

A generalizable framework for spatially explicit exploration of soil carbon sequestration on global marginal land

Ariane Albers (✉ albers.ariane@gmail.com)

TBI, Université de Toulouse, CNRS, INRAE, INSA

Angel Avadí

CIRAD, UPR Recyclage et risque

Lorie Hamelin

TBI, Université de Toulouse, CNRS, INRAE, INSA

Research Article

Keywords: climatic and edaphic, carbon sequestration, pedo-climatic constraints, plant species tolerances, soil organic carbon (SOC)

Posted Date: July 15th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-701807/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

A generalizable framework for spatially explicit exploration of soil carbon sequestration on global marginal land

Ariane Albers^{1,*}, Angel Avadí^{2,3}, Lorie Hamelin¹

¹ TBI, Université de Toulouse, CNRS, INRAE, INSA, Toulouse, France

² CIRAD, UPR Recyclage et risque, F-34398 Montpellier, France

³ Univ Montpellier, CIRAD, Montpellier, France

*Corresponding author: albers.ariane@gmail.com

Abstract

Designing land-based climate stabilisation strategies demands changes in land use management or new areas to grow biomass sustainably, without reducing food supply or damaging natural ecosystems. Here we integrated, into a generalizable framework, the mapping of a set of biophysical (climatic and edaphic) and land conservation constraints to identify the pairing of target areas with suitable plant species, and quantify their contributions to long-term soil carbon sequestration, including losses from rain-driven erosion. The proposed framework represents a refinement to previous global mapping exercises, which seldom consider pedo-climatic constraints, plant species tolerances and world protected areas. We demonstrated the applicability of the framework through a global case study by mapping target areas featuring low soil organic carbon stocks (≤ 50 t SOC ha⁻¹) and producing quantifiable soil carbon sequestration potentials of best matches. Preliminary target areas were mapped at a 30 arc-sec resolution, and then consolidated per geographical and environmental boundaries (geopolitical world regions x global ecological zones).

Introduction

The global top 2 m of soil feature a soil organic carbon (SOC) debt of 133 Gt¹, which has increased in the past two centuries due to increasing agricultural land use. This debt (i.e. the difference between the original pre-agriculture and the current amounts of SOC in exploited soils) demonstrates a strong correlation between land degradation and SOC losses, dependent on the degree of intensity and duration of soil exploitation. It has been recognised that the regions featuring historic SOC losses can now be considered as SOC sinks^{2,3}, with a potential to offset two thirds of today's SOC debt in the near-term¹.

The *4 per 1000* initiative (www.4p1000.org), launched at the COP21 Paris Climate Summit in 2015, is a multi-stakeholder international initiative aiming at increasing SOC sequestration through sustainable land management⁴. It is based on the premise that an annual mean SOC increase of 0.4% in the global agricultural topsoil (30 cm) would contribute to an annual global sequestration of 2.5 Gt C⁵ —an estimation subject to criticism, due to intrinsic data and model uncertainty⁶, yet representing a clear goal towards climate stabilisation. Soil carbon sequestration (SCS) thus constitutes a key element of negative emission technologies and practices that result in the net removal of greenhouse gases from the atmosphere⁷. SCS is increasingly seen as a key land-based mechanism relying on plant photosynthesis to stabilise the climate⁸, constituting both a mitigation measure (for climate change but also for ecosystems quality) and a way to induce additional negative emissions, as required by the Paris Agreement, to limit warming below 2°C⁹. Biophysical potentials for SCS have been variously estimated at 0.4-8.6 Gt CO₂ yr⁻¹¹⁰, while its technical feasibility may range as much as between 1-5 Gt CO₂ yr⁻¹¹¹ —reliant on local land use type (e.g. cropland), management (including external inputs) and pedoclimatic conditions, among other factors. Moreover, exploited soils featuring SOC stocks <30 t C ha⁻¹ are expected to be able to attain high SCS after introducing specific management practices⁵.

The bulk of agriculture's contribution to SCS is influenced by land use management (e.g. addition of organic matter via organic amendments and fertilisers, deployment of improved crop rotations and cover crops),

including the cultivation of specific crop types such as perennial and deep rooting species^{12,13}. Perennial species have received special attention, as they generally require less soil work, enlarge the C fraction in the soil micro-aggregates, and increase belowground C allocation¹⁴. In contrast to annually harvested crops, species with long life spans accumulate C over several years to decades in above- and belowground biomass, representing an organic C stock⁸. The extent to which crops in general contribute to SCS is subject to extensive research¹⁵.

The cultivation of dedicated biomass enabling SCS and eventually providing feedstock for various economic pathways is intrinsically connected with land demand, and thus with the availability of new areas to grow plants sustainably; i.e. with no negative effects (e.g. on the carbon debt, on food security, on ecosystem services, on biodiversity quality)¹⁶. In the past decade, research has focused on identifying land unsuitable for food production but potentially suitable for non-food crops (e.g. bioenergy, biomaterials), defined as **marginal land**¹⁷. Marginal land could, depending on its capacity to sustain biomass, be used for biomass cultivation aiming at either increasing SCS or providing feedstock for the bioeconomy. The expected benefits of exploiting marginal land are wide, ranging from soil quality improvements (e.g. soil fertility, soil structural stability)¹⁸ to carbon sequestration (e.g. 30-50% of C in cropland mineral soils has been lost due to degradation^{11,13}), through biodiversity conservation¹⁹ and eventual socio-economic development (e.g. employment, infrastructure, rural poverty alleviation)²⁰.

Key types of land classifiable as marginal include degraded and abandoned agricultural lands. **Degraded land**, which may include soils naturally characterised by low productivity (e.g. natural high salinity soils, or heathlands such as the Mediterranean *garrigue*), has received considerable attention, as a key constituency of marginal land²¹. In particular, research led by the International Soil Reference and Information Centre (ISRIC), throughout various projects such as GLASOD²², GLADA²³ and LADA²⁴, applied the use of a remotely sensed global normalised difference vegetation index (NDVI) as a proxy for land degradation due to different causes²⁵. Among the dominant typologies of degraded land, the following FAO classification — referred to as the FAO agricultural problem land approach²⁶, associated with the FAO/UNESCO Digital Soil Map of the World²⁷, now integrated into and superseded by the Harmonized World Soils Database (HWSD)²⁸ and into FAO's Global Agro-Ecological Zones (GAEZ)²⁹—, represents a synthesis of criteria: too cold (polar/boreal), alluvial soil in deserts, too dry, steep lands (dominant slope >30%), shallow lands, poorly drained, coarse texture, vertisols, infertile (e.g. nutrient-poor), saline/sodic, acid sulphate, and peats (organic soils). For **abandoned agricultural land**, another key constituency of marginal land, extensive research has proposed different definitions and mapping approaches³⁰⁻³². Agricultural land abandonment has been considered to be mainly driven by biophysical constraints, but also by reasons pertaining to farm structure, agricultural viability, as well as to changing population, political regimes, nature conservation and other regional contexts^{31,33}. These approaches generically consist of comparing satellite data corresponding to two different periods, and interpreting the land cover differences.

Several studies attempted to quantify and map marginal land use across spatial (global, national, regional and local data) and temporal (historic to current data) resolutions³⁴. At the global scale, it has been quantified by mapping current land cover and land suitability indices³⁵. However, it has not always been possible to identify the relative importance of agricultural (i.e. abandoned) and non-agricultural land types on the identification of marginal land, because of a combination of factors: analyses based exclusively on (bio)economic drivers, quality of underlying datasets and scales, and lack of detail on current land covers' spatial (i.e. granularity) and temporal resolutions¹⁶. Beyond the challenge of defining and identifying marginal lands, the design of SCS strategies involving biomass demands the correct pairing of suitable plant species with specific target areas, based on the compatibility of plant species with the target areas' biophysical characteristics. Such endeavour is not negligible, as demonstrated by the body of research evidence, as recently reviewed at the global scale¹⁶.

A recently proposed strategy to facilitate global scale soil climate mitigation ¹¹, suggests that detailed maps of C sequestration potentials including erosion, associated with simple analytical tools, would contribute to supporting the global implementation of SCS. One of the first examples of such an approach for quantifying global SOC dynamics is that by Morais et al. ³⁶, which was applied to spatially differentiated unique homogenous territorial units characterised in three main land use classes, namely cropland, grassland and forest land. Following the hypothesis that global soils with low initial SOC stocks feature high SCS potentials, we chose to explore and map global SOC-deficient marginal lands, featuring up to 50 t SOC ha⁻¹ at the topsoil (≤ 30 cm), equivalent to 1.2% C in soil, to test if the hypothesis applies to this specific condition. The framework proposed here thus complements that of Morais et al. (which focused on non-marginal land) and other attempted approaches focusing at least partially on marginal land for biomass production as bioenergy feedstock, usually at a regional scale ^{37,38}. This study had thus a double purpose: i) to propose a coherent generalizable framework to facilitate the mapping of marginal lands with low SOC stocks (≤ 50 t SOC ha⁻¹) at global scale and their matching with candidate plant species, and ii) to illustrate the feasibility and interest of the proposed framework by means of a proof-of-concept implementation producing a quantification of SCS resulting from suitable matches.

Results

Conceptual framework

The proposed framework allows quantifying global SCS potentials of biomass cultivation on target areas by accounting for both geographic- and plant species-dependent climatic and edaphic limitations. We combined and intersected georeferenced products and automatized the tasks of matching pre-selected “**biopumps**” (i.e. plant species featuring SCS enhancing capabilities and representing a potential source of feedstock for the bioeconomy) to target areas (i.e. aggregations of specific areas of marginal land, based on ecological zoning). Such matching was based on the pedoclimatic tolerances of the former to the prevailing conditions of the latter, and was followed by a computation of soil C sequestration and C losses due to erosion by water over the simulation time horizon 2020-2100, using well-established models. A R code combined with the data treatment strategy depicted in Figure 4, constitutes the proposed generalizable framework. The framework is able to produce estimations showing whether more soil C would be sequestered, in the long-term, than it would be lost to degradation and rain-driven erosion, if the matching plant species were systematically grown on the identified areas. To our knowledge, no previous global study on SCS considered such careful identification of marginal land (i.e. integrating pedoclimatic constraints, plant species tolerances, and land conservation), while balancing SOC sequestration and losses.

Global marginal lands and target areas

To define the extent of marginal land, elements from various studies ^{16,37,38} were combined to define marginal lands as land covers that are currently unused or underutilised by agriculture due to an aggregation of socio-economic and biophysical constraints, or human-induced land degradation, but which could potentially be suitable for sustainable biomass production. These areas are presented in Supplementary Results (Figure S1) with SOC stocks divided into five classes, following increments of 10 t SOC ha⁻¹. Identified marginal lands (in total 2 714 Mha), consist of bare areas (74.4%), sparsely vegetated (25.4%) and abandoned agricultural land (0.14%). Additional intermediate maps are also presented in Supplementary Results, depicting marginal land (Figure S2 to S6) and the criteria leading to the identification of target areas (Figure S7 to S10) under the retained constraints (e.g. protected areas, pedoclimatic constraints, global climate zones, and world regions).

Figure 1 illustrates the resulting consolidated target areas, defined by geographies established as the combination of climate zones defined in FAO’s Global Ecological Zones (GEZ) ³⁹ and geo-political world regions. Overall, 28 Mha of land were identified as target areas based on 152 possible combinations of 25

world regions and 19 GEZ (excluding water and polar). The target areas represent 1% of the preliminary global marginal lands due to biophysical (e.g. plant growth requirements) and land conservation limitations.

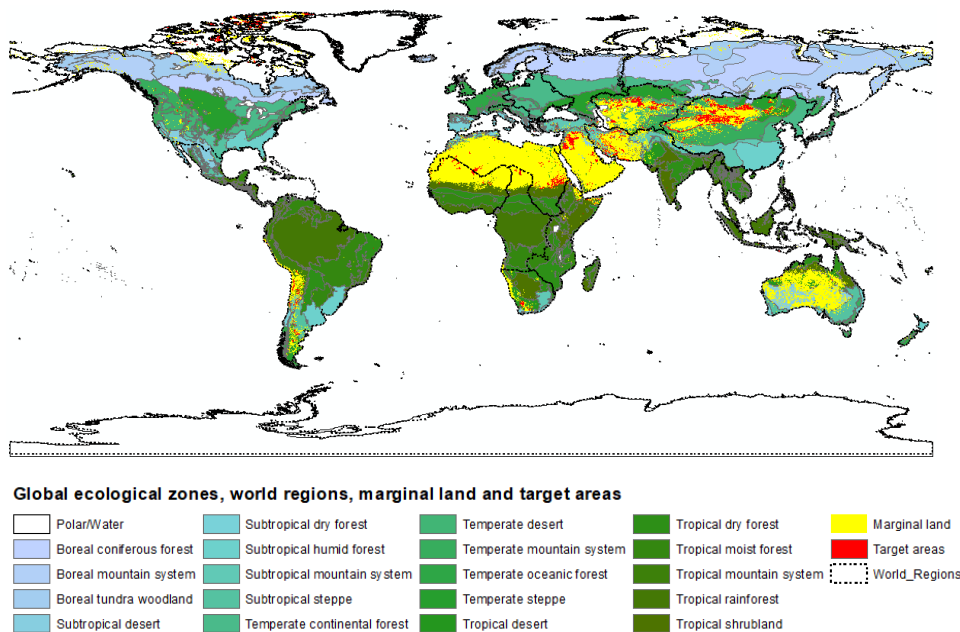


Figure 1. Identified target areas on marginal lands featuring up to 50 t SOC ha⁻¹ (at 30 cm topsoil) per global ecological zone and world region

The majority of target areas were found in Asia (Eastern, Central, Southern, and Western), Northern Africa, and South America (Figure 2), corresponding to GEZ in the temperate, tropical and subtropical desert and mountain systems, as well as temperate and subtropical steppe. Europe accounted for 0.1% of all target areas, largely present in the Southern Europe Region concordant with GEZ Subtropical dry forest. All values per world region and GEZ are listed in the Supplementary Results Table S1.

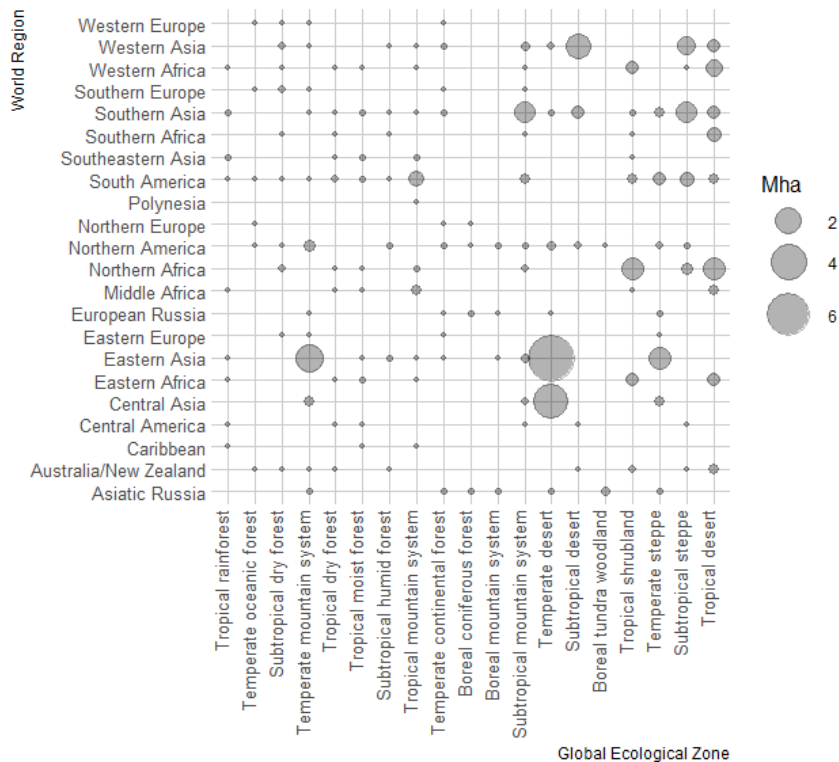


Figure 2. Relative geographical concentration of identified target areas

Suitable “biopumps” with soil carbon sequestration potentials

Initially, 432 species were identified in the FAO’s ECOCROP database⁴⁰, and their environmental requirements were determined to further assess the suitability to grow on the identified target areas. The eventual introduction of invasive species in the receiving ecosystems was not considered beyond the exclusion of world protected areas. Then, 50 biopumps were preselected by scoring and ranking SCS enhancing capabilities, primary yield productivity, and marginal land adaptability. SCS potentials are provided in Supplementary Results (Table S1 and Figure S10), including a ranking and scoring of biopumps in Table S2.

The matching exercise resulted in 460 viable combinations of target areas x biopumps (out of 65 664 theoretically possible matches). Overall, 128 combinations (28% of all matches) featured net SOC stocks >0 at the year 2100 (i.e. the sequestered C is larger than the eroded C), where 17 biopumps (associated to 67 species) were compatible within 12 world regions and 11 GEZ (Figure 3). The total area of target area-biopump pairs SOC stocks >0 comprised 0.56 Mha, representing about 2% of the initially identified target areas (28 Mha in total). The largest concentration in area was found in Western Africa (0.30 Mha), Australia/New Zealand (0.21 Mha) and South America (0.03 Mha), corresponding to the GEZ Tropical shrubland, Tropical desert and Tropical dry forest. Top biopumps with the highest SCS potentials, *prior to* the consideration of erosion, were hemp (*C. sativa spp. Indica*), cup plant (*S. perfoliatum*), neem (*M. azedarach var. australasica*), acacia (*Acacia karoo and Acacia nilotica*), and miscanthus (*spp.*) with 87, 54, 51, 18, 9 t C ha⁻¹, respectively. The biopumps with the highest number of pairings per region and/or GEZ, were acacia, eucalyptus, switchgrass, casuarina and cup plant.

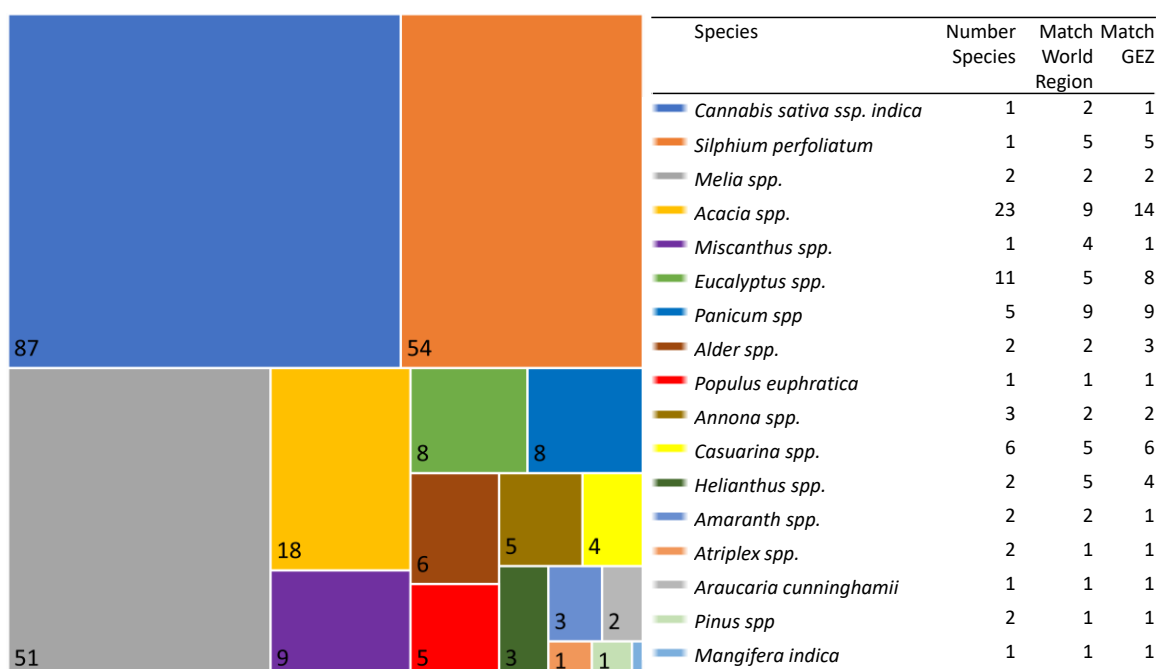


Figure 3. Top biopump soil carbon sequestration potentials (number in boxes in t C ha⁻¹) and preliminary compatibility with corresponding number of species and target areas per world regions and global ecological zones (GEZ)

Target areas-biopumps pairs with the highest sequestration potentials

From the previously identified 128 possible target area-biopump pairs with SOC >0, 6.5% (i.e. 30 matches) represent the selection of only one biopump per target area. Overall, 11 biopumps (associated with 12 species) were identified to yield the highest net SCS results at year 2100 (referred to as “best-cases”), whereas the most representative in number of matches were cup plant (*S. perfoliatum*) followed by

switchgrass (*P. maximum*) and acacia (*A. erioloba*). Data of the 30 best-case combinations are shown in the Supplementary Results Table S2.

A comparison of SOC stocks at the baseline year 2020 (referred to as “initial SOC”) with the final state of the simulation at year 2100 (referred to as “final SOC”) has shown, for these best-case combinations, that the observed SOC level improvements compared to initial SOC are significant (p-value<1%). Overall net SCS potentials ranged between 0.51-33.57 t C ha⁻¹ (equivalent to about 0.01-0.42 t C ha⁻¹·yr, when linearly annualised). Top 5 best-cases were Northern Africa x Tropical dry forest x *M. azedach*, Australia/New Zealand x Tropical dry forest x *M. azedarach var australasica*, Central America x Tropical dry forest x *C. sativa ssp. indica*, Eastern Africa x Tropical dry forest x *C. sativa ssp. indica*, and European Russia x Temperate mountain system x *S. perfoliatum*, with net SOC at 33.57, 31.07, 23.35, 22.88, 13.47 t ha⁻¹ representing 79%, 61%, 65%, 45% and 100% of final SOC *prior to* erosion, respectively (Table S2). There were two best-case exceptions, namely the matches: Western Europe x Temperate continental forest x *S. perfoliatum* and Western Africa x Tropical dry forest x *H. annuus*. These two situations, despite reaching a positive net SOC, corresponded to a reduction of the initial stock by 37% and 12%, respectively (Table S2), indicating that the implementation of biopumps would generate a net loss of SOC.

More generally, spatially dependent soil erosion affected the net SOC potentials significantly (p-value<1%). Long-term improvements in final SOC were not achieved for most originally identified target areas. The most affected target areas were Tropical moist forest in South and Central America, as well as Tropical dry forest in Western Africa, where the resulting net SOC represented respectively 1%, 15% and 4% of the final SOC (Table S2). We hypothesise that very strong seasonal precipitation, coupled with incomplete protection of the soil by canopy (expressed as the cover management factor in the erosion model), lead to these effects. From the resulting best target area x biopump pairs, one third of the results showed negligible soil C losses by erosion (up to 8% of final SOC values before erosion) located in European Russia (Boreal mountain system, Temperate mountain system and Boreal coniferous forest), Western Europe (Temperate mountain system and Temperate continental forest) and Northern Europe (Temperate continental forest).

Discussion

The framework design is generic enough to be implemented in other settings, regarding target area definition, plant species of interest, scopes (e.g. geographic boundaries), as well as temporal and spatial resolutions. The provided R script in ⁴¹ is usable in different situations as long as the required input data is organised in the prescribed way. Moreover, the framework is flexible to accommodate different SOC sequestration and erosion models, depending on the required outputs and data availability. For our example, we selected the RothC model, being an appropriate SOC model for the following reasons: (i) its demonstrated performance to replicate observed SOC changes in validation experiments, as compared with other models (described in Supplementary Methods), (ii) its applicability to a wide range of world climates and regions in combination with GIS products ^{3,36}; (iii) its recognition, being recommended as a standard spatialized SOC model at a 30 arcsec resolution by the FAO ⁴².

As a proof-of-concept we applied the framework on global marginal lands under well-defined criteria (e.g. soil constraints, land conservation and low initial SOC stocks) and were able to identify the plant species delivering highest net SCS —showing the feasibility and interest of the proposed framework to produce immediately useful results. The climatic and edaphic characteristics of the target areas were averaged at the world region x GEZ scale. Despite the loss of detail, we considered that the consolidated target areas were the more suitable geographical resolution to explore biopump compatibility at a manageable scale. Furthermore, the projected long-term SOC changes were based on near-term climate data, i.e. disregarding climate change effects on SOC stocks dependent on temperature, precipitation and evapotranspiration

variables. Both the uncertainties associated to the consolidation of target areas and future climate trajectories (e.g. as defined in the 6th Climate Model Intercomparison Project, CMIP6⁴³) are addressed in an upcoming publication by our team.

We identified abandoned agricultural land, as a key component of marginal land, of about 3.76 Mha for SOC stocks $\leq 50 \text{ t C ha}^{-1}$ over the years 2010-2018 (about 0.47 Mha yr^{-1}) based on ESA CCI⁴⁴. A global historic land use comparison over the period 1700-2000⁴⁵ determined permanent cropland conversions into other land covers of about 210 (about 0.7 Mha yr^{-1}) and 269 Mha (0.9 Mha yr^{-1}) based on the datasets HYDEv3.0⁴⁶ and SAGE⁴⁷ respectively. A recent study³³ quantified the world's current extent of land use change and developed a model called HILDA+ to provide annual harmonised global land use data with national inventories at near-present. Cropland use changes were identified at 1029 Mha over the period 1960-2015 (about 0.19 Mha yr^{-1}) with HYDEv3.2⁴⁸ and 1035 Mha over the period 1960-2011 (about 0.20 Mha yr^{-1}) with SAGE⁴⁷. A comparison of the former and the later with HILDA+ over the same periods showed higher results by 31% and 24%, respectively. The annualised values are comparable among different studies with different temporal resolutions; however, it generates uncertainty in the interannual variability as well as in scope and definition of abandoned agricultural land vs. cropland use change.

A key hypothesis for the global case study selection was that only SOC-deficient land covers are interesting for biopump implantations, as they have not reached C saturation and are likely to attain higher sequestration potentials⁵. However, SOC with $<0.5\%$ and $<0.75\%$ organic carbon (30 cm topsoil) are considered, respectively, as featuring severe to sub-severe soil fertility deficiency⁴⁹, which is equivalent to about 21 and $31.5 \text{ t SOC ha}^{-1}$ respectively, as per the equation in Ledo et al.⁵⁰. We disregarded the (sub-)severe soil fertility constraints for this exercise by cross-referencing global land covers with global SOC stocks $\leq 50 \text{ t C ha}^{-1}$ (about 1.2 % organic carbon). This soil constraint might even further reduce the identified target areas. However, we have not included it because the residual biomass, as input to the soil, contributes to soil fertility and other management practices might further improve the soil quality. Any land use management (or change) influences the evolution of SOC pools, which need to be specified in the soil model. In our example, we have not considered additional management practices other than plant-based C inputs. RothC is a non-saturating SOC model⁵¹, yet deemed to yield accurate predictions⁵¹ in cases when C-inputs are "low"⁵² (e.g. in the absence of organic fertilisation). A further refinement in the framework would involve adding a C-saturating model to better account for management practices, including organic fertilisation^{51,53,54}.

Cultivation practices in terms of resource requirements (e.g. nutrient supply, irrigation), including best management practices (e.g. organic amendments and fertilisers, cover crops), should be considered in more detailed assessments, as they have been shown to influence SCS rates⁵. Other specific agronomic requirements (e.g. to prioritise only low-input species on marginal land), as well as ecological impacts and invasion risks (of so-called alien or invasive species), need further assessment.

The preselection of the biopumps was primarily based on soil C sequestration and feedstock (yield) potentials. Only about 17% of all inventoried plant species were suitable for the identified target areas, yet representing about 34% of the preselected biopumps. The preselection criteria alone are not a qualification for SOC improvements and biomass productivity. The study has demonstrated that the net sequestration is a result of a combination of biophysical factors (e.g. soil, terrain and climate types, rain erosion, etc.), and that the performance of one crop on one target area is not the same on another. Several other elements would require consideration, such as local environmental challenges (e.g. compaction and low water retention, limited nutrients, low organic matter and potential phytotoxicity, weed and pests, etc.), trade-offs with competing land uses (e.g. forage and livestock, housing, conservation, recreation, etc.), and socio-economic market interferences (e.g. disruption of value chains or people's livelihoods).

Methods

We developed a framework to explore SCS potentials by various plant species and categories, and applied it to identifying and consolidating global target areas delimited by geopolitical and environmental boundaries. The proposed framework relies on the use of georeferenced products, corresponding to the needs of macro-level global models. It was structured around four main steps (Figure 4):

- 1) identifying target areas corresponding to land covers of interest, while considering biophysical constraints to biomass production and land conservation (in our example: global marginal land);
- 2) characterising target areas by their pedoclimatic and terrain conditions, and consolidating them into the required level of aggregation/granularity (in our example: consolidated target areas defined by geopolitical world regions and geocological zones),
- 3) selecting plant species or groups of species (in our example: promising biopumps) and determining their environmental requirements/tolerances, and
- 4) matching target areas-biopump pairs as determined by the compatibility of plant species to target areas' biophysical characteristics, and modelling soil C flows of the resulting target areas-biopump pairs.

The full list of exploited data and sources of georeferenced data is presented in the dataset ⁴¹ and Supplementary Methods Table S1.

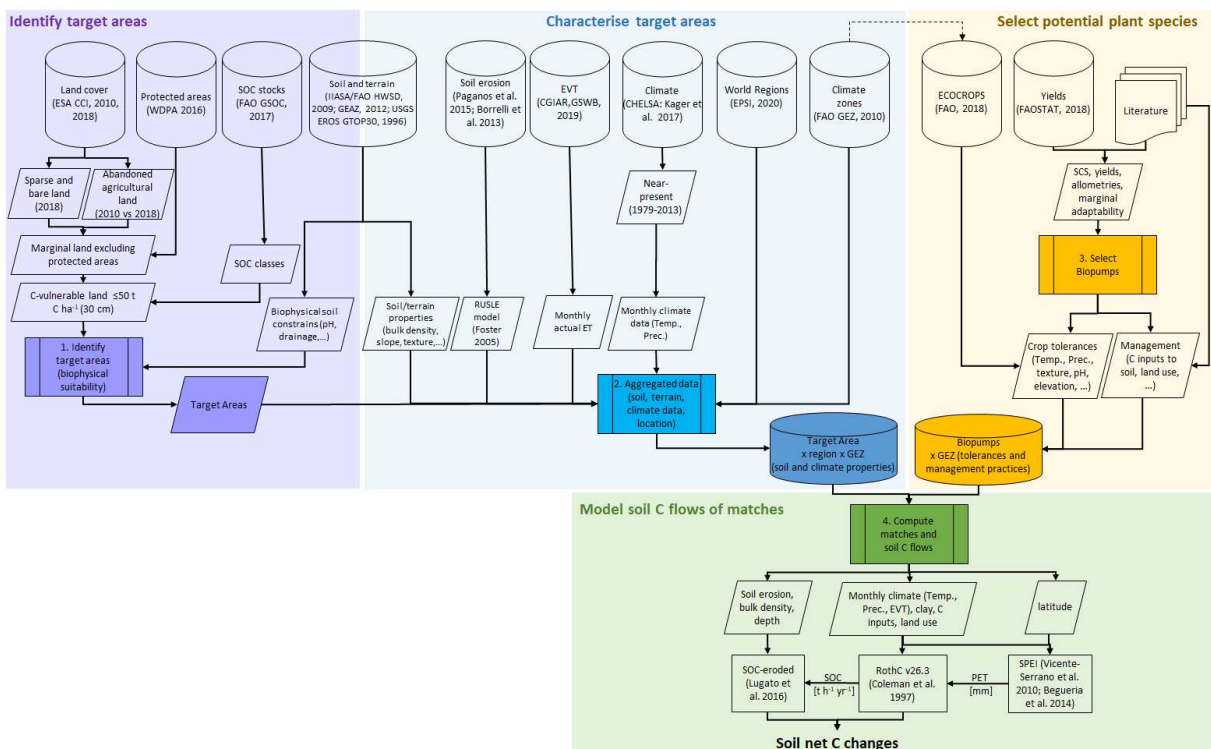


Figure 4. Stepwise framework and data sources for our case study. Abbreviations: Temp. = temperature, Prec. = precipitation; EVT = evapotranspiration; SPEI = standardised precipitation evapotranspiration index, PET = potential evapotranspiration, SOC = soil organic carbon

Identification of marginal land (steps 1)

Global marginal lands were mapped at a 30 arcsec (1 km) resolution, using a geographic information system (ArcGIS v10.6). We studied different definitions of marginal land to segregate land types into agricultural (potentially suitable for food production historically, currently or in future) and non-agricultural (unsuitable/unfavourable for food production). The selection was based on a combination of various criteria of soil constraints to biomass (plants-based) production from key marginal land studies (detailed in

Supplementary Methods, Table S2 and S3) to ensure that their associated biophysical constraints would be compatible with the environmental requirements for plant growth, and that chosen areas correspond to a land cover category that, if changed, would not contribute to further environmental degradation (including SOC losses). In our example, we followed five sub-steps to identify marginal lands:

In step 1.1, we explored the global land cover (LC) map by the European Space Agency Copernicus Climate Change Initiative (ESA-CCI) for the year 2018 at a 300 m (9.7 arcsec at the equator) resolution, which features 22 LC classes defined in the FAO Land Cover Classification System⁵⁵ (see Supplementary Methods, Table S4). We retained bare and sparsely vegetated (<15% of vegetation) areas, as all other LC classes likely feature food/feed production (e.g. irrigated or rainfed cropland, natural grasslands), densities of vegetation that would risk severe damage to natural ecosystems and SOC losses from land use change (e.g. forests and shrubland), or are vulnerable to crop growth due severe biophysical constraints (e.g. lichens and mosses).

In step 1.2, we identified recent abandoned agricultural land, understood as the share of marginal agricultural land that is unused or underutilised. The extent of abandonment was associated with land cover change from previous irrigated or rainfed cropland to other LC classes, by comparing ESA-CCI⁴⁴ georeferenced products from the years 2010 and 2018. We considered, elaborating upon the definition from Van Asselen and Verburg⁵⁶, a transition to mosaic cropland/natural vegetation (complemented with mosaic cropland/natural vegetation to semi-natural), grasslands, sparse vegetation, bare areas, mosaic herbaceous cover or shrubland.

In step 1.3, we refined the previous filter by analysing the global LC regarding SOC stocks [t SOC ha⁻¹] in topsoil (≤30 cm), based on the FAO's Global Soil Organic (GSOC v1.5) map⁵⁷ for the year 2017, and retaining areas featuring up to 50 t SOC ha⁻¹.

In step 1.4, we filtered out areas corresponding to geographically defined protected areas, particularly important for biodiversity and ecosystem services, covered in the World Database on Protected Areas (WDPA v1.6) map, a georeferenced product of the UN Environment Programme World Conservation Monitoring Centre (UNEP-WCMC)⁵⁸.

In step 1.5, we identified the biophysical constraints associated with the lands satisfying the previous two conditions, in order to discard areas featuring (sub-)severe soil and terrain restrictions that would exceed the tolerances of potential plant species. These constraints correspond to those associated with degraded lands in the georeferenced FAO Dominant Type of Problem Lands product for a generic perspective, complemented by the Harmonized World Soil Database (HWSD v1.21)²⁸ on soil properties.

Characterisation and consolidation of target areas (step 2)

The identified marginal lands were further characterised with their edaphic and climatic conditions for an iterative refinement in finding possible matches with biopumps' environmental tolerances, as well as to inform SOC models. In our example, we used the following georeferenced products:

- Soil properties (e.g. texture, clay content, pH, bulk density, depth) and terrain specification (slope and elevation) were obtained from HWSD v1.21²⁸ at a 30 arcsec resolution.
- Near present climate data (monthly temperature and precipitation) were retrieved from the Climatologies at High resolution for the Earth's Land Surface Areas (CHELSA v1.2)⁵⁹ at 30 arcsec for the years 1979 to 2013.
- Geographic boundaries were established with a Global Shapefile from the global administrative areas (GDAM v4) database⁶⁰. Layers for world regions (represented by the boundaries of 25 commonly recognised regions (Supplementary Methods, Table S5), as well as latitudes and longitude grids at 30 degrees, were retrieved as Esri ArcGIS Data & Maps products.

- Climate zones were classified as in FAO's Global Ecological Zones (GEZ)³⁹, consisting of 19 classes based on essential biological boundaries, i.e. with relatively homogeneous vegetation physiognomy (Supplementary Methods, Table S6).

Once all individual (contiguous pixels) marginal lands were characterised, it was necessary to consolidate them into larger target areas, according with the desired scope and granularity. In our example, we consolidated all marginal lands within the same GEZ (e.g. Tropical shrubland) and geo-political world region (e.g. West Africa), coarsely following⁶¹.

Identification and characterisation of plant species (step 3)

A preselection of potential biopumps was performed and the plant data recorded into a database to evaluate their suitability to grow on identified and consolidated target areas based on their pedoclimatic requirements, and to inform the SOC model (e.g. plant-based C inputs). Initially a list of 164 herbaceous and woody plants were considered, and data compiled from diverse data sources specific to agricultural perennial⁵⁰ and lignocellulosic bioenergy⁶² crops, including all 68 industrial crops initially considered in the EU H2020 MAGIC project (www.magic-h2020.eu/). The preselection was performed via a semi-quantitative multi-criteria analysis by scoring and ranking SCS performance, feedstock productivity, and marginal land suitability (detailed in in Supplementary Methods, Table S7).

Data on the pedoclimatic tolerances provided in⁴¹ for the matching were collected from FAO's ECOCROP database⁴⁰, per biopump, for the following variables: temperature [°C], precipitation [mm], pH, soil texture (coarse, medium, fine), altitude [m], and global climate zones. Other essential soil properties (e.g. soil drainage, rooting deep) were previously considered as limiting factors for determining target areas. The ECOCROP climate zone classification system based on Köppen⁶³ was harmonised with the FAO's GEZ⁶⁴ used for target areas consolidation, to enable the matching (harmonisation detailed in Supplementary Methods, Table S6). Moreover, several plant species with the same common name were recorded, showing different tolerances, to ensure a broader initial pool for the matching exercise, and because species specifications were lacking for some of the previously used datasets.

The land use class of biopumps was classified into three types based on the life form described in retained datasets^{40,50,65}: grass (e.g. bahiagrass, hemp, miscanthus, banana, bamboo, etc., including small woody or herbaceous shrubs (e.g. blueberry, etc.), crops (e.g. sugar cane, maize, loofah, etc.), and trees (e.g. short rotation coppice, orchards, tree-nuts, etc., as well as woody shrubs that are tree-like, e.g. due woody stem, lifetime, size and physiognomy of the plant). Uncertainties may have accumulated where the boundaries between grass and woody species are not clear.

Data provided in⁴¹ on plant C-inputs [$t\ C\ ha^{-1}\ yr^{-1}$] were computed by the fractioning and C partitioning approach⁶⁶, individually for above- (product, stem, and leaves) and below- (roots) ground compartments and C contents per fraction [%]⁶⁷. The total C-inputs were resolved by partitioning the C per fraction to the soil, depended on the lifecycle (annual or perennial) and lifetime in years. For perennial species several estimates are required, as variations arise from annualising the aboveground C-inputs (e.g. whether the leaves are deciduous or evergreen or the stem is considered a bioeconomy feedstock). Roots, on the contrary, remain over the entire rotation length in the soil. We adopted a linear approach by dividing the total C root by the lifetime in years, here assuming an uneven-aged approach⁶⁸.

Matching of target areas and plant species and computation of net sequestration of target areas-biopumps pairs (step 4)

Once both target areas and a list of candidate plant species were identified, it was necessary to determine matches that would be, *ad minimum*, biophysically possible. A clear way to achieve such pairings is to compare the pedoclimatic conditions prevailing in the target areas with those of the plants.

We developed a R (R Core Team ⁶⁹) engine provided in ⁴¹ to firstly identify the corresponding matches between potential biopumps and target areas, and subsequently run the models for both SOC stock changes and SOC losses to erosion per matched target area-biopump pair up to the year 2100. A match took place when the tolerance ranges of biopumps were within the values associated to the target areas (e.g. temperature, pH, elevation, climatic zone compatibility, etc.). The resulting combinations contain all the information gathered from the previous modules, which is a unique set of input data required to initiate the SOC model and compute the associated SOC stock changes.

We used them monthly time-stepped and processes-based Rothamsted C (RothC) model v26.3 ⁷⁰ to compute SCS (described in Supplementary Methods, including a model comparison in Table S8). To run RothC for a combination of multiple sites and biopumps we used the SoilR package v1.1 ⁷¹, which provides a library of functions and tools under the R environment. For the initialisation of the conceptual carbon pools we used pedotransfer functions (equation and constants provided in Supplementary Methods), which seemed appropriate, as other approaches (e.g. physico-chemical or model equilibrium analysis) are not measurable at the regional scale. Monthly temperature and precipitation data were retrieved from CHELSA ⁵⁹, and monthly evapotranspiration from CGIAR's High-Resolution Global Soil-Water Balance ⁷².

The SCS estimations were complemented with the method described in Lugato et al. ⁷³ to compute SOC erosion from soil erosion by water (detailed in Supplementary Methods). Input layers for soil erosion were used from the Global Soil Loss map at a 25 km (810 arcsec) resolution for the year 2012 ⁷⁴, based on a Revised Universal Soil Loss Equation (RUSLE)-based method ⁷⁵. Retained cover-management factors required by the SOC erosion method ⁷³ are listed in Supplementary Methods Table S9.

The p-value were computed via a paired sample t-test.

References

1. Sanderman, J., Hengl, T. & Fiske, G. J. Soil carbon debt of 12,000 years of human land use. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 9575–9580 (2017).
2. Smith, P. *et al.* How to measure, report and verify soil carbon change to realize the potential of soil carbon sequestration for atmospheric greenhouse gas removal. *Glob. Chang. Biol.* **26**, 219–241 (2020).
3. Gottschalk, P. *et al.* How will organic carbon stocks in mineral soils evolve under future climate? Global projections using RothC for a range of climate change scenarios. *Biogeosciences* **9**, 3151–3171 (2012).
4. Rumpel, C. *et al.* The 4p1000 initiative: Opportunities, limitations and challenges for implementing soil organic carbon sequestration as a sustainable development strategy. *Ambio* **49**, 350–360 (2020).
5. Minasny, B. *et al.* Soil carbon 4 per mille. *Geoderma* **292**, 59–86 (2017).
6. White, R. E., Davidson, B., Lam, S. K. & Chen, D. A critique of the paper 'Soil carbon 4 per mille' by Minasny et al. (2017). *Geoderma* **309**, 115–117 (2018).
7. Smith, P. *et al.* Biophysical and economic limits to negative CO₂ emissions. *Nat. Clim. Chang.* (2015) doi:10.1038/nclimate2870.
8. Paustian, K. *et al.* Quantifying carbon for agricultural soil management: from the current status toward a global soil information system. *Carbon Manag.* **10**, 567–587 (2019).
9. Brander, M., Ascui, F., Scott, V. & Tett, S. Carbon accounting for negative emissions technologies. *Clim. Policy* **0**, 1–19 (2021).
10. IPCC. *Climate Change and Land: An IPCC Special Report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems.* (2019). doi:10.4337/9781784710644.

11. Amelung, W. *et al.* Towards a global-scale soil climate mitigation strategy. *Nat. Commun.* **11**, 1–10 (2020).
12. Albers, A., Avadí, A., Benoist, A., Collet, P. & Hélias, A. Modelling dynamic soil organic carbon flows of annual and perennial energy crops to inform energy-transport policy scenarios in France. *Sci. Total Environ.* **718**, 1–19 (2020).
13. Paustian, K., Larson, E., Kent, J., Marx, E. & Swan, A. Soil C Sequestration as a Biological Negative Emission Strategy. *Front. Clim.* **1**, 1–11 (2019).
14. Chimento, C., Almagro, M. & Amaducci, S. Carbon sequestration potential in perennial bioenergy crops: The importance of organic matter inputs and its physical protection. *GCB Bioenergy* **8**, 111–121 (2016).
15. Stockmann, U. *et al.* The knowns, known unknowns and unknowns of sequestration of soil organic carbon. *Agric. Ecosyst. Environ.* **164**, 80–99 (2013).
16. Mellor, P., Lord, R. A., Joao, E., Thomas, R. & Hursthouse, A. Identifying non-agricultural marginal lands as a route to sustainable bioenergy provision - A review and holistic definition. *Renew. Sustain. Energy Rev.* **135**, (2020).
17. Kang, S. *et al.* Marginal Lands: Concept, Assessment and Management. *J. Agric. Sci.* **5**, 129–139 (2013).
18. Blanco-Canqui, H. Growing Dedicated Energy Crops on Marginal Lands and Ecosystem Services. *Soil Sci. Soc. Am. J.* **80**, 845–858 (2016).
19. Dauber, J. & Miyake, S. To integrate or to segregate food crop and energy crop cultivation at the landscape scale? Perspectives on biodiversity conservation in agriculture in Europe. *Energy. Sustain. Soc.* **6**, (2016).
20. Von Cossel, M. *et al.* Prospects of bioenergy cropping systems for a more social-ecologically sound bioeconomy. *Agronomy* vol. 9 (2019).
21. Gibbs, H. K. & Salmon, J. M. Mapping the world's degraded lands. *Appl. Geogr.* **57**, 12–21 (2015).
22. Oldeman, L. R., Hakkeling, R. T. A. & Sombroek, W. . G. *World map of the status of human-induced soil degradation: an explanatory note. Global Assessment of Soil Degradation GLASOD* (1991) doi:10.1016/0016-7061(92)90047-b.
23. Bai, Z. G., Dent, D. L., Olsson, L. & Schaepman, M. E. *An update of GLADA - Global assessment of land degradation and improvement. Zurich Open Repository and Archive* (2011) doi:report 2008/01.
24. FAO. *Land degradation assessment in drylands (LADA): Methodology and results. LADA Project* (2013).
25. Yengoh, G. T., Dent, D., Olsson, L., Tengberg, A. E. & Tucker III, C. J. *Use of the Normalized Difference Vegetation Index (NDVI) to Assess Land Degradation at Multiple Scales.* (2014) doi:10.1007/978-3-319-24112-8.
26. Eliasson, Å. *Review of Land Evaluation Methods for Quantifying Natural Constraints to Agriculture. JRC Scientific and Technical Reports* (2007).
27. Sanchez, P. A. *et al.* Digital soil map of the world. *Science (80-.)*. **325**, 680–681 (2009).
28. FAO/IIASA. *Harmonized World Soil Database (version 1.2).* FAO, Rome, Italy and IIASA, Laxenburg, Austria (2009).
29. IIASA/FAO. *Global Agro-ecological Zones (GAEZ v3.0).* (2012).
30. Goga, T. *et al.* A review of the application of remote sensing data for abandoned agricultural land identification with focus on Central and Eastern Europe. *Remote Sens.* **11**, (2019).
31. Perpiña Castillo, C. *et al.* *Agricultural land abandonment in the EU within 2015-2030. JRC Policy Insights* (2018).

32. Yin, H. *et al.* Mapping agricultural land abandonment from spatial and temporal segmentation of Landsat time series. *Remote Sens. Environ.* **210**, 12–24 (2018).
33. Winkler, K., Fuchs, R., Rounsevell, M. & Herold, M. Global land use changes are four times greater than previously estimated. *Nat. Commun.* **12**, 1–10 (2021).
34. Lewis, S. M. & Kelly, M. Mapping the potential for biofuel production on marginal lands: Differences in definitions, data and models across scales. *ISPRS Int. J. Geo-Information* **3**, 430–459 (2014).
35. Cai, A. *et al.* Long-term straw decomposition in agro-ecosystems described by a unified three-exponentiation equation with thermal time. *Sci. Total Environ.* **636**, 699–708 (2018).
36. Morais, T. G., Teixeira, R. F. M. & Domingos, T. Detailed global modelling of soil organic carbon in cropland, grassland and forest soils. *PLoS One* **14**, 1–27 (2019).
37. Milbrandt, A. & Overend, R. P. Assessment of Biomass Resources from Marginal Lands in APEC Economies. 52 (2009) doi:10.2172/968464.
38. Elbersen, B. *et al.* Deliverable 2.6 Methodological approaches to identify and map marginal land suitable for industrial crops in Europe. EU Horizon 2020; MAGIC; GA-No.: 727698 (2020).
39. FAO. *Global ecological zones for FAO forest reporting: 2010 Update. Forest resources Assessment Working Paper 179* (2012).
40. FAO. FAO ECOCROP: The Crop Environmental Requirements Database. <http://ecocrop.fao.org/ecocrop/srv/en/home>; available on: <https://github.com/supersistence/EcoCrop-ScrapeR> (2018).
41. Albers, A., Avadí, A. & Hamelin, L. Data to assess spatially and species-explicit plant-based soil carbon sequestration on global marginal land. (2021) doi:doi:10.48531/JBRU.CALMIP/A3CIFZ.
42. FAO. *Technical specifications and country guidelines for Global Soil Organic Carbon Sequestration Potential Map (GSOCseq). Pillar 4 Working Group & INSII vol. 42* (2020).
43. Eyring, V. *et al.* Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* **9**, 1937–1958 (2016).
44. ESA. Land Cover CCI Product User Guide Version 2. Tech. Rep. maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf (2017).
45. Campbell, J. E., Lobell, D. B., Genova, R. C. & Field, C. B. The global potential of bioenergy on abandoned agriculture lands. *Environ. Sci. Technol.* **42**, 5791–5794 (2008).
46. Goldewijk, K. K. Estimating global land use change over the past 300 years: The HYDE database. *Global Biogeochem. Cycles* **15**, 417–433 (2001).
47. Ramankutty, N. & Foley, J. A. Estimating historical changes in global land cover: Croplands from 1700 to 1992. *Global Biogeochem. Cycles* **13**, 997–1027 (1999).
48. Goldewijk, K. K., Beusen, A., Doelman, J. & Stehfest, E. Anthropogenic land use estimates for the Holocene; HYDE 3.2. *Earth Syst. Sci. Data Discuss.* **9**, 927–953 (2017).
49. Jones, R. *et al.* Updated common bio-physical criteria to define natural constraints for agriculture in Europe : definition and scientific justification for the common biophysical criteria : technical factsheets. (2012). doi:10.2788/91182.
50. Ledo, A. *et al.* A global, empirical, harmonised dataset of soil organic carbon changes under perennial crops. *Sci. Data* **6**, 1–7 (2019).
51. White, C. M., Kemanian, A. R. & Kaye, J. P. Implications of carbon saturation model structures for simulated nitrogen mineralization dynamics. *Biogeosciences* **11**, 6725–6738 (2014).
52. Heitkamp, F., Wendland, M., Offenberger, K. & Gerold, G. Implications of input estimation, residue quality and carbon saturation on the predictive power of the Rothamsted Carbon Model. *Geoderma* **170**, 168–175 (2012).

53. Kirschbaum, M. U. F., Moineau, G. Y. K., Hedley, C. B., Beare, M. H. & McNally, S. R. A conceptual model of carbon stabilisation based on patterns observed in different soils. *Soil Biol. Biochem.* **141**, 107683 (2020).
54. Stewart, C. E., Plante, A. F., Paustian, K., Conant, R. T. & Six, J. Soil Carbon Saturation: Linking Concept and Measurable Carbon Pools. *Soil Sci. Soc. Am. J.* **72**, 379–392 (2008).
55. Di Gregorio, A. *Land Cover Classification System. Classification concepts. Software version 3. October* (2016).
56. Van Asselen, S. & Verburg, P. H. Land cover change or land-use intensification: Simulating land system change with a global-scale land change model. *Glob. Chang. Biol.* **19**, 3648–3667 (2013).
57. FAO and ITPS. *Global Soil Organic Map V1.5: Technical Report.* (2020) doi:<https://doi.org/10.4060/ca7597en>.
58. UNEP-WCMC. *User Manual for the World Database on Protected Areas and world database on other effective area-based conservation measures : 1.6.* https://wdpa.s3-eu-west-1.amazonaws.com/WDPManual/English/WDPManual_WDOECM_Manual_1_6.pdf (2019).
59. Karger, D. N. *et al.* Data from: Climatologies at high resolution for the earth's land surface areas. *Dryad Digit. Repos.* (2018) doi:<http://dx.doi.org/doi:10.5061/dryad.kd1d4>.
60. GADM. Global Administrative (GADM) maps and data. https://gadm.org/download_world.html (2018).
61. Galland, V., Avadí, A., Bockstaller, C. & Paillat, J.-M. Data to inform the modelling of direct nitrogen field emissions from global agriculture. *Portail Data INRAE, V1* (2020).
62. Li, W., Ciais, P., Makowski, D. & Peng, S. Data descriptor: A global yield dataset for major lignocellulosic bioenergy crops based on field measurements. *Sci. Data* **5**, 1–10 (2018).
63. Köppen, W. *Grundrisse der Klimakunde.* (Walter de Gruyter Co., 1931).
64. Fischer, G. *et al.* *Global Agro-ecological Zones (GAEZ v3.0): Model Documentation.* (2012).
65. Monfreda, C., Ramankutty, N. & Foley, J. A. Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. *Global Biogeochem. Cycles* **22**, 1–19 (2008).
66. Bolinder, M. A., Janzen, H. H., Gregorich, E. G., Angers, D. A. & VandenBygaart, A. J. An approach for estimating net primary productivity and annual carbon inputs to soil for common agricultural crops in Canada. *Agric. Ecosyst. Environ.* **118**, 29–42 (2007).
67. Ma, S. *et al.* Variations and determinants of carbon content in plants: A global synthesis. *Biogeosciences* **15**, 693–702 (2018).
68. Albers, A. *Prise en compte de la dimension temporelle dans l'évaluation environnementale des produits de la biomasse : Modélisation dynamique du carbone. PhD thesis* (École doctorale GAIA – Biodiversité, Agriculture, Alimentation, Environnement, Terre, Eau Portée par l'Université de Montpellier, 2019).
69. R Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing* (2018).
70. Coleman, K. *et al.* Simulating trends in soil organic carbon in long-term experiments using RothC-26.3. *Geoderma* **81**, 29–44 (1997).
71. Sierra, C. A., Müller, M. & Trumbore, S. E. Models of soil organic matter decomposition: The SoilR package, version 1.0. *Geosci. Model Dev.* **5**, 1045–1060 (2012).
72. Trabucco, A. & Zomer, R. J. *Global High-Resolution Soil-Water Balance. figshare. Dataset.* vol. 2010 (2010).
73. Lugato, E., Paustian, K., Panagos, P., Jones, A. & Borrelli, P. Quantifying the erosion effect on current

carbon budget of European agricultural soils at high spatial resolution. *Glob. Chang. Biol.* **22**, 1976–1984 (2016).

74. Borrelli, P. *et al.* An assessment of the global impact of 21st century land use change on soil erosion. *Nat. Commun.* **8**, (2013).
75. Foster, R. G. *Revised Universal Soil Loss Equation – Version 2 (RUSLE2)*. (2005).

Acknowledgements

This project has received funding from the European Union’s Horizon 2020 Research and Innovation Programme under the Grant Agreement ID: 869192. This research is subject to further research in the frame of the NEGEM project (<https://www.negemproject.eu>). The authors thank Shivesh Karan for cross-referencing initial data on land cover and soil organic carbon classes.

Author contributions statement

A. Al. and L. H. conceived the study with inputs from A. Av.; A. Al. and A. Av. retrieved and treated all data; A. Al. analysed the results. All authors reviewed the manuscript.

Competing interests

The author(s) declare no competing interests.

Data availability

Dataset required for the framework implementation in ⁴¹ (<https://doi.org/10.48531/JBRU.CALMIP/A3CIFZ>).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMethods.pdf](#)
- [SupplementaryResults.pdf](#)