approximated with MC simulations).

The goal is to prove that $f_{x(t)}(x) = \lim_{N \to \infty} f_{x_N(t)}(x)$, under certain conditions.

Theorem 4.1 *Assume the following four hypotheses:*

H1 : $a, b \in \mathrm{L}^2([t_0, T] \times \Omega)$;

H2 : x_0 and $(\xi_1, \ldots, \xi_N, \eta_1, \ldots, \eta_N)$ are AC and independent, $N \geq 1$;

H3 : f_{x_0} is continuous on $\mathbb{R}$ and $f_{x_0}(x) \leq \dfrac{C}{|x|^{\frac{2}{3}}}$, for $x \neq 0$;

H4 : $\|\mathrm{e}^{-\frac{1}{3}K_a(t, \boldsymbol{\xi}_N)}\|_{\mathrm{L}^2(\Omega)} \leq C$, for all $N \geq 1$ and $t \in [t_0, T]$.

Then, for all $0 \neq x \in \mathbb{R}$ and $t \in [t_0, T]$, the sequence $\{f_{x_N(t)}(x)\}_{N=1}^{\infty}$ given by (6) converges to the density $f_{x(t)}(x)$ of the solution process $x(t, \omega)$ given by (2).

Theorem 4.2 *Assume that*

H1 : $a, b \in \mathrm{L}^2([t_0, T] \times \Omega)$;

H2 : $x_0, \eta_1, (\xi_1, \ldots, \xi_N, \eta_2, \ldots, \eta_N)$ are AC and independent, $N \geq 1$;

H3 : f_{η_1} is continuous and bounded on $\mathbb{R}$;

H4 : $\xi_1, \xi_2, \ldots$ have compact support in $[-A, A]$ and $\psi_1 > 0$ on (t_0, T).

Then, for each $0 \neq x \in \mathbb{R}$ and $t \in (t_0, T]$, the sequence $\{f_{x_N(t)}(x)\}_{N=1}^{\infty}$ given by (6) converges to the density $f_{x(t)}(x)$ of the solution process $x(t, \omega)$ given by (2).

5 Approximation of the expectation and variance of the solution process

The following result shows conditions under which the expectation and variance of $x(t)$ can be approximated.

Theorem 5.1 *If $a(t, \omega)$ is a Gaussian process or $\xi_1, \xi_2, \ldots$ have a common compact support, if $\|x_0\|_{\mathrm{L}^{2+s}(\Omega)} < \infty$ and if*

$$\|\mu_b\|_{\mathrm{L}^{6+s}([t_0, T])} + \sum_{i=1}^{\infty} \sqrt{\gamma_i} \|\psi_i\|_{\mathrm{L}^{6+s}([t_0, T])} \|\eta_i\|_{\mathrm{L}^{6+s}(\Omega)} < \infty \tag{7}$$

for some $s > 0$, then $x(t, \omega) \in \mathrm{L}^2(\Omega)$ and $x_N(t, \omega)$ tends in $\mathrm{L}^2(\Omega)$ to $x(t, \omega)$, for each $t \in [t_0, T]$. As a consequence, $\mathbb{E}[x_N(t, \omega)] \to \mathbb{E}[x(t, \omega)]$ and $\mathbb{V}[x_N(t, \omega)] \to \mathbb{V}[x(t, \omega)]$ as $N \to \infty$, for each $t \in [t_0, T]$.

6   Numerical example

We work on $[t_0, T] = [0, 1]$. Let

$$a(t, \omega) = \sum_{j=1}^{\infty} \frac{\sqrt{2}}{j^3} \sin(tj\pi)\xi_j(\omega), \tag{8}$$

where $\xi_1, \xi_2, \ldots$ are independent with distribution Uniform$(-\sqrt{3}, \sqrt{3})$. Let

$$b(t, \omega) = \sum_{i=1}^{\infty} \frac{\sqrt{2}}{i^4 + 6} \sin(ti\pi)\eta_i(\omega), \tag{9}$$

where $\eta_1, \eta_2, \ldots \sim \mathrm{Normal}(0, 1)$ are independent. Let $x_0 \sim \mathrm{Exponential}(2)$. It is assumed $x_0, \xi_1, \xi_2, \ldots$ and $\eta_1, \eta_2, \ldots$ to be independent.
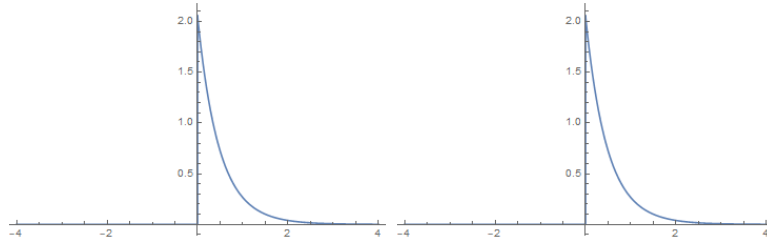
Figure 1: $f_{x_N(0.3)}(x)$ for $N = 5$ (left) and $N = 6$ (right) at $t = 0.3$. Observe the convergence.

N	1	2	3	4	5	6
$\mathbb{E}[x_N(0.3, \omega)]$	0.5071	0.5078	0.5076	0.5076	0.5076	0.5076
$\mathbb{V}[x_N(0.3, \omega)]$	0.2754	0.2769	0.2768	0.2768	0.2768	0.2768

Table 1: $\mathbb{E}[x_N(0.3, \omega)]$ and $\mathbb{V}[x_N(0.3, \omega)]$ for $N = 1, 2, 3, 4, 5, 6$.

References

[1] L. v. Bertalanffy. *Quantitative laws in metabolism and growth.* Q. Rev. Biol. 32(3) (1957) 217–231.

[2] Craig R. White, Roger S. Seymour. *Mammalian basal metabolic rate is proportional to body mass 2/3.* (2003) Proceedings of the National Academy of Sciences of the United States of America 100:4046-4049. Doi: 10.1073/pnas.0436428100.

[3] Jayanth R. Banavar, John Damuth, Amos Maritan, Andrea Rinaldo. *Ontogenetic growth. Modelling universality and scaling.* Nature volume 420, page 626 (2002). Doi:10.1038/420626a.

[4] J. Calatayud, J.-C. Cortés, M. Jornet. *Approximation of the probability density function of the randomized heat equation with non-homogeneous boundary conditions.* arXiv:1805.03738 (2018).

A Gauss-Legendre Product Quadrature for the Neutron Transport Equation

A. Bernal *, S. Morató, R. Miró and G. Verdú

Institute for Industrial, Radiophysical and Environmental Safety (ISIRYM)

Universitat Politècnica de València, Valencia, Spain

November 30, 2018

1   Introduction

The prediction of the neutron distribution inside nuclear reactors is important because of several reasons. First, the power generated inside the nuclear reactor core is related with the neutron distribution. Consequently, an accurate prediction of the power distribution is crucial to determine an accurate prediction of the thermal-hydraulic behavior. Thus, one makes sure that these values are under the safety limits to avoid any kind of accident and damaging the components of the nuclear reactor core. Second, the neutron flux irradiates the different structural components of the nuclear reactor core, including the fuel, which influences strongly the mechanical resistance of them [1].

The Neutron Transport Equation describes accurately the transport of neutrons in any control domain, so one can solve this equation to get an accurate prediction of the neutron distribution inside the nuclear reactor core. This equation is an integrodifferential equation containing spatial and time derivatives terms and it depends on seven independent variables: three of the space domain, two of the direction domain, one of the energy domain and one of time domain [1]. Even if one considers only the steady state, the

---

*e-mail: abernal@iqn.upv.es

solution of this equation in nuclear reactor cores is not straightforward. In fact, one should use numerical methods to solve it.

These numerical methods discretize each variable of the Neutron Transport Equation: energy, space and direction. In order to deal with the energy dependence, one commonly applies an energy multi-group approximation, obtaining a set of equations depending on the number of energy groups [2]. In this case, the final solution will be the sum of the solutions for each group. As regards the direction, the Discrete Ordinates method is the method most used to discretize the direction variables. This approach solves the Neutron Transport Equation for a set of selected directions (quadrature sets), obtaining a set of directional equations and solutions for each equation which are the angular flux. The final solution is the weighted sum of all the directional solutions. Finally, space discretization is solved by classical methods, such as Finite Difference, Finite Element or Finite Volume Methods.

In this work, the authors focused on the selected directions used in the Discrete Ordinates Formulation. Each direction can be described by a polar and azimuthal angle. The value of these directions and their weights is vital to obtain accurate results of the neutron flux distribution. These weights are used to perform different numerical integrations of the variables depending on the direction variables. There are a number of quadrature sets used for this purpose, such as Level-Symmetric, or Legendre-Chebyshev [3]. All these approaches might obtain accurate solutions of the neutron flux for a large number of directions. However, the larger the number of directions, the larger the number of equations, and consequently the computational cost.

In addition, the authors studied the numerical integration of the previous quadrature sets in numerical integration of several functions. The authors realized that these sets might provide wrong results. Therefore, they developed a new quadrature set which is based in a product quadrature, for the polar and azimuthal variable. The authors tested this new quadrature set for integrating numerically functions depending on direction variables. These values are also compared with the values obtained with the other sets mentioned above. Finally, the authors tested the different quadrature sets for solving the Neutron Transport Equation.

The outline of the rest of the paper is as follows. Section 2 explains the new product quadrature. Section 3 shows the results.

2 Method

The authors developed a new product quadrature using the Gauss-Legendre quadrature for the polar $(\cos(\theta))$ and azimuthal (φ) variables. First, the method considers N_p collocation points for $\cos(\theta)$. This value N_p is an even number, with the aim of conserving the symmetry with respect to θ. For each one of the N_p collocation points, the method uses the weights w_i^x and collocation points x_i of the Gauss-Legendre quadrature.

Second, the method performs four integrals with respect to φ, one in each quadrant of the domain of φ. For each quadrant (q), one chooses N_a collocation points y_j and weights w_j^y of the Gauss-Legendre quadrature. Then, one performs a change of interval from the Gauss-Legendre domain to the domain of each quadrant. To sum up, Equations 1-3 show the collocation points and weights, which are normalized to 8.

$$\cos(\theta_i) = x_i \qquad , 1 \leq i \leq N_p \tag{1}$$

$$\varphi_{q,j} = \frac{\pi}{4} y_j + \frac{(2q-1)\pi}{4} \qquad , 1 \leq j \leq N_a \qquad , 1 \leq q \leq 4 \tag{2}$$

$$w_{i,q,j} = \frac{8 \cdot w_i^x \cdot w_j^y}{\sum_{i=1}^{N_p} \sum_{q=1}^{4} \sum_{j=1}^{N_a} w_i^x \cdot w_j^y} \qquad , 1 \leq j \leq N_a \qquad , 1 \leq q \leq 4 \qquad , 1 \leq i \leq N_p \tag{3}$$

3 Results

On the one hand, this section tests this new quadrature set for integrating numerically functions depending on direction variables. These values are also compared with the values obtained with the following quadrature sets: Level-Symmetric (S_n), Legendre-Chebyshev $(P_n T_n)$ and product quadrature based on Gauss-Legendre quadrature for the polar variable and equal weights for the azimuthal variable $(P_n EW)$.

On the other hand, the authors tested the different quadrature sets for solving the Neutron Transport Equation, with the Discrete Ordinates and Finite Difference Method. This section also includes a sensitivity analysis of the different quadrature sets and number of directions.

First, the following integral is evaluated: $I = \frac{1}{4\pi} \int_{-1}^{1} \exp(\mu)d\mu \int_{0}^{2\pi} \exp(\varphi)d\varphi$. One can integrate this term analytically, and evaluate the numerical integration errors with relative errors. Table 1 shows the relative errors for the new product quadrature, where N_d is the number of directions. Table 2 displays the relative errors for other quadrature sets.

Table 1: Relative errors (%) of the numerical integration with the product quadrature based on Gauss-Legendre

N_d	$N_p/2$	N_a	$Error(\%)$
8	1	1	9.88
16	2	1	0.81
32	2	2	6.32
24	3	1	3.48
48	3	2	3.35
72	3	3	3.45
32	4	1	4.48
64	4	2	1.95
96	4	3	2.30
128	4	4	2.13
40	5	1	4.96
80	5	2	1.22
120	5	3	1.62
160	5	4	1.55
200	5	5	1.46
48	6	1	5.22
96	6	2	0.79
144	6	3	1.19
192	6	4	1.18
240	6	5	1.11
288	6	6	1.07

From these tables, one draws two conclusions. First, the new product quadrature obtains better results than other quadrature sets for a lower number of collocation points. Second, the results of the new product quadrature depend on the number of collocation points for both variables. In fact, one

Table 2: Relative errors (%) of the numerical integration with different quadrature sets

N_d	S_n	P_nT_n	P_nEW
8	9.88	9.88	9.88
24	5.12	4.47	1.53
48	3.34	2.65	0.26
80	2.45	1.77	0.80
120	1.94	1.27	0.98
168	1.58	0.96	1.02
224	1.35	0.75	1.01
288	1.16	0.61	0.97

obtains better results if the number of collocation points for the polar and azimuthal variables is similar.

As regards the Neutron Transport Equation, the authors tested the different quadrature sets with a formulation based on the Discrete Ordinates and Finite Difference Method. This formulation was applied to a homogeneous 2D reactor. The geometry of the reactor is a rectangle of 40 cm × 40 cm, which is modeled with a structured mesh of 10 × 10. The Neutron Transport Equation uses the the two-energy group formulation with isotropic scattering.

The authors calculated the eigenvalue problem of the Neutron Transport Equation of this reactor, with vacuum boundary conditions. The largest eigenvalue is evaluated for the different quadrature sets. Table 3 shows the eigenvalue for the new product quadrature, where N_d is the number of directions. Table 4 displays the eigenvalue for other quadrature sets.

One draws two conclusions from Tables 3 and 4. First, the new product quadrature obtains better results than other quadrature sets for a lower number of collocation points. Actually, the eigenvalue calculated with 16 directions with the new product quadrature is the same as the eigenvalue obtained with other quadrature sets with 220 directions. Second, the results obtained with the new product quadrature depend on the number of collocation points for both variables. In particular, one obtains excellent results with $N_p \geq 2$ and $N_a \geq 2$.

Table 3: Eigenvalue for the product quadrature based on Gauss-Legendre

N_d	$N_p/2$	N_a	*Eigenvalue*
1	1	4	1.099438
1	2	8	1.104771
1	3	12	1.104438
2	1	8	1.108929
2	2	16	1.114042
2	3	24	1.113738
3	1	12	1.109215
3	2	24	1.114346
3	3	36	1.114043

Table 4: Eigenvalue for different quadrature sets

N_d	S_n	P_nT_n	P_nEW
4	1.099438	1.099438	1.099438
12	1.112082	1.111766	1.110148
24	1.113045	1.113058	1.111119
40	1.113477	1.113504	1.111668
60	1.113649	1.113710	1.112036
84	1.113762	1.113822	1.112188
112	1.113824	1.113888	1.112373
144	1.113872	1.113933	1.112558
180	1.113902	1.113965	1.112710
220	1.113929	1.113986	1.112802

References

[1] D.G. Cacuci, Handbook of Nuclear Engineering v. 3. Springer US, 2010.

[2] W.M. Stacey, Nuclear Reactor Physics. John Wiley & Sons, 2007.

[3] A. Hébert, Applied Reactor Physics. Presses internationales Polytechnique, 2009.

PGD path planning for dynamic obstacle
robotic problems

L. Hilario[1] *, N. Montés[1], M.C. Mora[2], E. Nadal[3], A. Falcó[1], F. Chinesta[4]
and J.L.Duval[5]

(1) ESI International Chair Universidad Cardenal Herrera-CEU, CEU Universities,

C/ San Bartolomé, 55 Alfara del Patriarca 46115 (Spain),

(2) Universitat Jaume I, Castellón,

(3) Universitat Politcnica de Valncia,

(4) ENSAM Paris Tech,

(5) ESI Group,

November 19, 2018

1   Introduction

A fundamental robotics task is to plan collision-free motions among a set of static and known obstacles from a start to a goal position. The geometric construction of this planning strategy is computationally hard and hence unfeasible for its use in real-time (RT) applications [12]. This motion planning (or the piano mover's) problem has motivated many works in the field of robotics.

In the above context, one of the most popular algorithms is the so-called Artificial Potential Field technique (APF) [6, 12, 13].This technique is very fast for real-time applications, except when the robot is trapped in a deadlock (a local minima of the potential function). The solution of this problem lies in the use of harmonic functions to generate the potential field [7]. Despite

---

*e-mail: luciah@uchceu.es

their attractive properties, path planning based on harmonic functions has some drawbacks that have prevented the extensive use of this methodology, as indicated in [5].

Lately, a novel approach called the Proper Generalized Decomposition (PGD) has appeared to approximate the solutions of non-linear convex variational problems [3]. In our previous work, [11], [4], was presented for the first time, the Proper Generalized Decomposition method to solve the motion planning problem. In that work, the PGD was designed just for static obstacles and computed as a vademecum for all Start and goal combinations.

However, in a realistic scenario, it is necessary to take into account dynamic obstacles. The goal of this work is to solve this problem applying PGD considering dynamic obstacles as an extra parameter.

2 PGD-Vademecum for Path Planning in static environment

For the path planning application proposed here, the space is not decomposed in X and Y but parameters in the model are set as additional extra-coordinates, that is, a PGD-Vademecum, see [1] and [2]. In our previous works,[8],[9],[10], [11] the additional extra-parameters are considered in the source term, being all the posible combinations of the start and goal configurations.

2.1 Source term definition

Consider the functions $g_S : \Omega_{\underline{X}} \times \Omega_{\underline{S}} \to \mathbb{R}$ and $g_T : \Omega_{\underline{X}} \times \Omega_{\underline{T}} \to \mathbb{R}$ as 2D Gaussian density distributions centered in the start $\underline{S} = (s_1, s_2) \in \Omega_{\underline{S}}$ and target configurations $\underline{T} = (t_1, t_2) \in \Omega_{\underline{T}}$, respectively. Both functions are assume to have equal variance given by a diagonal matrix $\Sigma = \text{diag}(r, r)$ for some $r > 0, r \in \mathbb{R}$. More precisely, we can write:

$$g_S = g_S((x, y); (s_1, s_2), r) = (2\pi r)^{-1} e^{-\frac{1}{2r}((x-s_1)^2 + (y-s_2)^2)},$$

$$g_T = g_T((x, y); (t_1, t_2), r) = (2\pi r)^{-1} e^{-\frac{1}{2r}((x-t_1)^2 + (y-t_2)^2)}$$

and hence $\Omega_{\underline{X}} = \Omega_x \times \Omega_y$, $\Omega_{\underline{S}} = \Omega_s \times \Omega_r$ and $\Omega_{\underline{T}} = \Omega_t \times \Omega_r$.

Here $\Omega_{\underline{X}} = \Omega_{\underline{S}} = \Omega_{\underline{T}} \subset \mathbb{R}^2$.

Let's assume that the source term f is non-uniform, that is, $f = g_S - g_T$ when $(x, y) \in \Omega_{\underline{X}}$ and zero otherwise. Then, the Poisson equation is now

$$-\Delta u(\underline{X}, \underline{S}, \underline{T}) \quad = f(\underline{X}, \underline{S}, \underline{T}) \tag{1}$$

2.2 PGD-Vademecum solution

The PGD-Vademecum is constructed considering that the solution of the potential field u can be constructed as a finite sum of terms, each one consisting of the product of three functions: a function R of the environment $\underline{X}$, a function W of the start configuration $\underline{S}$ and a function K of the target or goal configuration $\underline{T}$:

$$u^{n-1}(\underline{X}, \underline{S}, \underline{T}) = \sum_{i=1}^{n-1} R_i(\underline{X}) \cdot W_i(\underline{S}) \cdot K_i(\underline{T}) \tag{2}$$

and where the enrichment step is given by

$$u^n = u^{n-1} + R(\underline{X}) \cdot W(\underline{S}) \cdot K(\underline{T}). \tag{3}$$

3 PGD-Vademecum for Path Planning in a dynamic environment

In some practical situations, it is not enough that the target modifies its position since the environment could, spontaneously, change. With the previous formulation, but modifying the spatial domain $\Omega_{\underline{X}}$, considering now the new position of the obstacle. However, an obstacle could be seen as a region of the space towards which the vehicle must not to go. Mathematically this can be also obtained by modifying the properties of the initial domain $\Omega_{\underline{X}}$, i.e. defining the flux as $-K(\underline{X})\nabla u(\underline{X})$. Higher values of $K(\underline{X})$ will imply attraction of the vehicle while smaller values of $K(\underline{X})$ will provoke a repulsion to the vehicle. Note that in previous sections we consider $K(\underline{X}) \equiv 1$.

Using this formulation, a different set of obstacles can be modelled by a different definition of function K. Without losing in generality, let assume that all possible obstacle configurations can be modelled by a single parameter $p \in \Omega_p$ (one can always add more parameters to define more complex obstacles since the PGD permits to solve high dimensional problems easily). Therefore function

$$K\left(\underline{X},p\right) = \sum_{i=1}^{M} K^x(\underline{X})K^p(p) \tag{4}$$

models all possible obstacle configurations. Then the Laplace problem can be rewritten as follows:

$$\nabla \cdot (-K(\underline{X},p)\nabla u) = f \tag{5}$$

Then, the use of the PGD technology for solving equation 5 produces a solution as follows

$$u\left(\underline{X},p\right) = \sum_{i=1}^{N} X_i\left(\underline{X}\right)P_i(p) \tag{6}$$

This solution represents the potential field for any position of the space and for any position of the obstacle. Thus, the new path planning will only require to post process this solution when the obstacle configuration changes.

References

[1] Chinesta F, Leygue A, Bordeu F, Aguado JV, Cueto E, Gonzalez D, Alfaro I, Ammar A and Huerta A (2013) *PGD-Based computational vademecum for efficient Design, Optimization and Control* Archives of Computational Methods in Engineering. Vol 20, Is 1, pp 31-49.

[2] Chinesta F, keunings R and Leygue A (2014). *The Proper Generalized Decomposition for Advanced Numerical Simulations. A primer* Springer Briefs in Applied Science and Technology.

[3] Falcó A and Nouy A (2012) *Proper Generalized Decomposition for Nonlinear Convex Problems in Tensor Banach Spaces* Numerische Mathematik 121(3): 503-530,.

[4] A.Falcó, N. Montés, F. Chinesta, L. Hilario, M.C. Mora *On the Existence of a Progressive Variational Vademecum based on the Proper Generalized Decomposition for a Class of Elliptic Parameterized Problems*, Journal of Computational and Applied Mathematics, 2018, vol.330, 1093-1107

[5] Garrido S, Moreno L, Blanco D and Martín Monar F (2010) *Robotic Motion Using Harmonic Functions and Finite Elements* J Intell Robot Syst, 59:57-73.

[6] Khatib O (1986) *Real-time obstacle avoidance for manipulators and mobile robots* Int. J. Robot Res. 5(1): 90-98.

[7] Kim J and Khosla P (1992) *Real-time obstacle avoidance using harmonic potencial functions* IEEE Trans. Robotics and Automation, 8(3): 338-349.

[8] Montés N, Chinesta F, Falcó A, Mora MC, Hilario L (Under review) *A PGD based Method for Global Path Planning. A primer* IEEE, International Conference on Robots and Systems 2018.

[9] Montés N, Chinesta F, Falcó A, Mora MC, Hilario L, Rosillo N (Under review) *A Proper Generalized Decomposition-based framework for Potential-guided Robot Path Planning.* International Journal of Robotics Research.

[10] Montés N, Chinesta F, Falcó A, Mora MC, Hilario L, Rosillo N (2017) *Applications for Proper Generalized Decomposition method in motion planning robotics systems* Workshop on Reduced Basis, POD and PGD Model Reduction Techniques, 2017.

[11] Montés N, Chinesta F, Falcó A, Mora MC, Hilario L, Rosillo N (2017) *Youtube Video link.*

[12] Reif JH (1976) *Complexity of the mover's problem and generalizations.* IEEE Symp. Found. Comput. Sci., pp. 421-427.

[13] Rimon E and Koditschek D (1992) *Exact robot navigation using artificial potential functions* IEEE Transactions on Robotics and Automation 8(5): 501-518.

Modeling the rise of the Precariat in Spain

Elena de la Poza-Plaza[1*], Acxel E. Fernández[2], Lucas Jódar[3], Paloma Merello[4]

[1] Centro de Ingeniería Económica, Universitat Politècnica de València, 46022 Valencia, Spain.

[2,3] Instituto de Matemática Multidisciplinar, Universitat Politècnica de València, 46022, Valencia, Spain.

[4] Department of Accounting, University of Valencia, 46071, Valencia, Spain.

*e-mail[1] : elpopla@esp.upv.es

1. Introduction

In 2017, the decrease of the Spanish level of unemployment announced the end of the crisis. However, despite the positive figures of the Spanish GDP indicator, the risk of poverty has not stopped increasing since 2007. Thus, a new phenomenon is emerging, the flattering of the Spanish middle class produced by the so-called "gig economy" characterized as a race to the bottom in wages and labour rights bringing the impoverishment of workers (Barbieri, 2009).

In addition, the creation of new contracts produced by the economic recovery lacks not only the proper economic conditions, but also missing suitable conditions. As a result, the new contracts are mainly temporary/ involuntary part-time or false self-employed rather than permanent contracts. A new concept emerges fiercely in western countries, the precariat (Standing, 2012). In the Spanish context, the precariat population embraces those individuals who lack of wages higher than 1,250 euros/month and cannot live autonomously in dignified living conditions (Nachtwey, 2017).

The importance of the problem is evident and is in force; the crisis has led to the destruction of the Spanish middle class, due to the increasing divergence of salaries, traditionally fixed in Spain, and the consequent imbalance in the welfare of society, accelerated by the process of robotization, and the emergence of the digital economy, that is to say, the massive destruction of jobs and the creation of others, but to a lesser extent that demand a high qualification.

Compared to previous studies in which the poverty threshold in Spain is analyzed from information derived from surveys and statistical inference (Eurostat, 2016; Felgueroso *et al*., 2018) and which offers only a partial and fixed image of the Spanish social reality, this work proposes a model that allows quantifying the population at risk of precariousness in Spain with a dynamic approach. Thus, we build a compartmental mathematical model to identify and quantify the size of the Spanish population at risk to become precariat in the period of time 2017-2021 (De la Poza y Jódar, 2018, De la Poza *et al*., 2016).

The drivers taken into account to build the model are economic, demographic, substance abuse, socio-legal, psychological and technological. The proposed model

is built for the particular case of Spain but it is applicable to any European country when data is available. The relevance of this study relies on reporting the problem to public authorities responsible for addressing policies to stop this trend (Rifkin, 2011).

2. Model

The dynamic population model (Haddad *et al*., 2002; Goldthorpe, 2016) quantifies the amount of people from 18 years old who lives in precarious condition in Spain.

Thus, the Spanish population is classified into 6 categories according to their level of precariat:

$Z(n)$: zero risk subpopulation. It represents the population over 18 years old that is idle (lives on income) or is retired with sufficient income in the n-th semester after December 2017, ($n = 0$). Z is assumed to remain constant for the period of study (Spanish Statistics Institute, 2016).

$PL(n)$: professional training students and university students (Bachelor and Master students) who are older than 18 years old.; PL also includes those who will embrace the job market after graduating or hold part-time jobs, (in Spain 30% part-time workers, OECD, 2018).

$E(n)$: entrepreneurs.

$FNP(n)$: fixed employed people including civil servants, public employees and employees of large corporations with a level of monthly gross amount of incomes higher than 1,250 Euros, (Spanish Statistics Institute, 2016).

$P(n)$: precarious people. Those who are false self-employed workers, non-registered household workers, all types of temporary and part-time employees (excluding those who are students at the same time, 30%, OECD, 2018), retired people with low pensions and long-term unemployed.

$HIM(n)$: marginalized subpopulation lacking of social and economic integration, such as gypsies, refugees, undocumented immigrants, convicted prisoners (Spanish Statistics Institute, 2016).

The individuals transit to lower or higher levels of precariousness by the conjunction of factors; those factors are explained by economic, legal, socio-demographic and psychological causes.

Starting by the retirement transit (α_r); $E(n)$ and $FNP(n)$ individuals transit to $Z(n+1)$ when getting retired perceiving convenient pension incomes; however, 85% of $E(n)$ and 100% $P(n)$ individuals transit to $P(n+1)$ when retire.

The next transit is explained by the economy; according to the IMF and Funcas forecast, the economic growth will create 200,000 new jobs per year during the period of study. $PL(n)$ individuals transit to $P(n+1)$ due to the economic effect (β_e), but also a small proportion of $PL(n)$ achieves a permanent contract (with compensations higher to 1,250 euros) becoming $FNP(n+1)$. This transit is assumed

constant due to the Spanish high level of unemployment for the period of study. Linked to the economy, it is the emergence of start-ups and other companies explaining the entrepreneurship transit generating a shift from PL(n) and P(n) to E(n+1), (β_{en}).

Regarding the demographic transit, the variables considered are the birth rate, the death rate, the immigration and emigration assumed constant for the period of study. As a result, the net emigration result amounted to 44,563 individuals by semester in 2017, which is split into the six categories (Spanish Statistics Institute). Also, each semester 30,000 HIM individuals regularize their labor status, transiting to P(n+1), (β_{rgb}).

Next it is the adulthood transit (β_a), which explains the PL(n) individuals transit to the rest of categories when becoming 26 years old.
Related to the robotization process, its effect is double. It destroys jobs, so FNP(n) transits to P(n+1) (Cameron, 2018), but also fosters the creation of qualified jobs, promoting the transit from PL(n) to FNP(n+1).

In addition, the emotional status of individuals impacts negatively on their precariousness; thus, P(n) individuals transit to HIM (n+1). This transit is modeled through the combination of three factors: long-term unemployed with emotional stress and abuse of drugs and alcohol.
Finally, it is considered that the Government will rise the increase of the minimum wages per month to 1,000 euros; as a result, a proportion of the permanent contracts (assumed as P(n) individuals) will transit to FNP(n+1), (α_{rl}).

Following, the compartment dynamic model to quantify the precarious population is expressed:

$$Z(n + 1) - Z(n) = \alpha_{rE}(n) \cdot E(n) + \alpha_{rFNP}(n)FNP(n) + b_{dZ} + b_{rlZ}(n)$$

$$\begin{aligned} PL(n + 1) - PL(n) &= -(\beta_{enE}(n) + \beta_{eP}(n) + \beta_{eFNP}(n) + \beta_{aE} + \beta_{aFNP} + \beta_{aP} \\ &\quad + \beta_{rbFNP}(n)) \cdot PL(n) + b_{dPL} + b_{rlPL}(n) \end{aligned}$$

$$\begin{aligned} E(n + 1) - E(n) &= -(\alpha_{rE}(n) + \alpha_{rP}(n)) \cdot E(n) + \beta_{enE}(n) \cdot P(n) + (\beta_{enE}(n) + \beta_{aE}) \\ &\quad \cdot PL(N) + b_{dPL} + b_{rlE}(n) \end{aligned}$$

$$\begin{aligned} FNP(n + 1) - FNP(n) &= -(\beta_{rbFNP}(n) + \alpha_{rFNP}(n)) \cdot FNP(n) \\ &\quad + (\beta_{eFNP}(n) + \beta_{aFNP} + \beta_{rbFNP}) \cdot PL(n) \quad + b_{dFNP} + b_{rlFNP}(n) \end{aligned}$$

$$\begin{aligned} P(n + 1) - P(n) &= \beta_{rbFNP}(n) \cdot FNP(n) + \alpha_{rP}(n) \cdot E(n) + (\beta_{aP}(n) \\ &\quad + \beta_{aE} + \beta_{aFNP}) \cdot PL(N) + \beta_{rgbHIM}(n) \cdot HIM(n) \\ &\quad - (\alpha_{rlFNP}(n) + \alpha_{emHIM} + \beta_{enE}(n)) \cdot P(n) + b_{dP} + b_{rlP}(n) \end{aligned}$$

$$\begin{aligned} HIM(n + 1) - HIM(n) &= -\beta_{rgbHIM}(n) \cdot HIM(n) + \alpha_{emHIM}(n) \cdot P(n) + b_{dHIM} \\ &\quad + b_{rlHIM}(n) \end{aligned}$$

3. Results

By computing the model, the subpopulation values are estimated by semester. Table 1 shows the results at the beginning of the study, n=0 (December 2017) and at the end, n=8 (December 2018).

Table 1. Subpopulations forecast at n=0 and n=8.

	Semester	Z	PL	E	FNP	P	HIM
Dec 2017	0	7,473,115	1,869,121	3,231,279	12,096,000	14,938,735	1,922,065
Dec 2021	8	7,317,440	1,233,312	3,450,388	11,948,933	16,455,709	1,920,113

Results show how the precarious subpopulation grows for the 4 years period of study, representing close to the 40% of the Spanish population in 2021. Conversely to the economic recovery, the precarious do increase over time. On contrast, the pre-labor category decreases.

4. Conclusions

The study shows the deterioration of the Spanish population living standards despite the improvement experienced by the macro indicators. The model estimates evidence of the destruction of the Spanish middle class and the impoverishment of society.
The stop to this social problem requires the policy makers' action to incentive permanent contracts with suitable economic compensations but also new formulas to link the employees' wages to the firm performance.

References

[1] Barbieri, P. (2009). Flexible Employment and Inequality in Europe, European Sociological Review, Volume 25, Issue 6, 1, pp. 621–628.

[2] Cameron, E. Will robots really steal our jobs?, Price Waterhouse Coopers, 2018. www.pwc.co.uk/economics

[3] De la Poza, E., L. Jódar. (2018). A Short-Term Population Model of the Suicide Risk: The Case of Spain. Cult Med Psychiatry https://doi.org/10.1007/s11013-018-9589-4

[4] De la Poza, E., L. Jódar, and S. Barreda. (2016). Mathematical Modeling of Hidden Intimate Partner Violence in Spain: A Quantitative and Qualitative Approach. Abstract and Applied Analysis. https://doi.org/10.1155/2016/8372493 .Eurostat,

Income Inequality in the EU, 2016.
http://ec.europa.eu/eurostat/web/products-eurostat-news/-/EDN-20180426-1

[5] Felgueroso, F., Millán, A., Torres, M. Población Especialmente Vulnerable ante el Empleo en España. Cuantificación y Caracterización. Estudios sobre la Economía Española, 2017, FEDEA, Madrid, 2018. http://documentos.fedea.net/pubs/eee/eee2017-07.pdf Funcas, https://www.funcas.es/Indicadores/Indicadores.aspx?Id=1

[6] Goldthorpe, J.H. (2016). Sociology as a Population Science. Cambridge: Cambridge Univ. Press.

[7] Haddad, W. M., Chellaboina, V. and Nersesov, S. G. (2002). Hybrid nonnegative and compartmental dynamical systems, *Math. Probl. Eng*., vol. 8, no. 6, pp. 493–

[8] International Monetary Fund, www.imf.com

[9] Nachtwey, O. *La sociedad del descenso*. Madrid: Paidos, 2017.

[10] OECD Labor force Statistics (2018). Statistics [online]. Available at: http://www.oecd-ilibrary.org/employment/oecd-labour-force-statistics_23083387

[11] Spanish Institute of Statistics www.ine.es

[12] Standing, G. *The Precariat. The New Dangerous Class.* London: Bloomsbury, 2012.

[13] Rifkin, J. *The Zero Marginal Cost Society*. New York: Palgrave Macmillan, 2011.