



HAL
open science

A new approach to estimate soil organic carbon content targets in European croplands topsoils

Lorenza Pacini, Pierre Arbelet, Songchao Chen, Aurélie Bacq-Labreuil, Christophe Calvaruso, Florian Schneider, D. Arrouays, Nicolas P. A. Saby, Lauric Cécillon, Pierre Barré

► To cite this version:

Lorenza Pacini, Pierre Arbelet, Songchao Chen, Aurélie Bacq-Labreuil, Christophe Calvaruso, et al.. A new approach to estimate soil organic carbon content targets in European croplands topsoils. *Science of the Total Environment*, 2023, 900, pp.165811. 10.1016/j.scitotenv.2023.165811 . hal-04193166

HAL Id: hal-04193166

<https://hal.science/hal-04193166>

Submitted on 1 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Journal Pre-proof

A new approach to estimate soil organic carbon content targets in European croplands topsoils

Lorenza Pacini, Pierre Arbelet, Songchao Chen, Aurélie Bacq-Labreuil, Christophe Calvaruso, Florian Schneider, Dominique Arrouays, Nicolas P.A. Saby, Lauric Cécillon, Pierre Barré



PII: S0048-9697(23)04436-4

DOI: <https://doi.org/10.1016/j.scitotenv.2023.165811>

Reference: STOTEN 165811

To appear in: *Science of the Total Environment*

Received date: 16 March 2023

Revised date: 21 July 2023

Accepted date: 24 July 2023

Please cite this article as: L. Pacini, P. Arbelet, S. Chen, et al., A new approach to estimate soil organic carbon content targets in European croplands topsoils, *Science of the Total Environment* (2023), <https://doi.org/10.1016/j.scitotenv.2023.165811>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2023 Published by Elsevier B.V.

A new approach to estimate soil organic carbon content targets in European croplands topsoils.

Lorenza Pacini^{1,2}, Pierre Arbelet², Songchao Chen³, Aurélie Bacq-Labreuil², Christophe Calvaruso², Florian Schneider⁴, Dominique Arrouays⁵, Nicolas P.A. Saby⁵, Lauric Cécillon¹, Pierre Barré¹

Institutional affiliations:

¹Laboratoire de Géologie, École normale supérieure, CNRS, Université PSL, IPSL, 75005 Paris, France

²Greenback SA (commercial name: Genesis), 75003 Paris, France

³ZJU-Hangzhou Global Scientific and Technological Innovation Center, Hangzhou 311200, China

⁴Thünen Institute of Climate-Smart Agriculture, 38116 Braunschweig, Germany

⁵INRAE, Info&Sols Unit, 45075 Orléans, France

Corresponding author: Lorenza Pacini - lorenza.pacini@cnrs.fr, Laboratoire de Géologie, École normale supérieure, 24 rue Lhomond 75005 Paris, France

Abstract

Adopting land management practices that increase the stock of soil organic carbon (SOC) in croplands is widely promoted as a win-win strategy to enhance soil health and mitigate climate change. In this context, the definition of reference SOC content and stock values is needed to provide reliable targets to farmers, policymakers, and stakeholders. In this study, we used the LUCAS dataset to compare different methods for evaluating reference SOC content and stock values in European croplands topsoils (0-

20 cm depth). Methods gave generally similar estimates although being built on very different assumptions. In the absence of an objective criterion to establish which approach is the most suitable to determine SOC reference values, we propose an ensemble modelling approach that consists in extracting the estimates using different relevant methods and retaining the median value among them. Interestingly, this approach led us to select values from the three different approaches with similar frequencies. Using estimated bulk density values, we obtained a first rough estimate of 3.5 Gt C of SOC storage potential in the cropland topsoils that we interpret as a long-term aspirational target that would be reachable only under extreme changes in agricultural practices. The use of additional methods in the ensemble modelling approach and more valid statistical spatial estimates may further refine our approach designed for the estimation of SOC reference values for croplands.

Keywords: Soil organic carbon, carbon storage, climate mitigation, LUCAS dataset, data-driven modelling, ensemble modelling.

1. Introduction

Soil organic matter plays a fundamental role in providing ecosystem services such as biomass production and flood and erosion mitigation (Comerford et al., 2013; Johannes et al., 2017). Moreover, soil organic carbon (SOC), the main component of soil organic matter, is one of the main global carbon pools (Lal, 2004a). However, the conversion of natural vegetation to agricultural land has caused a depletion in SOC stock (tons of carbon per unit surface), threatening ecosystem functioning, and causing greenhouse gas emissions (Lal, 2004a; Lal, 2004b; Macías and Camps Arbestain, 2010; Poeplau et al., 2011; Dutta and Dutta, 2016; Sanderman et al., 2017; Garnier et al., 2022). Accordingly,

increasing SOC content (grams of carbon per kilogram of fine earth) in soils is widely promoted as a means to improve soil health and mitigate climate change (Lal, 2004b; Bispo et al., 2017; Shukla et al., 2019; Powlson et al., 2022) and programs encouraging carbon sequestration in soils are being developed (Chabbi et al., 2017; IPBES, 2018; Shukla et al., 2019; Montanarella and Panagos, 2021; Oldfield et al., 2022). As a consequence, the definition of reference SOC content values is essential to provide reliable targets to farmers, policymakers and stakeholders.

Reference SOC content values may depend on the scope of the specific program or initiative: preserving soil health or soil-based ecosystem services might require minimal SOC content values that ensure specific soil functions (Bonfante et al., 2020; Wood and Blankinship, 2022; Loveland, 2023); on the other hand, as a first approximation, we can hypothesise that the higher the soil carbon content is the better for climate mitigation and soil health, as long as SOC storing practices do not generate enhanced emissions of greenhouse gases (Powlson et al., 2011). Thus, SOC targets for climate mitigation do not necessarily coincide with targets for soil health. Moreover, determining reference values requires defining under what conditions they should be reached (Smith, 2004). Some approaches aim at estimating maximal reachable SOC contents (Hassink, 1997; Chen, Arrouays, Angers, Martin and Walter, 2019; Georgiou et al., 2022), while others add constraints, for example restricting to SOC contents actually observed in field measurements (Schneider, Poeplau and Don, 2021; Drexler et al., 2022; Guillaume et al., 2022) or keeping similar land cover (Chen, Arrouays, Angers, Chenu et al., 2019). Additionally, while the necessity of considering the influence of climatic and pedological variables on SOC storage potentials is generally recognized in these studies, they high-

ly differ in the set of parameters and how they are employed.

As a consequence, SOC storage potential estimates Δ SOC, calculated as the difference between a reference value and the actual SOC content of a soil and expressed as stocks or concentrations differences, depend on several modelling choices, namely the algorithm(s) used for the estimation of the reference values and the choice of a reference system, the pedoclimatic variables that are taken into consideration, and the constraints under which the SOC content is expected to evolve.

The goal of this study is to explore, for the first time and on a very large dataset of more than 20000 measurements (LUCAS dataset, Orgiazzi et al., 2018), the consequences of these modelling choices on the determination of SOC reference values for croplands topsoils (0-20 cm depth) of the European Union (EU27) and the United Kingdom. To this goal, we compared the estimates resulting from three modelling approaches: (i) a proposed novel method that uses SOC contents in natural systems as references for given pedoclimates; (ii) the data-driven reciprocal modelling approach developed by Schneider, Poeplau and Don et al. (2021), that estimates changes in SOC content upon conversion to grassland (Schneider, Poeplau and Don, 2021); and (iii) the carbon landscape zones approach developed by Chen et al. (2019), that uses best-performing croplands in given pedoclimates and for given net primary production levels as references (Chen, Arrouays, Angers, Chenu et al., 2019). We compared both the SOC reference values obtained with the different models and the resulting Δ SOC. We then combined the three methods using an ensemble modelling framework. The ensemble modelling framework allows to seek consensus and mitigate local over- or under-estimations from single models.

2. Methods

2.1 Data

2.1.1 Soil data

Soil data (0-20cm) were extracted from the LUCAS Topsoil dataset (EUROSTAT, 2009; EUROSTAT, 2012; European Commission, 2013; ESDAC, 2013; EUROSTAT, 2015; ESDAC, 2020). Specifically, we used four subsets:

- Dataset n. 1 - LUCAS 2009 and 2012: topsoil data and land cover data of mineral soils ($\text{SOC} < 120 \text{ gC kg}^{-1}$) with cropland, grassland, or woodland land cover (15251 sites). Sites were sampled in 2009, apart from sites in Romania, Bulgaria, Malta, and Cyprus that were sampled in 2012.
- Dataset n. 2 - LUCAS 2015: topsoil data and land cover data of mineral soils ($\text{SOC} < 120 \text{ gC kg}^{-1}$) with cropland, grassland, or woodland land cover that were also sampled in LUCAS 2009/12 (15251 sites).
- Dataset n. 3 - LUCAS croplands: croplands of dataset n. 2 that were also croplands in dataset n. 1. (7324 sites).
- Dataset n. 3bis: subset of dataset n. 3 including only cropland sites that fell in the predictor space of the "data-driven reciprocal modelling" model (see Section 2.3) as described by Schneider, Poeplau and Don (2021) and for which net primary production data could be extracted (see Section 2.1.3) (6777 sites).

We used SOC content measurements from LUCAS 2015 and texture measurements from LUCAS 2009/12, because the texture was not re-measured in the second LUCAS campaign for resampled sites (European Commission, 2020). The texture was measured using laser diffraction in LUCAS 2015 instead of sieving and sedimentation used in

LUCAS 2009/12 (European Commission, 2020). To ensure a homogeneous dataset, we decided to use only texture data from the LUCAS 2009/12, i.e. we discarded sites that were sampled in the LUCAS 2015 campaign only.

LUCAS data are openly available in ESDAC and Eurostat at the following urls. LUCAS 2009/12 Topsoil (ESDAC, 2013): <https://esdac.jrc.ec.europa.eu/content/lucas-2009-topsoil-data>; LUCAS 2015 Topsoil (ESDAC, 2020): <https://esdac.jrc.ec.europa.eu/content/lucas2015-topsoil-data>, LUCAS 2009 land cover data (EUROSTAT, 2009): <https://ec.europa.eu/eurostat/web/lucas/data/primary-data/2009>; LUCAS 2012 land cover data (EUROSTAT, 2012): https://ec.europa.eu/eurostat/web/lucas/data/primary_data/2012; LUCAS 2015 land cover data (EUROSTAT, 2015): <https://ec.europa.eu/eurostat/web/lucas/data/primary-data/2015>.

2.1.3 Climate data

Climate data at the LUCAS sites were extracted from the Climatic Research Unit dataset (Harris et al., 2020), available at <https://catalogue.ceda.ac.uk/uuid/89e1e34ec3554dc98594a5732622bce9> (last accessed 6th October 2022). Specifically, the data that we extracted were the following:

- Mean annual temperature: average over 30 years (1985 to 2015);
- Standard deviation of monthly average temperature over the year: average over 30 years (1985 to 2015);
- Total annual precipitation: average over 30 years (1985 to 2015);
- Aridity index: difference between the potential evapotranspiration and the total annual precipitation, average over 30 years (1985 to 2015);

- Mean monthly temperature: average over 10 years (2005 to 2015);
- Mean monthly precipitation: average over 10 years (2005 to 2015);
- Monthly potential evapotranspiration: average over 10 years (2005 to 2015).

Monthly data were averaged over 10 years because they were used in the "carbon landscape zones" method together with net primary production data extracted over 10 years.

2.1.3 Net primary production data

Net primary production (NPP) data at 8974 LUCAS 2015 croplands sites were extracted from the MODIS/Terra MOD17A2H dataset (Running et al., 2015). The data were accessed using the Google Earth Engine, product MOD17A2H (https://developers.google.com/earth-engine/datasets/catalog/MODIS_006_MOD17A2H#description, last accessed 6th October 2022). For each month of the year, the monthly NPP was extracted as the mean daily NPP (from 8-days data covering the month) multiplied by the number of days in the month. Data were averaged over 10 years (2005 to 2015). By default, the NPP was extracted with a spatial resolution of 1 km. However, data could not be found for 105 sites. For these points, we looked for data at an increasingly larger spatial resolution (2 km: 64 sites, 3 km: 10 sites, 4 km: 6 sites). Data was not found for 25 sites.

2.2 Conversion of soil organic carbon contents to stocks

To ensure the comparability of the results across different methods and with literature data, we have expressed reference values and SOC storage potentials in terms of SOC content (grams of organic carbon per kilogram of fine earth, gC kg^{-1}) and in terms of SOC stock (tons of organic carbon per hectare, Mg ha^{-1}).

The SOC content values and the reference values SOC_{ref} expressed in $gC\ kg^{-1}$ were converted to SOC stock values ($Mg\ ha^{-1}$) using the formula in eq. 1 proposed by Schneider, Poeplau and Don. (2021) and Lugato, Panagos, Bampa et al. (2014):

$$SOC_{stock} = (SOC_{content}/10) * BD * Depth * (1 - Coarse_{volume\ fraction}) \quad (1)$$

where $Depth = 20cm$, $SOC_{content}$ is expressed in $gC\ kg^{-1}$, $Coarse_{volume\ fraction}$ is the volume fraction of coarse fragments (unitless), and BD is the bulk density of the fine earth, expressed in $g\ cm^{-3}$. The BD was estimated using the pedotransfer function in eq. 2 proposed by Hollis et al. (2012):

$$BD = 0.80806 + 0.823844 * e^{-0.27993 \times (SOC_{content}/10) + 0.014065 * sand - 0.0010299 * clay} \quad (2)$$

where $sand$ and $clay$ are the mass percentage of sand and clay, respectively, and $SOC_{content}$ is expressed in $gC\ kg^{-1}$. This function fitted the bulk density of 333 cultivated topsoils with a root mean square error of $0.13\ g\ cm^{-3}$ in the original study (Hollis et al., 2012).

The volume fraction of coarse fragments was estimated from the mass fraction of coarse fragments available in the LUCAS dataset using eq. 3 (Pacini et al., 2023):

$$Coarse_{volume\ fraction} = \frac{Coarse_{mass\ fraction}}{\frac{\rho_{coarse}}{R} + \left(1 - \frac{\rho_{coarse}}{BD}\right) Coarse_{mass\ fraction}} \quad (3)$$

where $Coarse_{mass\ fraction}$ is the mass fraction of coarse fragments (unitless). We assumed a constant density $\rho_{coarse} = 2.6\ g\ cm^{-3}$ for the coarse fragments.

2.3 Estimating soil organic carbon content reference values

2.3.1 Natural references per pedoclimate

The "natural reference per pedoclimate" is a novel approach based on pedoclimatic contextualisation of SOC content values. Pedoclimatic contextualization was performed by clustering the LUCAS 2015 dataset (dataset n. 2) according to its pedoclimatic fea-

tures.

As pedological features, we selected the clay and silt contents and the carbonate content. As climate features we selected the mean annual temperature, the standard deviation of monthly average temperature over the year, the total annual precipitation, and the aridity index (difference between the potential evapotranspiration and the total annual precipitation). These pedoclimatic features have been selected because they are known to influence the SOC content (Roussel and Bourmeau, 2001; Arrouays et al., 2006; Fernández-Ugalde and Tóth, 2017; Chenu et al., 2013; Büchi et al., 2022) and, as a first approximation, are independent of human activity. The goal is to isolate the factors that influence the SOC content regardless of the land management practices so that “intrinsic” SOC potentials independent of the human activity can be estimated. Because of this reason, the soil pH and C:N ratio were not retained as predictors, even though they influence the SOC content (Rasmussen et al., 2018).

The clustering was performed in two parallel steps using the Python library Scikit-learn (version 1.1.1) (Pedregosa et al., 2011). First, sites were clustered according to the four climatic features: mean annual temperature, standard deviation of monthly average temperature over the year, total annual precipitation, and aridity index. Second, sites were clustered according to their texture (clay and silt content) and carbonate content. Then, the two clustering results are combined in a unique label "x-y" with x being the climate label and y the soil label of the site (as an example, pedoclimatic cluster 2-0 means climate cluster n. 2 and soil cluster n. 0). The combined label defines the pedoclimatic cluster of a site. Performing two clustering procedures in parallel has the advantage of providing a higher interpretability of the results, because it allows to inde-

pendently distinguish the type of climate and the type of soil of a given site.

The clustering algorithm (Gaussian mixture or agglomerative clustering) and the number of clusters for each of the clustering procedures were chosen according to the following criteria:

- interpretability: not too many clusters to ensure enough discrimination between the clusters and thus the possibility to characterise them;
- representativity of the clusters: not too many clusters to limit the number of small-population clusters, and choose algorithms that promote homogeneous cluster populations;
- clustering performance: high silhouette score and low Davies Bouldin score for agglomerative clustering, low Bayesian information criterion (BIC score) for Gaussian mixture;
- geographic definition: for climate clustering only, ensure that the clusters are geographically well-defined (visual examination).

We discarded agglomerative clustering in the climate clustering step because it resulted in Davies Bouldin scores near 1, meaning that clusters are similar one to another. Instead, we used the Gaussian mixture clustering algorithm, with full covariance type and fixing the number of clusters to eleven. The BIC score decreased with the number of clusters in the range we tested (3 to 19 clusters). However, we decided to limit the number of clusters to reduce the total number of pedoclimatic clusters after combination with the soil clustering. We guided the choice of the number of clusters by visually checking that climatic clusters were not highly dispersed geographically. Supplementary Figure 1A reports the climate clusters obtained using 10 to 12 clusters with the Gaus-

an mixture algorithm and using 11 clusters with the agglomerative clustering algorithm: we selected the Gaussian mixture algorithm with 11 clusters because it provides the most well-defined clusters. The resulting model outputs a "climate-label" for a given site, which is an integer between 0 and 10.

For the soil clustering step, we selected the agglomerative clustering algorithm, with ward linkage and fixing the number of clusters to four. The number of clusters was set to four because the Davies Bouldin score was minimal with four clusters and the silhouette score decreased for more than five clusters. The resulting model outputs a "soil-label" for a given site, which is an integer between 0 and 3. We did not use Gaussian mixture clustering because it resulted in less well-defined clusters in terms of carbonate content and texture (compare Supplementary Figure 1B with Table 2 and Figure 2).

After combining the climate and soil clustering results, 44 pedoclimatic clusters were defined, with 3 to 1659 different sites per cluster (clusters 5-0 and 4-1, respectively).

As a first approximation, grasslands and woodlands can be considered as natural systems, where the SOC content has been much less affected by human activity compared to croplands (Poeplau et al., 2011; Sanderman et al, 2017). Thus, they represent a reference of reachable SOC content values in a given pedoclimatic context under natural vegetation. Accordingly, a single SOC reference value expressed as SOC content (SOCref) was defined per pedoclimatic cluster based on the distribution of SOC content in grasslands and woodlands in that cluster. For each pedoclimatic cluster, we selected grasslands and woodlands in dataset n. 2 that belong to the cluster and we tested several percentiles of SOC content as a potential SOCref value. We tested the percentiles $p = 0.4, 0.45, 0.5, \text{ and } 0.55$, i.e. the SOCref of a pedoclimatic cluster is the minimal SOC

of the best 60% (55%, 50%, 45% respectively) grasslands and woodlands in the pedoclimatic cluster. Following the procedure proposed by Chen, Arrouays, Anger, Chenu et al. (2019) to estimate the confidence intervals for the SOCref values, we performed bootstrapping: the percentile of each cluster was calculated after sampling 85% of the grasslands and woodland with replacement. The procedure was repeated 100 times and the SOCref value for the pedoclimatic cluster was defined as the mean of the 100 percentile values of SOC.

Additionally, we used the following procedure to populate pedoclimatic clusters containing fewer than 30 grasslands and woodlands prior to random sampling. First, a K-nearest neighbours model was trained using all grasslands and woodlands and based on the pedoclimatic features. We used the implementation of the Python library Scikit-learn (version 1.1.1) (Pedregosa et al., 2011). The model takes the pedoclimatic features of a site as input and outputs the K identifiers of the grassland or woodland sites that are the closest to the input site in the features space, as well as the corresponding K distances in the features space. Then, we set $K=1$ and extracted the first neighbours of all the sites that are already in the cluster. These are candidate sites to populate the cluster. The candidate sites were sorted according to their distances from the sites already in the cluster. If the cluster initially contained N sites, then $30-N$ sites had to be added and N_c candidate sites (with $N_c \leq N$) had been extracted. If $N_c \geq 30-N$, then the first $30-N$ candidates were added to the cluster. Otherwise, all N_c sites were added to the cluster and the procedure was repeated with $K=2$ (second neighbours), $K=3$ (third neighbours), etc., until the populated cluster contained 30 sites. Seventeen pedoclimatic clusters over 44 were populated according to this procedure (0-3, 1-0, 2-0, 2-3, 3-0, 3-3,

4-0, 4-3, 5-0, 5-3, 6-0, 6-3, 7-0, 7-3, 8-0, 8-3 and 9-0).

We converted the SOCref values from content to stock using the formula described in section 2.2.

Finally, for each site in the LUCAS 2015 croplands database (database n. 3), we extracted the pedoclimatic cluster to which it belongs and calculated the SOC storage potential as the difference between the SOCref of the pedoclimatic cluster and actual SOC content of the site.

2.3.2 Data-driven reciprocal modelling

The data-driven reciprocal modelling method in its original implementation is described by Schneider, Poeplau and Don (2021) and applied to the prediction of SOC stock storage potentials of LUCAS 2015 croplands if they were converted to grasslands. In summary, the method consists of two steps. First, a random forest model that predicts the SOC stock from pedoclimatic features is trained on LUCAS 2015 grasslands. Then, the trained model is used to predict SOC stock in croplands that belong to the model's domain of applicability. It is estimated that the residuals of the prediction, i.e. the difference between the predicted SOC stock and the actual SOC stock, represent the potential change in SOC stock that would be observed if the cropland site were converted to grassland. Thus, the predicted "grassland-equivalent" SOC stocks can be interpreted as reference values SOCref for the LUCAS croplands that take into account the pedoclimatic features of the site.

In their case study, Schneider, Poeplau and Don (2021) used the following pedoclimatic features as predictors of SOC stocks: soil texture, pH, C:N ratio, carbonates content, soil group, elevation, slope, orientation of the slope (called "aspect" in the LUCAS da-

taset), and climate data. In this study, to allow for comparison with the "natural references per pedoclimate" method, we trained a modified version of the model. The first difference is that we removed the pH and C:N ratio from the list of predictors. The second difference is that we trained the model for the prediction of SOC contents instead of SOC stocks and only then we converted the observed SOC contents and the predicted SOCref values into stocks.

2.3.3 Carbon landscape zones

In 2019, Chen et al. proposed a method to estimate the SOC storage potential of French croplands based on the concept of carbon landscape zones (Chen, Arrouays, Angers, Chenu et al., 2019). In summary, the method consists of grouping cropland sites into carbon landscape zones based on their pedoclimatic features (clay content and climatic decomposition index) and on the NPP. Then, a reference SOC stock value, SOCref, is defined for each carbon landscape zone as a specific percentile of the observed SOC stock values among croplands that belong to the carbon landscape zone. We tested four percentiles: 0.3, 0.85, 0.9, or 0.95, i.e. the SOCref of a carbon landscape zone is the minimal SOC of the best 20% (15%, 10%, 5%, respectively) croplands in the carbon landscape zone. As in the "natural references per pedoclimate" method, the SOCref values are calculated by averaging over 100 random bootstrapping procedures. The most important difference with respect to the "natural references per pedoclimate" and the "data-driven reciprocal modelling" methods is that the SOCref values are estimated relative to other croplands and not to natural systems. Moreover, NPP is used as a feature, differently from the other methods.

The method was originally developed for estimating SOC storage potential of croplands

at the national scale in France (Chen, Arrouays, Angers, Chenu et al., 2019), using the French soil monitoring network (RMQS) dataset (Jolivet et al., 2006). Here, we applied the same methodology to the LUCAS croplands dataset (database n. 3) with two differences. The first difference from the original methodology is that the clustering was performed using pedoclimatic features and NPP extracted at the LUCAS cropland sites only and not over the whole territory of the EU27 and the United Kingdom. This choice was made for two reasons: first, to reduce the computational effort without reducing the accuracy of the climatic and NPP data, and second, to use only measured texture data and not spatial predictions. The second difference is that the SOCref values were extracted for each carbon landscape zone from the SOC content statistics and not from the SOC stock statistics. Only then, the conversion to SOC stocks was performed.

The number of carbon landscape zones was set to twenty to minimise the Bayesian information criterion (BIC) while keeping a minimal number of ten LUCAS croplands in each carbon landscape zone. The BIC values as a function of the number of carbon landscape zones are shown in Supplementary Figure 3A.

2.4 Ensemble modelling

We combined the "natural references per pedoclimate", "data-driven reciprocal modelling" and "carbon landscape zones" results using an ensemble modelling approach that consists in keeping the median estimate among the three, for each cropland site.

2.5 Estimation of the cumulative soil organic carbon storage potential of European croplands

We determined the median SOCref and then estimated the total SOC storage potential in croplands topsoils of the European Union and the United Kingdom by first computing

the median of the Δ SOC of all LUCAS sites per country, expressed in Mg ha^{-1} , and secondly by multiplying it by the country's croplands area and finally by summing the national estimates over the countries. The country croplands area data were obtained from Eurostat (Main farmland use by NUTS 2 regions, https://ec.europa.eu/eurostat/databrowser/view/EF_LUS_MAIN_custom_3779332/default/table?lang=en, last visited 7th November 2022). We used data from 2016 and summed the arable land and permanent crop areas. It should be noted that the median Δ SOC per country is not an unbiased estimate, because LUCAS sampling sites were not chosen using probability sampling (Teuling et al., 2021). Nevertheless, we used the median Δ SOC per country for simplicity.

2.6 Correlation coefficients calculation

We calculated the Spearman correlation coefficients between the estimates obtained using the three different approaches and between the estimates and the geographical coordinates of the sites using the Python library Scipy version 1.11.1.

3. Results and Discussion

3.1 Natural references per pedoclimate

3.1.1 Pedoclimatic clusters

The pedoclimatic clusters for the LUCAS dataset are shown in Figure 1A and the geographical regions to which the LUCAS sites of each pedoclimatic cluster belongs are reported in Table 1. Table 2 reports the carbonates content statistics and Figure 2 shows the texture triangles of sites based on the soil cluster. From these distributions, at first order, the soil clusters correspond to calcareous soils (soil cluster n. 0), sandy soils (soil cluster n. 1), clay soils (soil cluster n. 3), and soils with average pedological

features (soil cluster n. 2) (Table 1). Using only four classes of soils results in a very rough pedological classification; however, this allows for populating all pedoclimatic clusters after combining the soil and climate labels.

We found that the SOC contents distributions in the 0-20 cm depth are different across different pedoclimatic clusters and for the same land cover (cropland, grassland, or woodland) (Supplementary Figure 2A), supporting the need for the pedoclimatic contextualization of SOC content measurements prior to the definition of reference SOC values. As an example, considering only clusters including at least 30 croplands, the lowest values of SOC in croplands are found in the cluster 7-1 (sandy soils in central Spain, median SOC = 6 gC kg⁻¹) and the highest values are found in cluster 5-3 (clay soils in Scandinavia, median SOC = 30 gC kg⁻¹) (Table 1). This is consistent with highest expected SOC contents in clay soils (Martin et al., 2011; Poeplau et al., 2020) and in colder climates (de Brogniez et al., 2014).

We also found that none of the LUCAS croplands and only nine grasslands and two woodlands belonged to the cluster 2-3, that means that soil of type 3 is rare in the geographical region of the climate cluster 2 (Portugal, north of Spain, Brittany, Ireland, Scotland, or the west of England). This pedoclimatic cluster is discarded in the following.

3.1.2 Estimation of soil organic carbon reference values

For each pedoclimatic cluster, we defined a reference SOC value, SOC_{ref}, for croplands as a percentile of the observed SOC content among grasslands and woodlands.

The percentile used to define the SOC_{ref} values should be low enough to take into account that the reduced carbon input in croplands compared to grasslands and wood-

lands inhibits the possibility of reaching the highest SOC contents observed in natural systems (Powlson et al., 2022), i.e. the SOC_{ref} values should be reachable under cropland land cover. However, the percentile should also be high enough to discard young grasslands and woodlands that have recently been converted from croplands. We tested the percentiles $p = 0.4, 0.45, 0.5, \text{ and } 0.55$ and we found that the corresponding SOC_{ref} values do not diverge when the percentile is changed within this range (Supplementary Figure 3). We choose to use $p = 0.5$, i.e. the SOC_{ref} corresponding to median SOC content values for grasslands and woodlands in a pedoclimatic cluster. The choice of the percentile is necessarily arbitrary; however, using median values of grassland SOC contents as reference is consistent with the observation from long-term experiments at Rothamsted Research (UK) that 40 to 70% of the SOC stock of topsoils (0-23 cm depth) under natural vegetation could be reached under arable land use undergoing well-chosen agricultural practices (Powlson et al., 2022).

To allow for comparison with the "data-driven reciprocal modelling" and the "carbon landscape zones" approach, we have restricted the analysis to the dataset n.3bis.

Figure 3A shows the SOC_{ref} values of the LUCAS cropland sites. As expected, there is a high variability depending on the pedoclimatic cluster, with SOC_{ref} ranging from 10 gC kg⁻¹ to 61 gC kg⁻¹ (15 Mg ha⁻¹ to 167 Mg ha⁻¹ in terms of stock). The lowest SOC_{ref} values are observed in central Spain and the highest SOC_{ref} values are observed mostly in the United Kingdom, Ireland, Brittany (north-west of France), and Portugal. It should be noted that the SOC_{ref} values expressed as SOC stocks are not constant within a pedoclimatic cluster, because the conversion from SOC content to SOC stock is done based on the specific cropland site texture and SOC content. A high variability

of SOCref is consistent with the observation that SOC distributions are distinct across pedoclimatic clusters (Supplementary Figure 2A).

We have estimated the Δ SOC of croplands as the difference between the reference value SOCref and the actual SOC. Figure 4A shows the results on LUCAS croplands, expressed both in terms of SOC content and in terms of SOC stock. Broad geographical trends are visible: lower SOC storage potentials are found in Spain and in eastern Europe and highest values are observed in France, Scotland, Portugal, and Italy. This may suggest that the pedoclimatic clustering did not allow to completely get rid of the influence of pedoclimatic variables on SOC storage and isolating the effect of agricultural practices on SOC storage. However, heterogeneous values of Δ SOC are observed within the pedoclimatic clusters and there is no trend of Δ SOC with latitude (Spearman correlation coefficient = 0.1). This reassures on the efficacy of the pedoclimatic clustering method in isolating the effect of agricultural practices. It is possible that the observed broad geographical trends reflect spatial patterns of adaption of beneficial or detrimental agricultural practices and, probably, of land-use history (Heikkinen et al., 2013; Van Wesemael et al., 2010)

3.2 Data-driven reciprocal modelling

We have extracted the reference values SOCref and the resulting Δ SOC using the "data-driven reciprocal modelling" for the LUCAS croplands dataset (dataset n. 3bis). The results are shown in Figure 3B (SOCref values) and Figure 4B (Δ SOC values), expressed in terms of SOC content and SOC stocks. Compared to the "natural references per pedoclimate" method, a clearer influence of the latitude on the reference values is observed: higher SOCref in northern Europe and lower values in southern Europe (Fig-

ure 3B). This trend is found also for the Δ SOC values (Figure 4B). The existence of a trend of Δ SOC values with geographic coordinates may imply that the “Data-driven reciprocal modeling” method is slightly less appropriate in normalizing for the effect of pedoclimatic variables on SOC storage compared to the pedoclimatic clustering. Nevertheless, it has the advantage of providing continuous estimates of SOC_{ref} values as a function of pedoclimatic variables, contrarily to the clustering approaches that assign the same SOC_{ref} value to all croplands that belong to the same cluster, raising issues for the estimations at the frontier between clusters (“natural references per pedoclimate” and “carbon landscape zones” approaches).

3.3 Carbon landscape zones

3.3.1 Carbon landscape zones

The carbon landscape zones of the EU27 and United Kingdom are shown in Figure 1B. The frontiers among the carbon landscape zones resemble the frontiers among the climate clusters found with the pedoclimatic clustering method (Figure 1). However, some differences exist and are related to the fact that different pedoclimatic features are used and that the carbon landscape zones clustering also incorporates information on the NPP.

Similarly to what was found for the pedoclimatic clusters, we found that the SOC content distributions of croplands of different carbon landscape zones are distinct (Supplementary Figure 2B), supporting the ability of the “carbon landscape zones” method to contextualise SOC content measurements. Indeed, different distributions of SOC are expected under different pedoclimatic contexts (here one carbon landscape zone is one context). Nevertheless, some overlap among the distributions is observed, consistent

with a wide range of SOC values within the same carbon landscape zone depending on the adopted agricultural practices.

3.3.2 Estimation of soil organic carbon reference values

For each carbon landscape zone, we calculated a reference value SOC_{ref} as a specific percentile of the SOC content of croplands that belong to it. We tested four percentiles: 0.8, 0.85, 0.9, or 0.95 (Supplementary Figure 3B). We found that the SOC_{ref} values diverge when the percentile is increased from 0.9 to 0.95, thus we fixed the percentile to 0.9 and calculated the reference values SOC_{ref} for all carbon landscape zones expressed as SOC content. Then, the reference values SOC_{ref} were converted to stocks. The results are shown in Figure 3C. The resulting estimations of SOC storage potentials ΔSOC are shown in Figure 4C. Compared to the methods described above, the "carbon landscape zones" approach issues the most homogeneous estimations for the reference values across Europe, with the SOC_{ref} values ranging from 17 to 33 gC kg^{-1} (16 Mg ha^{-1} to 107 Mg ha^{-1} in terms of stock). As a consequence, no ΔSOC values above 30 gC kg^{-1} are observed. No clear pattern of ΔSOC is observed with latitude and longitude. The fact that the SOC_{ref} estimates with this method, that uses croplands as references, are more homogeneous compared to the values estimated with the other two methods, that use grasslands and woodlands as references, may be related to the fact that cropland soil are less heterogeneous than forest and grassland soils. Indeed, forest soils can be stonier and vary greatly even on a short spatial scale (Barré et al., 2017).

3.4 Reachability of soil organic carbon reference values estimated with the novel "natural references per pedoclimate" method

We investigated the possibility for a cropland to reach the SOC_{ref} value of its pedoclimatic cluster obtained using this method by assessing whether cropland sites with negative Δ SOC exist in each pedoclimatic cluster. We found that in dataset n. 3bis, 700 croplands over 6777 have negative Δ SOC evaluated with the "natural references per pedoclimate" method and that only four pedoclimatic clusters, 2-0, 5-0, 8-3, and 10-3, do not contain any cropland with negative Δ SOC (Supplementary Figure 5A). This suggests that the SOC_{ref} values estimated using the "natural references per pedoclimate" are generally reachable in croplands. Nevertheless, even though we selected only sites that were classified as croplands in both LUCAS 2009 and LUCAS 2015 campaigns, we cannot discard the possibility that croplands with the highest SOC concentrations per pedoclimate correspond to sites that were under a different land cover before 2009. Determining under what conditions the SOC_{ref} values can be reached is impossible without historic data on land cover, agricultural practices, and associated SOC dynamics.

The pedoclimatic cluster 5-0 contains only one cropland and none of the approaches predict a positive Δ SOC for it, while the pedoclimatic clusters 2-0 and 8-3 contain only seven and eleven croplands, respectively, among which only one and four croplands, respectively, have negative predicted Δ SOC using the "data-driven reciprocal modeling" and the "carbon landscape zones" approaches (Supplementary Figure 5A). Thus, the absence of negative Δ SOC using "natural references per pedoclimate" in the clusters 5-0, 2-0, and 8-3 is most probably due to the small number of occurrences of croplands in these pedoclimatic contexts. For comparison, the median number of croplands per pedoclimatic cluster is 79. Thus, the observed absence of negative

Δ SOC values is likely due to undersampling.

The pedoclimatic cluster 10-3 contains 44 croplands of the dataset n. 3bis. The "data-driven reciprocal modelling" and the "carbon landscape zones" approaches predict negative Δ SOC values in this pedoclimatic context for two and seven croplands, respectively (Supplementary Figure 5A). The pedoclimatic cluster 10-3 includes clay soils mostly located in France, England, Belgium, the Netherlands, and west Germany (Table 1). Supplementary Table 1 reports the statistics for the SOC content in cluster 10-3 for croplands, grasslands, and woodlands. The statistics show that few cases of very high values of SOC content are found among the grasslands and woodlands in this cluster. Relatively high levels of SOC are expected because of the large clay fraction in soil cluster label 3 (Figure 2); however, these extreme cases may raise the reference value SOC_{ref} to a value (61 gC kg⁻¹) that could be unreachable in croplands in the pedoclimatic cluster 10-3. Nevertheless, the minimum Δ SOC observed in this cluster is 0.8 gC kg⁻¹ suggesting that croplands in cluster 10-3 can actually reach such high reference values. It is possible that the "carbon landscape zones" method, that uses croplands as references, is more adapted for the specific case of the pedoclimatic cluster 10-3.

3.5 Comparison of the soil organic carbon reference values estimated with the three methods

The pairwise Spearman correlation coefficients between the SOC_{ref} values estimated with the three methods are all between 0.4 and 0.6 (p-value < 10⁻⁶). Thus, all methods produce weakly but significantly correlated SOC_{ref} estimations.

Some correlation between the SOC_{ref} values estimated with the different methods is expected from their common goal. The fact that the correlation is weak between the

"natural references per pedoclimate", "carbon landscape zones", and "data-driven reciprocal modelling" methods is not surprising because of the different hypotheses underlying the three methods. Indeed, the three methods used different land covers as references, are based on different algorithms and use different pedoclimatic variables.

The estimated reference values SOC_{ref} have similar medians across the three methods (between 26 and 28 gC kg⁻¹ in terms of SOC content or between 64 and 72 Mg ha⁻¹ in terms of SOC stock) and follow the order "carbon landscape zones" < "natural references per pedoclimate" < "data-driven reciprocal modelling" (Figure 3, median values). Accordingly, the total SOC storage potential for the European Union and United Kingdom is 3.1 Gt C, 3.4 Gt C and 4.1 Gt C using the estimations from the "carbon landscape zones", "natural references per pedoclimate", and "data-driven reciprocal modelling" methods, respectively.

Nevertheless, the differences between the "carbon landscape zones", "natural references per pedoclimate", and "data-driven reciprocal modelling" methods are not homogeneous in space. Figure 5 shows the pairwise differences between the reference values SOC_{ref} for the SOC content of croplands evaluated with the three methods. Overall, in central and northern Europe (colder climate) the reference values follow the order "carbon landscape zones" < "natural references per pedoclimate" < "data-driven reciprocal modelling", while in southern Europe (warmer climate) the reference values follow the order "data-driven reciprocal modelling" < "carbon landscape zones" < "natural references per pedoclimate". This shows that the "data-driven reciprocal modelling" is the most sensitive to climate, as it accentuates the expected trend of higher SOC contents in colder climates, where grassland productivity is higher (Smit et al., 2008) and the

SOC mineralization rate might be lower (Clivot et al., 2019). Nevertheless, we observe three main exceptions:

- Brittany (France): the "natural references per pedoclimate" method predicts higher reference values than the "data-driven reciprocal modelling" and the "carbon landscape zones" methods, with differences of more than 20 gC kg^{-1} between the "natural references per pedoclimate" and the "carbon landscape zones" methods (circle in Figure 5B).
- Central Spain (southern Europe): the "carbon landscape zones" method predicts higher reference values than both the "natural references per pedoclimate" and the "data-driven reciprocal modelling" methods (triangle in Figure 5B). The "natural references per pedoclimate" and the "data-driven reciprocal modelling" methods predict similar values (Figure 5A).
- Denmark and north of Germany (northern Europe): the "carbon landscape zones" method predicts higher reference values than the "natural references per pedoclimate" method (rectangle in Figure 5B).

These peculiar cases are discussed in the following section.

3.6 The peculiar cases of Brittany, central Spain, Denmark and the north of Germany.

3.6.1 Brittany

The "natural references per pedoclimate" method predicted larger SOC_{ref} values in Brittany, specially with respect to the "carbon landscape zones" method (Figure 5B). This region belongs to the climate cluster n. 2 and comprises sites of mainly soil cluster n. 2 that is characterized by average pedological features (Table 1). Cluster 2-2 also con-

tains grassland sites located in the United Kingdom and Ireland, as well as woodland sites located in the north of Spain that have particularly high values of SOC content, nearing the threshold of 120 gC kg^{-1} that we used to select mineral soils (Supplementary Figure 6). This results in a high SOC_{ref} value (46 gC kg^{-1}) that may be impossible to reach in Brittany. The "carbon landscape zones" method associated most of Brittany's cropland sites to the carbon landscape zone n. 5, that does not include croplands of Ireland, the United Kingdom, and the north of Spain. The difference most probably comes from the fact that the "carbon landscape zones" method takes into account the NPP, whose values are distinct in Brittany compared to Ireland and the United Kingdom (Supplementary Figure 4C, principal components of NPP). Moreover, the principal components of the climatic decomposition index (CDI) clearly distinguish Brittany from the north of Spain (Supplementary Figure 4C), while the difference between their climates is lost in the pedoclimatic clustering. These results show that the pedoclimatic clustering of the "natural references per pedoclimate" method probably fails to pick-up the peculiar pedoclimate of Brittany. The "data-driven reciprocal modelling" method predicted both moderate and high reference values in Brittany (Figure 3), intermediate between the predictions obtained with the natural references per pedoclimate" and "carbon landscape zones" methods (Figures 4A and 4C). The fact that the "data-driven reciprocal modelling" provides estimates intermediate between the two methods suggests that it may be the least biased. However, it is impossible to establish with certainty which estimates are the closest to reality without experimental data on SOC dynamics under well-characterised agricultural practices in this region. Such data would be particularly valuable, but are currently, to our best knowledge, non-existent.

3.6.2 Central Spain

In central Spain, the SOC_{ref} values estimated using the three methods follow the order "natural references per pedoclimate" ~ "data-driven reciprocal modelling" < "carbon landscape zones", while in most of southern Europe the "natural references per pedoclimate" predicted the highest values (Figure 4). Central Spain is an arid region where irrigation is widely employed (Wriedt et al., 2009). Highly productive irrigated croplands versus unproductive grasslands and woodlands in this area may explain why the "carbon landscape zones" method, which uses best-performing croplands as references, estimated higher SOC storage potentials than the methods that use natural references. In this case, the estimations obtained using the "natural references per pedoclimate" or the "data-driven reciprocal modelling" approaches may be more representative of the potential SOC content reachable without irrigation.

3.6.3 Denmark and the north of Germany

Differently from the rest of northern Europe, the "carbon landscape zones" method predicted higher reference values than the "natural references per pedoclimate" method in Denmark and the north of Germany (rectangle in Figure 5B). Moreover, the "natural references per pedoclimate" and the "data-driven reciprocal modelling" approaches yielded estimations that differ by more than 20 gC kg⁻¹ in this area (Figure 5A). Recent work has highlighted that black sands soils, characterised by sand content > 80% and C:N ratio > 13 and present in the north of Germany, Denmark, Belgium, and the Netherlands, should be treated separately when assessing SOC storage potentials because of their unusually high SOC contents due to historic heathland and peatland land cover (Poeplau et al., 2020; Schneider, Amelung and Don, 2021; Dexlet et al., 2022). In the

dataset n.3bis, we find 66 black sands soils according to this definition (sand content > 80% and C:N ratio > 13), most of which are in Denmark and the north of Germany (Supplementary Figure 7).

A local model adapted to the specificity of black sands soils may be needed to correctly assess SOC reference values in this area, for example the "data-driven reciprocal modelling" method in its original implementation that includes the C:N ratio as a predictive feature (Schneider, Poeplau and Don, 2021), or the benchmarking of German Agricultural Soils Inventory proposed by Drexler et al. (2022) where black sands soils are isolated in a specific stratum.

3.7 Ensemble modelling

The examples reported in the previous sections (pedoclimatic cluster 10-3, Brittany, Central Spain, Denmark, and North of Germany) show that the hypotheses underlying each method have a potentially large influence on the estimates of the local SOC storage potentials. These peculiar cases could be identified using the comparison across different methods, while the use of a unique method would have not allowed to pinpoint possible over- or underestimations of the SOC storage potentials.

Determining which model performs the best is hard, if not impossible, as validating the SOC storage potential estimates would require monitoring SOC dynamics in the long-run across different pedoclimates and under different land management practices with high spatial accuracy. Reliable data of this kind are scarce at local scale and non-existent at the scale of Europe.

Nevertheless, farmers, policymakers and stakeholders need quantifiable SOC targets that can be widely applied and are ambitious yet reachable. Perhaps, in the absence of

any possible validation procedure, the most conservative approach is to evaluate the SOC reference values using different methods and seek consensus, e.g. by taking the median estimate. This is an ensemble modelling technique that is generally used to provide robustness to the estimations and has been recently applied, for instance, to the estimation of the SOC dynamics in long-term bare-fallow soils (Farina et al., 2021).

Figure 6A-D show the reference values SOC_{ref} and the corresponding SOC storage potentials estimates ΔSOC for the LUCAS croplands evaluated as the median of the estimates from the "natural references per pedoclimate", "data-driven reciprocal modelling", and "carbon landscape zones" methods. This corresponds to always selecting the intermediate of the three estimated values. We found that each method was selected for around a third of the LUCAS croplands, with no evident spatial pattern (Figure 6E).

The proposed ensemble modelling approach has the advantage of discarding extremely high and low estimates from single methods that are probably erroneous. As an example, the estimates from the "data-driven reciprocal modelling" approach are kept for almost all croplands in Brittany (Figure 6E), where the "natural references per pedoclimate" and the "carbon landscape zones" approaches provide highly divergent values (Figure 5). Similarly, the estimates from the "carbon landscape zones" approach were discarded in the ensemble modelling results for central Spain (Figure 6E), where using croplands as references may bias the SOC_{ref} estimations to higher values due to frequent irrigation (section 3.6.2). On the contrary, the estimates from the "carbon landscape zones" approach are kept in Denmark and the north of Germany (Figure 6E), where the estimates from the other methods diverge (Figure 5).

3.8 Application to long-term experiment sites

Despite the impossibility of validating the SOC_{ref} estimations at large scale, experi-

mental data from long-term experiments are useful to provide insights on whether estimated SOC_{ref} values are reachable under specific but well-characterized conditions. To this goal, we have applied the ensemble modelling approach to the estimation of target SOC_{ref} values for two long-term experiment sites: the Ultuna site in Sweden (Kätterer et al., 2011) and the Broadbalk site in the United Kingdom (Rothamsted Research, 2021). Then, we calculated the ratio between the SOC content measured in the experimental plots under different treatments and the reference value SOC_{ref} (SOC data from 2009 for Ultuna (Kätterer et al., 2011) and from 2015 for Broadbalk (Rothamsted Research, 2021)).

We found that the estimated SOC_{ref} is reached in Ultuna under peat amendment with and without nitrogen fertilization (ratio > 1) and is approached under sewage sludge amendment, farmyard manure amendment, and sawdust amendment with nitrogen fertilization (ratio = 0.85, 0.71 and 0.70, respectively). Similarly, the estimated SOC_{ref} value is approached in Broadbalk under continued farmyard manure (ratio = 0.95 when applied since 1843 and ratio = 0.85 when applied since 1885 with nitrogen fertilization). The results for all treatments for both sites are reported in Supplementary Table 2.

These results suggest that the estimated SOC_{ref} values may not be reachable unless high organic matter input is provided over several decades, at least in the pedoclimatic contexts to which the Ultuna and Broadbalk sites belong (pedoclimatic clusters 1-2 and 10-2, respectively). Nevertheless, these preliminary results should be confirmed through the analysis of a dataset of long-term experiments data from different pedoclimatic context and it would also be important to evaluate the method outside LTEs, which are not necessarily representative of real farm conditions. If these results are confirmed, it will

probably be necessary to propose target SOC values that are lower than the SOC_{ref} values calculated here.

3.9 A first rough estimate of soil organic carbon storage potential estimation in European croplands using our novel framework

Across all EU27 countries and the United Kingdom, we obtain a rough estimate of 3.5 Gt C of SOC storage potential in the cropland topsoils using the ensemble modelling framework.

In our work, we used the sample medians of the LUCAS sites per country as estimates of the subpopulation medians assuming a random sampling design. However, LUCAS points location were not selected using a probability sampling design (de Gruijter et al., 2006) which do not allow valid estimation using designed-based estimates. Further work will need to implement model-dependent inference using for instance a spatial model (kriging) (Teuling et al., 2021). Nevertheless, to our knowledge, no unbiased estimate of total SOC stocks at the EU scale and its associate uncertainty has been produced yet. This absence is mainly attributable to the heterogeneity of sampling designs between countries and to the lack of a EU-wide probability samples. However, Some European countries having implemented a probability sampling design (for example, a systematic random sampling like a grid, e.g. France (Chen et al., 2018), Germany (Grüneberg et al., 2014; Poeplau et al., 2020), Denmark (Adhikari et al., 2014), England and Wales (Bellamy et al., 2005), or stratified simple random sampling, e.g. the Netherlands (Knotter et al., 2022)), are able to produce a national designed-based estimates of the population mean value or a population total stock together with a reliable and valid estimate of their uncertainty.

According to current estimations of SOC storage rates, e.g. 41-95 Mt C yr⁻¹ in the European Union (Bellassen et al., 2022) or 16-19 Mt C yr⁻¹ for EU15 countries only (Freibauer et al., 2004), the SOC storage potential of 3.5 Gt C issued from our ensemble modelling approach could be reached in 37 to 85 years if practices including land-use change (agroforestry, cover crops, substituting fodder crops with grass) were fully adopted in the European Union and the United Kingdom. However, much longer times would be needed to reach the potential if less ambitious practices are adopted. For instance, this potential could be reached in almost 300 years with only the introduction of cover crops (Bellassen et al., 2022), under the condition that the storage rate is maintained. These numbers represent bio-physical estimates only and do not take into consideration any social, political, and economic constraints associated with the adoption of appropriate agricultural practices and land-use (Arunderson and Biardeau, 2018).

Thus, the 3.5 Gt C potential for SOC storage in European croplands can only be considered as a long-term aspirational target that would be reachable only under extreme changes in agricultural practices, including land-cover or land-use change. This is not surprising, as the ensemble modelling approach includes the "data Driven reciprocal modelling" method that was developed to predict changes in SOC upon conversion to grassland, and the "natural references per pedoclimate" that uses grasslands and woodlands as references. However, the "carbon landscape zones" method, that uses croplands as references, predicts higher SOC reference values in southern Europe, showing that the modelling choices impact on the estimations is context-dependent. This is consistent with each method being kept for around a third of the croplands in the ensemble modelling (Figure 6E).

Nevertheless, attention should be paid to the estimates expressed as SOC stocks, that are of more practical significance than SOC contents, yet are highly uncertain when bulk density measurements are not available. Indeed, we needed to perform several approximations when converting SOC contents to stocks: the use of a pedotransfer function to estimate the bulk density, the use of a constant coarse fraction density for the conversion of coarse mass fractions to coarse volume fractions, and not accounting for the presence of stones, whose fraction is not available in the LUCAS Topsoil 2009 and 2015 datasets. Additionally, when SOC storage potentials Δ SOC are estimated relative to SOC measurements in best-performing sites, observing non-null Δ SOC stocks may be due to actual differences in SOC content but also to differences in bulk density and/or coarse fragments content.

Because of the discussed limitations, the 3.5 Gt C estimate is an indicator of the order of magnitude of SOC storage potential in Europe, however its exact value should be handled with care.

3.10 Future developments

In the future, our results should be challenged against long-term experiments and real farm data belonging to different pedoclimates and associated to detailed land-management practices. Moreover, the estimation of the associated uncertainty should be included and the robustness of the estimations could be further improved by adding to the pool other methods that are based on different hypotheses and other scientific knowledge. It would also be possible to include local models that could participate in the ensemble modelling only in their region of applicability. National inventories, e.g. the RMQS in France (Jolivet et al., 2006) or the German Agricultural Soil Inventory (Poeplau et al., 2020), have the advantage of generally including bulk density meas-

urements as well as both topsoil and subsoil SOC contents and could provide more accurate estimations of SOC stocks potentials. Moreover, local models may be necessary for peculiar types of soils (e.g. the black sands soils in Denmark and the north of Germany), to compensate for artifacts introduced by human activity (e.g. irrigation in arid regions such as central Spain) and to take into account the fact that the rate of SOC storage (or loss) upon changes in management practices depends on the pedoclimatic context (Lugato, Bampa, Panagos et al., 2014; Chenu et al., 2019).

Finally, the same approach could be applied to estimate SOC storage potentials at deeper soil depth, in cases where SOC content and/or stocks are measured along the soil profile. This is the case for example for the German Agricultural Soil Inventory (Poeplau et al., 2020).

4. Conclusion

In summary, we found that different approaches based on different hypotheses, provided generally comparable estimates of SOC_{ref} values but with local discrepancies. The proposed ensemble modelling framework allows combining estimates from different approaches. It discards extreme estimations from single methods, which are likely erroneous. Using the ensemble modelling output and approximate bulk density values, we obtained a first rough estimate of 3.5 Gt C of SOC storage potential in the European cropland topsoils. Nevertheless, this is probably an overestimation that should be interpreted as a long-term aspirational target. The use of additional methods and datasets in the ensemble modelling approach and more valid statistical spatial estimates may further refine our approach.

Acknowledgements

The authors thank Denis Angers for his advice on the first version of this manuscript and the anonymous reviewers for their constructive comments that have improved the manuscript.

Funding

This study was supported by France Relance - project CARBOTOPIY. We do not have any conflicts of interest to declare.

Data availability statement

The data that support the findings of this study are openly available in ESDAC and Eurostat at the following url's. LUCAS 2009/12 Topsoil³¹: <https://esdac.jrc.ec.europa.eu/content/lucas-2009-topsoil-data>; LUCAS 2015 Topsoil³²: <https://esdac.jrc.ec.europa.eu/content/lucas2015-topsoil-data>; LUCAS 2009 land cover data³³: <https://ec.europa.eu/eurostat/web/lucas/data/primary-data/2009>; LUCAS 2012 land cover data³⁴: <https://ec.europa.eu/eurostat/web/lucas/data/primary-data/2012>; LUCAS 2015 land cover data³⁵: <https://ec.europa.eu/eurostat/web/lucas/data/primary-data/2015>.

Code availability

The code is available at https://github.com/greenbackdev/soc_targets_europe.

References

Adhikari, K., Hartemink, A. E., Minasny, B., Bou Kheir, R., Greve, M. B., Greve, M. H., 2014. Digital Mapping of Soil Organic Carbon Contents and Stocks in Denmark. PLoS ONE, 9 (8), e105519. <https://doi.org/10.1371/journal.pone.0105519>.

Amundson, R., Biardeau, L., 2018. Soil Carbon Sequestration Is an Elusive Climate Mitigation Tool. *Proc. Natl. Acad. Sci.*, 115 (46), 11652–11656. <https://doi.org/10.1073/pnas.1815901115>.

Arrouays, D., Saby, N., Walter, C., Lemercier, B., Schwartz, C., 2006. Relationships between Particle-Size Distribution and Organic Carbon in French Arable Topsoils. *Soil Use Manag.*, 22 (1), 48–51. <https://doi.org/10.1111/j.1475-2743.2006.00020.x>.

Barre, P., Durand, H., Chenu, C., Meunier, P., Montagne, D., Castel, G., Billiou, D., Soucemarianadin, L. and Cecillon, L., 2017. Geological control of soil organic carbon and nitrogen stocks at the landscape scale. *Geoderma*, 285, pp.50-56.

Bellamy, P. H., Loveland, P. J., Bradley, R. I., Lark, R. M., Kirk, G. J. D., 2005. Carbon Losses from All Soils across England and Wales 1978–2003. *Nature*, 437 (7056), 245–248. <https://doi.org/10.1038/nature04048>.

Bellassen, V., Angers, D., Kowalczyński, T., Olesen, A., 2022 Soil Carbon Is the Blind Spot of European National GHG Inventories. *Nat. Clim. Change*, 12 (4), 324–331. <https://doi.org/10.1038/s41558-022-01321-9>.

Bispo, A., Andersen, L., Angers, D. A., Bernoux, M., Brossard, M., Cécillon, L., Comans, R. N. J., Harmsen, J., Jonassen, K., Lamé, F., Lhuillery, C., Maly, S., Martin, E., McElnea, A. E., Sakai, H., Watabe, Y., Eglin, T. K., 2017. Accounting for Carbon Stocks in Soils and Measuring GHGs Emission Fluxes from Soils: Do We Have the Necessary Standards? *Front. Environ. Sci.*, 5, 41. <https://doi.org/10.3389/fenvs.2017.00041>.

de Brogniez, D., Ballabio, C., Stevens, A., Jones, R. J. A., Montanarella, L., & van Wesemael, B., 2015. A map of the topsoil organic carbon content of Europe generated by a generalized additive model. *European Journal of Soil Science*, 66(1), 121-134.

Bonfante, A., Basile, A., Bouma, J., 2020. Targeting the Soil Quality and Soil Health Concepts When Aiming for the United Nations Sustainable Development Goals and the EU Green Deal. *SOIL*, 6 (2), 453–466. <https://doi.org/10.5194/soil-6-453-2020>.

Büchi, L., Walder, F., Banerjee, S., Colombi, T., van der Heijden, M. G. A., Keller, T., Charles, R., Six, J., 2022. Pedoclimatic Factors and Management Determine Soil Organic Carbon and Aggregation in Farmer Fields at a Regional Scale. *Geoderma*, 409, 115632. <https://doi.org/10.1016/j.geoderma.2021.115632>.

Chabbi, A., Lehmann, J., Ciais, P., Loescher, H. W., Cotruac, M. F., Don, A., SanClements, M., Schipper, L., Six, J., Smith, P., Rumpel, C., 2017. Aligning Agriculture and Climate Policy. *Nat. Clim. Change*, 7 (5), 307–309. <https://doi.org/10.1038/nclimate3286>.

Chen, S., Martin, M. P., Saby, N. P. A., Walter, C., Angers, D. A., Arrouays, D., 2018. Fine Resolution Map of Top- and Subsoil Carbon Sequestration Potential in France. *Sci. Total Environ.*, 630, 389–400. <https://doi.org/10.1016/j.scitotenv.2018.02.209>.

Chen, S., Arrouays, D., Angers, D. A., Chenu, C., Barré, P., Martin, M. P., Saby, N. P. A., Walter, C., 2019. National Estimation of Soil Organic Carbon Storage Potential for Arable Soils: A Data-Driven Approach Coupled with Carbon-Landscape Zones. *Sci. Total Environ.*, 666, 355–367. <https://doi.org/10.1016/j.scitotenv.2019.02.249>.

Chen, S., Arrouays, D., Angers, D. A., Martin, M. P., Walter, C., 2019. Soil Carbon Stocks under Different Land Uses and the Applicability of the Soil Carbon Saturation Concept. *Soil Tillage Res.*, 188, 53–58. <https://doi.org/10.1016/j.still.2018.11.001>.

Chenu, C., Angers, D. A., Barré, P., Derrien, D., Arrouays, D., Balesdent, J., 2019. Increasing Organic Stocks in Agricultural Soils: Knowledge Gaps and Potential Innova-

tions. *Soil Tillage Res.*, 188, 41–52. <https://doi.org/10.1016/j.still.2018.04.011>.

Clivot, H., Mouny, J.-C., Duparque, A., Dinh, J.-L., Denoroy, P., Houot, S., Vertès, F., Trochard, R., Bouthier, A., Sagot, S., Mary, B., 2019. Modeling Soil Organic Carbon Evolution in Long-Term Arable Experiments with AMG Model. *Environ. Model. Softw.*, 118, 99–113. <https://doi.org/10.1016/j.envsoft.2019.04.004>.

Comerford, N. B., Franzluebbers, A. J., Stromberger, M. E., Morris, L., Markewitz, D., Moore, R., 2013. Assessment and Evaluation of Soil Ecosystem Services. *Soil Horiz.* 54 (3), 0. <https://doi.org/10.2136/sh12-10-0028>.

de Gruijter, J. J., Bierkens, M. F. P., Brus, D. J., Knoters, M., 2006. Sampling for Natural Resource Monitoring, Eds., Springer: Berlin, Heidelberg, pp 13–25. https://doi.org/10.1007/3-540-33161-1_2.

Drexler, S., Broll, G., Flessa, H., Don, A., 2022. Benchmarking Soil Organic Carbon to Support Agricultural Carbon Management: A German Case Study#. *J. Plant Nutr. Soil Sci.*, 185 (3), 427–440. <https://doi.org/10.1002/jpln.202200007>.

Dutta, H., Dutta, A., 2016. The Microbial Aspect of Climate Change. *Energy Ecol. Environ.*, 1 (4), 209–232. <https://doi.org/10.1007/s40974-016-0034-7>.

ESDAC, 2013. European Soil Data Centre (ESDAC), European Commission, Joint Research Centre. LUCAS 2009 TOPSOIL Data. <https://esdac.jrc.ec.europa.eu/content/lucas-2009-topsoil-data>.

ESDAC, 2020. European Soil Data Centre (ESDAC), European Commission, Joint Research Centre. LUCAS 2015 TOPSOIL Data. <https://esdac.jrc.ec.europa.eu/content/lucas2015-topsoil-data>.

European Commission, 2013. European Commission, Joint Research Centre. Institute

for Environment and Sustainability. LUCAS Topsoil Survey: Methodology, Data and Results. Publications Office: LU.

European Commission, 2020. European Commission, Joint Research Centre. LUCAS 2015 Topsoil Survey: Presentation of Dataset and Results. Publications Office: LU.

EUROSTAT, 2009. European Commission, Joint Research Centre. LUCAS Micro Data 2009. <https://ec.europa.eu/eurostat/web/lucas/data/primary-data/2009>.

EUROSTAT, 2012. European Commission, Joint Research Centre. LUCAS Micro Data 2012. <https://ec.europa.eu/eurostat/web/lucas/data/primary-data/2012>.

EUROSTAT, 2015. European Commission, Joint Research Centre. LUCAS Micro Data 2015. <https://ec.europa.eu/eurostat/web/lucas/data/primary-data/2015>.

Farina, R., Sándor, R., Abdalla, M., Álvarez-Fuentes, J., Bechini, L., Bolinder, M. A., Brilli, L., Chenu, C., Clivot, H., De Antoni Migliorati, M., Di Bene, C., Dorich, C. D., Ehrhardt, F., Ferchaud, F., Fitton, N., Francaviglia, R., Franko, U., Giltrap, D. L., Grant, B. B., Guenet, B., Harrison, M. T., Hirschbaum, M. U. F., Kuka, K., Kulmala, L., Liski, J., McGrath, M. J., Meier, E., Monicetti, L., Moyano, F., Nendel, C., Recous, S., Reibold, N., Shepherd, A., Smith, W. N., Smith, P., Soussana, J., Stella, T., Taghizadeh-Toosi, A., Tsutsikh, E., Benocchi, G., 2021. Ensemble Modelling, Uncertainty and Robust Predictions of Organic Carbon in Long-term Bare-fallow Soils. *Glob. Change Biol.*, 27 (4), 904–928. <https://doi.org/10.1111/gcb.15441>.

Fernández-Ugalde, O., Tóth, G., 2017. Pedotransfer Functions for Predicting Organic Carbon in Subsurface Horizons of European Soils: Predicting Organic Carbon in Subsurface Horizons in Europe. *Eur. J. Soil Sci.*, 68 (5), 716–725. <https://doi.org/10.1111/ejss.12464>.

Freibauer, A., Rounsevell, M. D. A., Smith, P., Verhagen, J., 2004. Carbon Sequestration in the Agricultural Soils of Europe. *Geoderma*, 122 (1), 1–23. <https://doi.org/10.1016/j.geoderma.2004.01.021>.

Garnier, J., Billen, G., Tournebise, J., Barré, P., Mary, B., Baudin, F., 2022. Storage or Loss of Soil Active Carbon in Cropland Soils: The Effect of Agricultural Practices and Hydrology. *Geoderma*, 407, 115538. <https://doi.org/10.1016/j.geoderma.2021.115538>.

Georgiou, K., Jackson, R. B., Vindušková, O., Abramoff, R. Z., Ahlström, A., Feng, W., Harden, J. W., Pellegrini, A. F. A., Polley, H. W., Soong, J. L., Riley, W. J., Torn, M. S., 2022. Global Stocks and Capacity of Mineral-Associated Soil Organic Carbon. *Nat. Commun.*, 13 (1), 3797. <https://doi.org/10.1038/s41467-022-31540-9>.

Grüneberg, E., Ziche, D., Wellbrock, N., 2014. Organic Carbon Stocks and Sequestration Rates of Forest Soils in Germany. *Glob. Change Biol.*, 20 (8), 2644–2662. <https://doi.org/10.1111/gcb.12558>.

Guillaume, T., Makowski, D., Liholova, Z., Bragazza, L., Sallaku, F., Sinaj, S., 2022. Soil Organic Carbon Saturation in Cropland-Grassland Systems: Storage Potential and Soil Quality. *Geoderma*, 407, 115529. <https://doi.org/10.1016/j.geoderma.2021.115529>.

Harris, I., Osborn, T. J., Jones, P., Lister, D., 2020. Version 4 of the CRU TS Monthly High-Resolution Gridded Multivariate Climate Dataset. *Sci. Data*, 7 (1), 109. <https://doi.org/10.1038/s41597-020-0453-3>.

Hassink, J., 1997. The Capacity of Soils to Preserve Organic C and N by Their Association with Clay and Silt Particles. *Plant Soil*, 191 (1), 77–87. <https://doi.org/10.1023/A:1004213929699>.

Heikkinen, J., Ketoja, E., Nuutinen, V., & Regina, K., 2013. Declining trend of

carbon in Finnish cropland soils in 1974–2009. *Global change biology*, 19(5), 1456-1469.

Hollis, J. M., Hannam, J., Bellamy, P. H., 2012. Empirically-Derived Pedotransfer Functions for Predicting Bulk Density in European Soils. *Eur. J. Soil Sci.*, 63 (1), 96–109. <https://doi.org/10.1111/j.1365-2389.2011.01412.x>.

IPBES, 2018. Summary for Policymakers of the Assessment Report on Land Degradation and Restoration of the Intergovernmental SciencePolicy Platform on Biodiversity and Ecosystem Services.

Johannes, A., Matter, A., Schulin, R., Weiskopf, P., Raveye, P. C., Boivin, P., 2017. Optimal Organic Carbon Values for Soil Structure Quality of Arable Soils. Does Clay Content Matter? *Geoderma*, 302, 14–21. <https://doi.org/10.1016/j.geoderma.2017.04.021>.

Jolivet, C. C., Boulonne, L., Ratié, C., 2006. Manuel du Réseau de Mesures de la Qualité des Sols (RMQS), 188.

Kätterer, T., Bolinder, M. A., Andrén, O., Kirchmann, H., Menichetti, L., 2011. Roots Contribute More to Refractory Soil Organic Matter than Above-Ground Crop Residues, as Revealed by a Long-Term Field Experiment. *Agriculture, Ecosystems & Environment*, 141 (1–2), 184–192. <https://doi.org/10.1016/j.agee.2011.02.029>.

Knotters, M., Teuling, K., Reijneveld, A., Lesschen, J. P., Kuikman, P., 2022. Changes in Organic Matter Contents and Carbon Stocks in Dutch Soils, 1998–2018. *Geoderma*, 414, 115751. <https://doi.org/10.1016/j.geoderma.2022.115751>.

Lal, R., 2004a. Soil Carbon Sequestration to Mitigate Climate Change. *Geoderma*, 123 (1–2), 1–22. <https://doi.org/10.1016/j.geoderma.2004.01.032>.

Lal, R., 2004b. Soil Carbon Sequestration Impacts on Global Climate Change and Food Security. *Science*, 304, 6.

Loveland, P., 2003. Is There a Critical Level of Organic Matter in the Agricultural Soils of Temperate Regions: A Review. *Soil Tillage Res.*, 70 (1), 1–18. [https://doi.org/10.1016/S0167-1987\(02\)00139-3](https://doi.org/10.1016/S0167-1987(02)00139-3).

Lugato, E., Bampa, F., Panagos, P., Montanarella, L., Jones, A., 2014. Potential Carbon Sequestration of European Arable Soils Estimated by Modelling a Comprehensive Set of Management Practices. *Glob. Change Biol.*, 20 (11), 3557–3567. <https://doi.org/10.1111/gcb.12551>.

Lugato, E., Panagos, P., Bampa, F., Jones, A., Montanarella, L., 2014. A New Baseline of Organic Carbon Stock in European Agricultural Soils Using a Modelling Approach. *Glob. Change Biol.*, 20 (1), 313–326. <https://doi.org/10.1111/gcb.12292>.

Macías, F., Camps Arbestain, M., 2010. Soil Carbon Sequestration in a Changing Global Environment. *Mitig. Adapt. Strateg. Glob. Change*, 15 (6), 511–529. <https://doi.org/10.1007/s11027-010-9231-4>.

Martin M., Wattenbach M., Smith, P., Meersmans, J., Jolivet, C., Boulonne, L., Arrouays, D., 2011, Spatial distribution of soil organic carbon stocks in France. *Biogeosciences*, 8, 1053–1065

Montanarella, L., Panagos, P., 2021. The Relevance of Sustainable Soil Management within the European Green Deal. *Land Use Policy*, 100, 104950. <https://doi.org/10.1016/j.landusepol.2020.104950>.

Oldfield, E. E., Eagle, A. J., Rubin, R. L., Rudek, J., Sanderman, J., Gordon, D. R., 2022. Crediting Agricultural Soil Carbon Sequestration. *Science*, 375 (6586), 1222–

1225. <https://doi.org/10.1126/science.abl7991>.

Orgiazzi, A., Ballabio, C., Panagos, P., Jones, A., Fernández-Ugalde, O., 2018. LUCAS Soil, the Largest Expandable Soil Dataset for Europe: A Review. *Eur. J. Soil Sci.*, 69 (1), 140–153. <https://doi.org/10.1111/ejss.12499>.

Pacini, L., Yunta, F., Jones, A., Montanarella, L., Barrè, P., Saia, S., Chen, S., Schillaci, C., 2023. Fine Earth Soil Bulk Density at 0.2m Depth from Land Use and Coverage Area Frame Survey (LUCAS) Soil 2018. *Eur. J. Soil Sci.*, e13391. <https://doi.org/10.1111/ejss.13391>.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, O., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.*, 12 (85), 2825–2830.

Poeplau, C., Don, A., Vesterdal, L., Leifeld, J., Van Wesemael, B., Schumacher, J., Gensior, A., 2011. Temporal Dynamics of Soil Organic Carbon after Land-Use Change in the Temperate Zone - Carbon Response Functions as a Model Approach: SOIL ORGANIC CARBON AND LAND-USE CHANGE. *Glob. Change Biol.*, 17 (7), 2415–2427. <https://doi.org/10.1111/j.1365-2486.2011.02408.x>.

Poeplau, C., Jacobs, A., Don, A., Vos, C., Schneider, F., Wittnebel, M., Tiemeyer, B., Heidkamp, A., Prietz, R., Flessa, H., 2020. Stocks of Organic Carbon in German Agricultural Soils—Key Results of the First Comprehensive Inventory. *J. Plant Nutr. Soil Sci.*, 183 (6), 665–681. <https://doi.org/10.1002/jpln.202000113>.

Powlson, D. S., Whitmore, A. P., Goulding, K. W. T., 2011. Soil Carbon Sequestration to Mitigate Climate Change: A Critical Re-Examination to Identify the True and the False.

Eur. J. Soil Sci., 62 (1), 42–55. <https://doi.org/10.1111/j.1365-2389.2010.01342.x>.

Powlson, D. S., Poulton, P. R., Glendining, M. J., Macdonald, A. J., Goulding, K. W. T., 2022. Is It Possible to Attain the Same Soil Organic Matter Content in Arable Agricultural Soils as under Natural Vegetation? *Outlook Agric.*, 51 (1), 91–104. <https://doi.org/10.1177/00307270221082113>.

Rasmussen, C., Heckman, K., Wieder, W. R., Keiluweit, M., Lawrence, C. R., Berhe, A. A., Blankinship, J. C., Crow, S. E., Druhan, J. L., Hicks Pries, C. E., Marin-Spiotta, E., Plante, A. F., Schädel, C., Schimel, J. P., Sierra, C. A., Thompson, A., Wagai, R., 2018. Beyond Clay: Towards an Improved Set of Variables for Predicting Soil Organic Matter Content. *Biogeochemistry*, 137 (3), 297–306. <https://doi.org/10.1007/s10533-018-0424-3>.

Rothamsted Research, 2021. Broadbalk soil organic carbon content 1843-2015 Electronic Rothamsted Archive, <https://doi.org/10.23637/KeyRefOABKsoc-02>.

Roussel, O., Bourmeau, E., 2001. Évaluation du déficit en matière organique des sols français et des besoins potentiels en amendements organiques. *Étude Gest. Sols*, 18.

Running, S., Mu, Q., Zhen, M., 2015. MOD17A2H MODIS/Terra Gross Primary Productivity 8-Day L4 Global 500m SIN Grid V006. <https://doi.org/10.5067/MODIS/MOD17A2H.006>.

Sanderman, J., Hengl, T., Fiske, G. J., 2017. Soil Carbon Debt of 12,000 Years of Human Land Use. *Proc. Natl. Acad. Sci.*, 114 (36), 9575–9580. <https://doi.org/10.1073/pnas.1706103114>.

Schneider, F., Amelung, W., Don, A., 2021. Origin of Carbon in Agricultural Soil Profiles Deduced from Depth Gradients of C:N Ratios, Carbon Fractions, $\Delta^{13}\text{C}$ and $\Delta^{15}\text{N}$ Val-

ues. *Plant Soil*, 460 (1–2), 123–148. <https://doi.org/10.1007/s11104-020-04769-w>.

Schneider, F., Poeplau, C., Don, A., 2021. Predicting Ecosystem Responses by Data-driven Reciprocal Modelling. *Glob. Change Biol.*, 27 (21), 5670–5679. <https://doi.org/10.1111/gcb.15817>.

Shukla, P. R., Skea, J., Buendia, E. C., Masson-Delmotte, V., Pörtner, H.-O., Roberts, D. C., Zhai, P., Slade, R., Connors, S., Diemen, R. van, Ferrat, M., Haughey, E., Luz, S., Neogi, S., Pathak, M., Petzold, J., Pereira, J. P., Vyas, P., Huntley, E., Kissick, K., Belkacemi, M., Malley, J., 2019. IPCC, 2019: 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.

Smit, H. J., Metzger, M. J., Ewert, F., 2008. Spatial Distribution of Grassland Productivity and Land Use in Europe. *Agric. Syst.*, 98 (3), 208–219. <https://doi.org/10.1016/j.agry.2008.07.004>.

Smith, P., 2004. Carbon Sequestration in Croplands: The Potential in Europe and the Global Context. *Eur. J. Agron.*, 20 (3), 229–236. <https://doi.org/10.1016/j.eja.2003.08.002>.

Teuling K., Kempen B., Knotters M., Saby N., Brus D., Vašát R., van Egmond F., Bispo A., 2021. Sampling theory for mapping and monitoring purposes. In van Egmond F.M, Fantappiè M,. Report on harmonized procedures for creation of databases and maps (6.1). EJP SOIL.

https://ejpsoil.eu/fileadmin/projects/ejpsoil/WP6/EJP_SOIL_D6.1_Report_on_harmonized_procedures_for_creation_of_databases_and_maps_final.pdf

Van Wesemael, B., Paustian, K., Meersmans, J., Goidts, E., Barancikova, G., & Easter,

M., 2010. Agricultural management explains historic changes in regional soil carbon stocks. *Proceedings of the National Academy of Sciences*, 107(33), 14926-14930.

Wood, S. A., Blankinship, J. C., 2022. Making Soil Health Science Practical: Guiding Research for Agronomic and Environmental Benefits. *Soil Biol. Biochem.*, 172, 108776. <https://doi.org/10.1016/j.soilbio.2022.108776>.

Wriedt, G., van der Velde, M., Aloe, A., Bouraoui, F., 2009. A European Irrigation Map for Spatially Distributed Agricultural Modelling. *Agric. Water Manag.*, 96 (5), 771–789. <https://doi.org/10.1016/j.agwat.2008.10.012>.

Table captions

Table 1: Interpretation of the pedoclimatic clusters. Geographical regions to which the LUCAS sites of each climatic cluster belong and description of the soil clusters.

Climate cluster number	Geographical region
0	North Spain; South France; Central and South Italy; Greece.
1	South Sweden; Estonia; Latvia; Lithuania; East Poland; Czech Republic; border between the Czech Republic and Germany, Austria and Poland; Slovakia; North and Central Romania; few sites in Bulgaria and Greece.
2	Atlantic coast of Spain and Portugal; Brittany; Ireland; Scotland; Wales; North and West England.
3	Central and East France; South Belgium; south-west Germany; North Italy; few sites in continental Greece.
4	Germany; Denmark; South Sweden; Poland; few sites in Latvia and Lithuania.
5	Sweden; Finland; few sites in Estonia, Latvia and Lithuania.
6	South Portugal; South Spain; Corse; Sardinia; Sicily; Crete.
7	Central Spain.
8	North Italy; Slovenia; North Greece; West Bulgaria; few sites in France,

	Sweden, Austria, Czech Republic, Slovakia, Poland and Romania.
9	Hungary; Romania; Bulgaria; few sites in Slovakia and Greece.
10	France; England; Belgium; Netherlands; West Germany; few sites in Italy.
Soil cluster number	Description
0	Calcareous soils.
1	Sandy soils.
2	Soils with average pedological features.
3	Clay soils.

Table 2: Carbonates content of LUCAS sites depending on the soil cluster. Std: standard deviation; Q1: first quartile; Q3: third quartile. Values are expressed in [gCaCO₃/kg] and rounded to the nearest integer.

Soil cluster	Number of sites	Mean	Std	Min	Q1	Median	Q3	Max
0	1845	345	153	37	228	315	438	976
1	6284	4	17	0	0	0	1	241
2	6072	16	34	0	0	1	9	206
3	1050	27	54	0	0	1	17	259

Figure captions

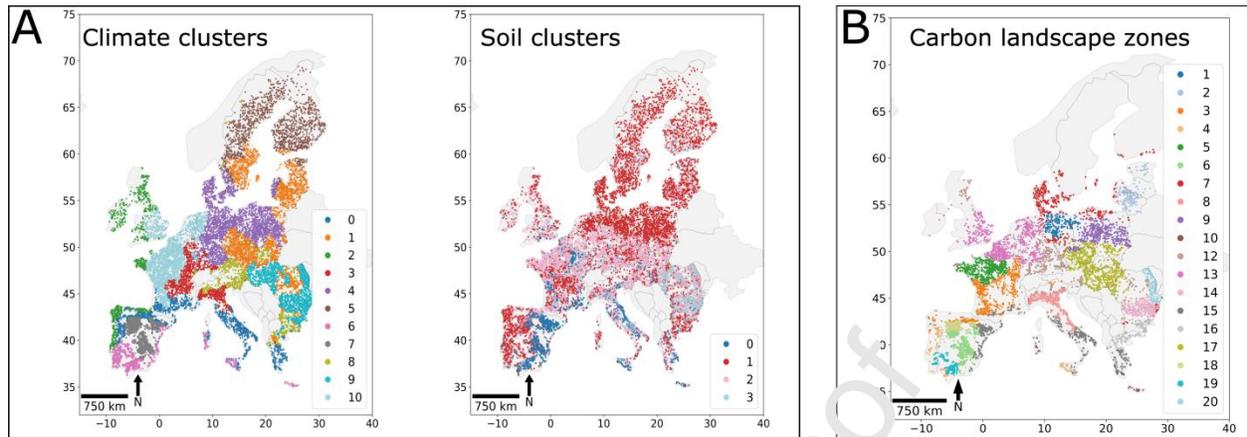
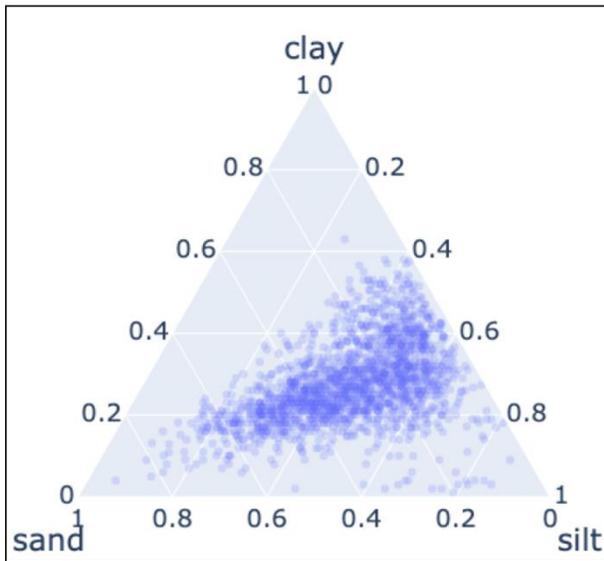
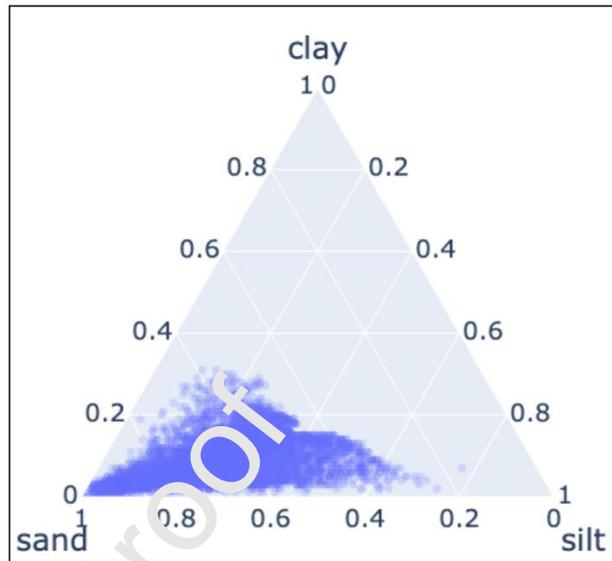


Figure 1: Pedoclimatic clusters and carbon landscape zones of the LUCAS dataset. A: Pedoclimatic clusters of the LUCAS dataset (natural references per pedoclimate approach). Each point represents a LUCAS site. **1:** The sites are colored according to the climate cluster label. **2:** The sites are colored according to the soil cluster label. **B: Carbon landscape zones of the LUCAS croplands dataset (carbon landscape zones approach).** Each point represents a LUCAS cropland site. Refer to the electronic version for colors.

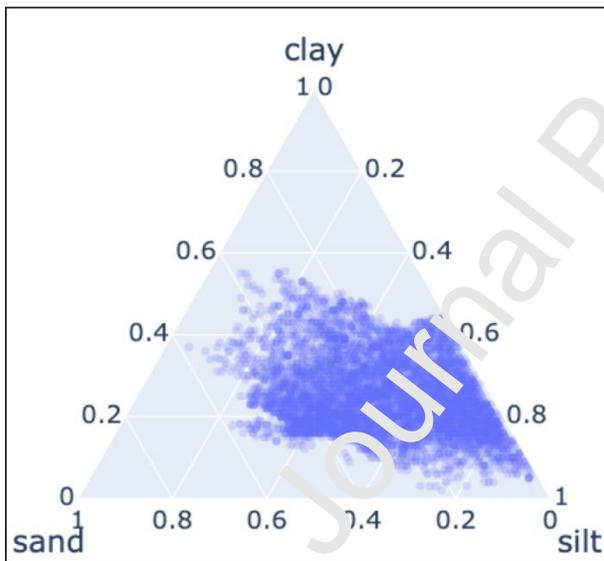
Cluster soil 0



Cluster soil 1



Cluster soil 2



Cluster soil 3

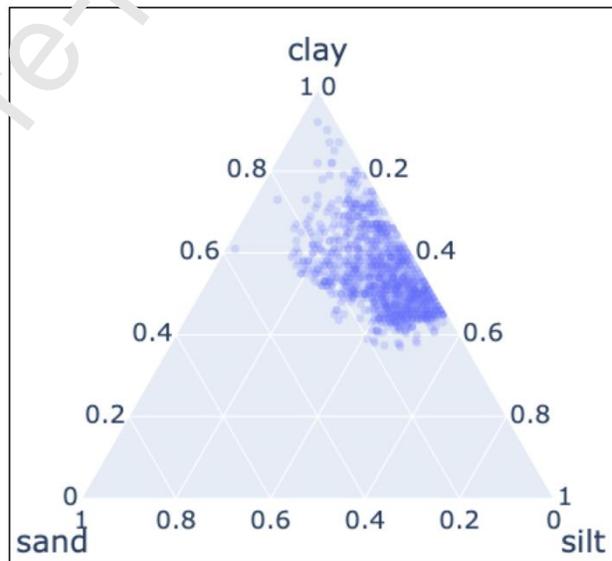


Figure 2: Texture triangles of the soil clusters (natural references per pedoclimate approach). Each point represents the topsoil value of a LUCAS site. Refer to the electronic version for colors.

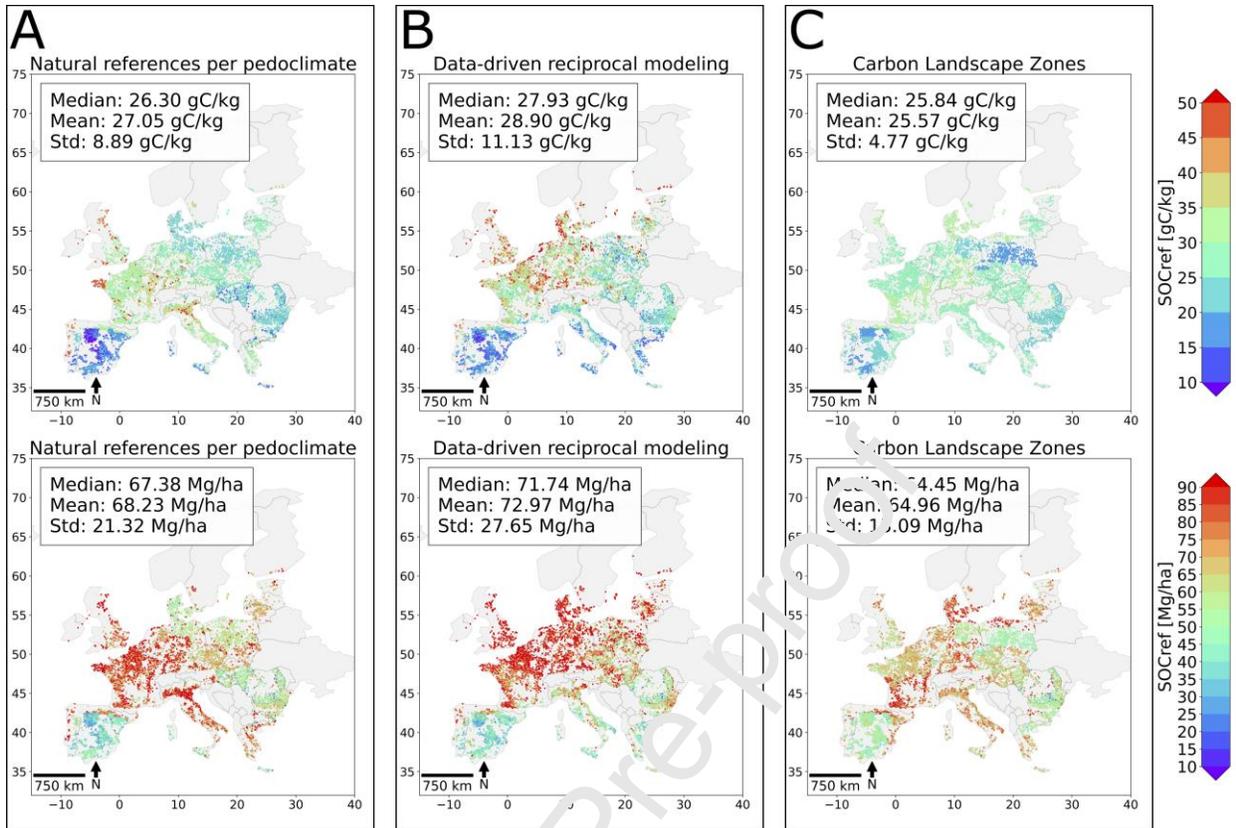


Figure 3: SOC reference values for LUCAS croplands extracted using the different methods. Data are shown for croplands of dataset n. 3bis to allow comparison across all methods. **A: "natural references per pedoclimate" approach.** Top: reference values expressed as SOC content. Bottom: reference values expressed as SOC stock. **B: "data-driven reciprocal modeling" approach.** Top: reference values expressed as SOC content. Bottom: reference values expressed as SOC stock. **C: "carbon landscape zones" approach.** Top: reference values expressed as SOC content. Bottom: reference values expressed as SOC stock. Refer to the electronic version for colors.

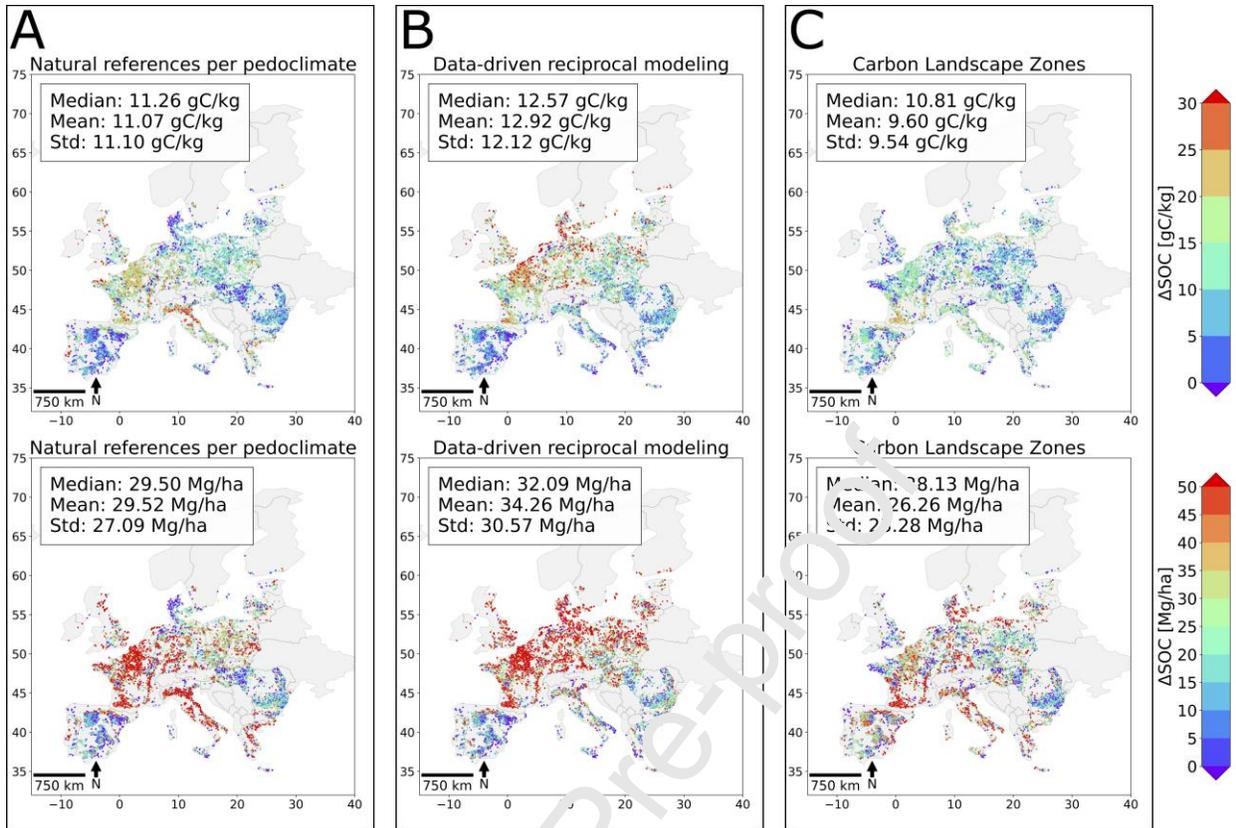


Figure 4: SOC storage potentials for LUCAS croplands estimated using the different methods. Data are shown for croplands of dataset n. 3bis to allow comparison across all methods. The SOC storage potential is defined as the difference between the reference SOC and the actual SOC. **A: "natural references per pedoclimate" approach.** Top: SOC storage potential expressed as SOC content. Bottom: SOC storage potential expressed as SOC stock. **B: "data-driven reciprocal modelling" approach.** Top: SOC storage potential expressed as SOC content. Bottom: SOC storage potential expressed as SOC stock. **C: "carbon landscape zones" approach.** Top: SOC storage potential expressed as SOC content. Bottom: SOC storage potential expressed as SOC stock. Refer to the electronic version for colors.

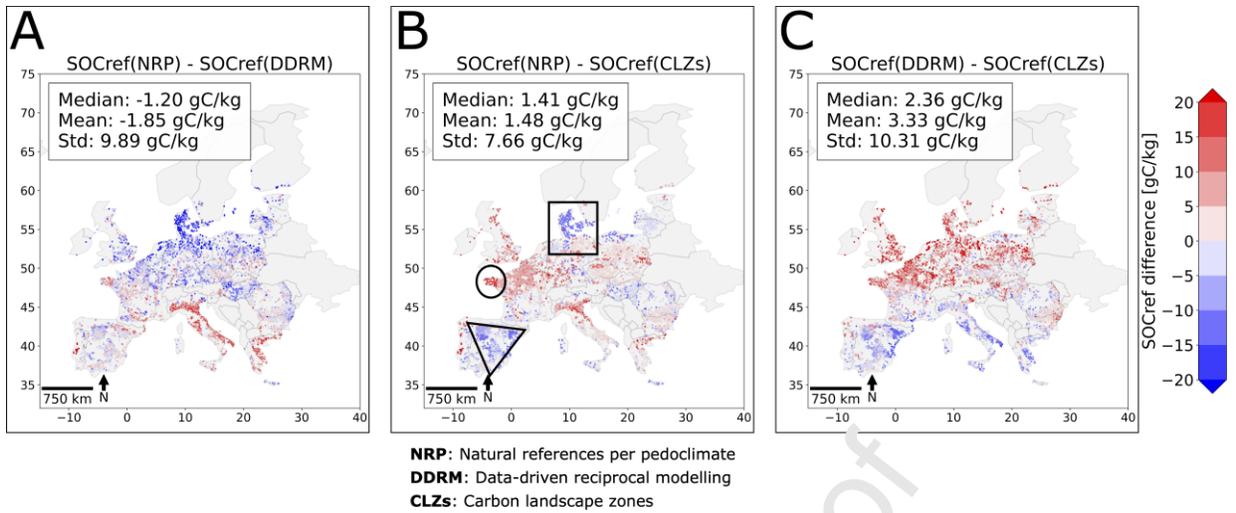


Figure 5: Differences in SOC reference values for UCAS cropland sites extracted using the different methods. Values are expressed as SOC contents. **A:** "natural references per pedoclimate" minus "data-driven reciprocal modelling". **B:** "natural references per pedoclimate" minus "carbon landscape zones". **C:** "data-driven reciprocal modelling" minus "carbon landscape zones". Refer to the electronic version for colors.

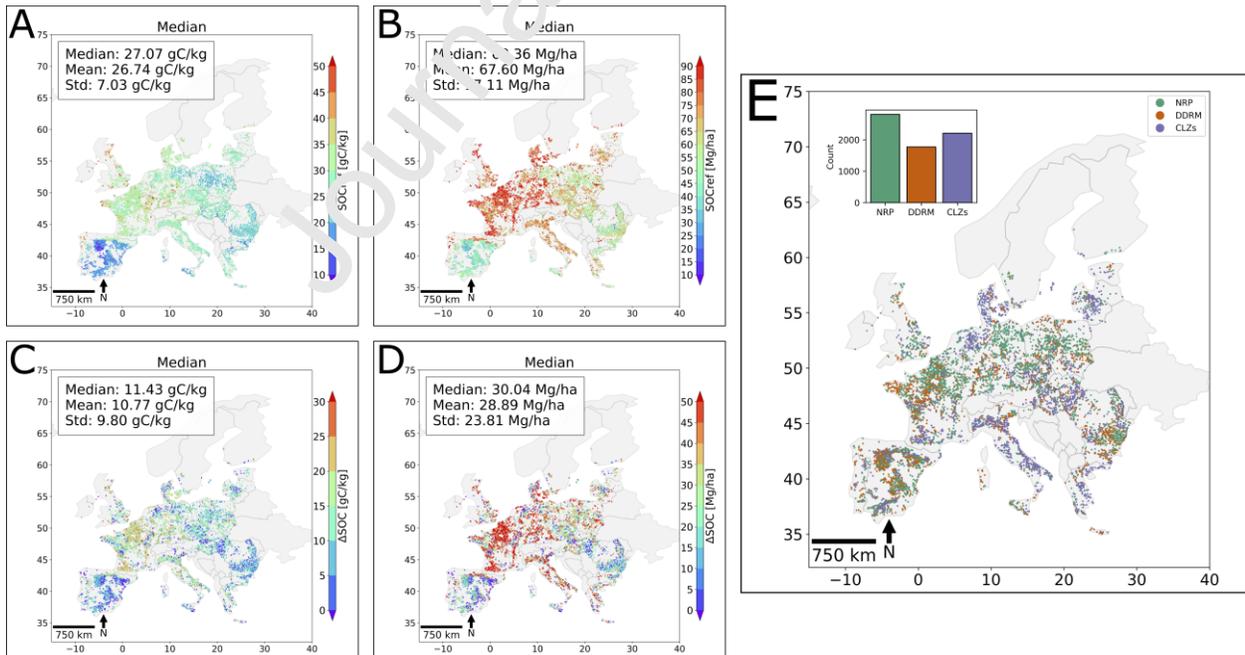


Figure 6: Ensemble modelling : median SOC reference values and median SOC

storage potentials for LUCAS croplands. Data are shown for croplands of dataset n. 3bis to allow comparison across all methods. The median is evaluated across the methods "natural references per pedoclimate", "data-driven reciprocal modelling" and "carbon landscape zones". **A: median SOC reference values expressed as SOC content. B: median SOC reference values expressed as SOC stock. C: median SOC storage potential expressed as SOC content. D: median SOC storage potential expressed as SOC stock. E: method kept when taking the median SOCref.** NRP: "natural references per pedoclimate", DDRM: data-driven reciprocal modelling, CLZ: "carbon landscape zones". Refer to the electronic version for colors.

CRedit author statement

Lorenza Pacini: Conceptualization, Methodology, Software, Investigation, Writing – Original draft preparation. **Pierre Arbelet:** Conceptualization, Supervision. **Songchao Chen:** Software, Methodology. **Aurélie Bacq-Labreuil:** Conceptualization, Funding acquisition. **Christophe Calvaruso:** Conceptualization. **Florian Schneider:** Software, Methodology. **Dominique Arrouays:** Supervision. **Nicolas P.A. Saby:** Supervision. **Lauric Cécillon:** Conceptualization, Supervision, Funding acquisition. **Pierre Barré:** Conceptualization, Supervision, Funding acquisition. All authors have contributed to Writing – review & editing and have approved the final version of the draft.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

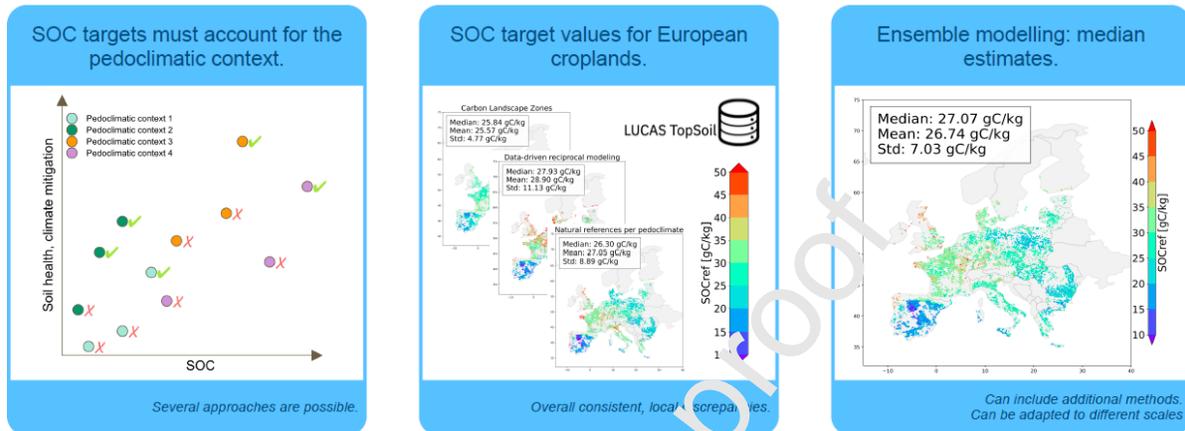
The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Journal Pre-proof

Graphical abstract

A new approach to estimate soil organic carbon content targets in European croplands topsoils.

Ensemble modelling provides robust soil organic carbon (SOC) targets to guide policymakers, farmers and stakeholders.



Reference: Lorenza Pacini, Pierre Aubelet, Songchao Chen, Aurélie Bacq-Labreuil, Christophe Calvaruso, Florian Schneider, Dominique Arrouay, Nicolas P.A. Saby, Lauric Cécillon and Pierre Barré, A new approach to estimate soil organic carbon content targets in European croplands topsoils. *Science of the Total Environment*, 2023.

Highlights

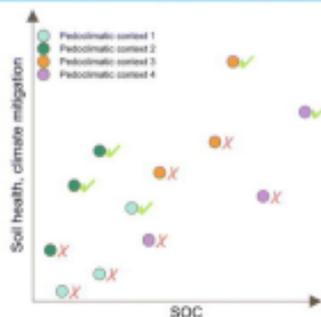
- Targets for soil organic carbon are estimated with different approaches.
- Targets are expressed as SOC content (g/kg) or stock (Mg/ha).
- Targets for 0-20cm topsoils of EU28 croplands are estimated using LUCAS data.
- The obtained estimations are overall consistent but with local discrepancies.
- More robust estimations are obtained using ensemble modeling (median estimation).

Journal Pre-proof

A new approach to estimate soil organic carbon content targets in European croplands topsoils.

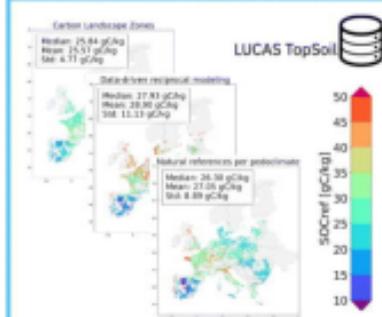
Ensemble modelling provides robust soil organic carbon (SOC) targets to guide policymakers, farmers and stakeholders.

SOC targets must account for the pedoclimatic context.



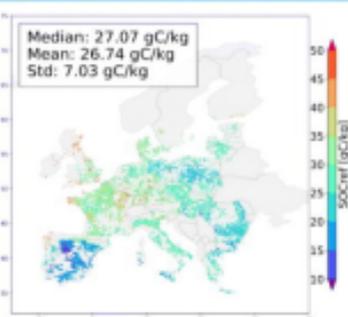
Several approaches are possible.

SOC target values for European croplands.



Overall consistent, local discrepancies.

Ensemble modelling: median estimates.



Can include additional methods.
Can be adapted to different scales.



Reference: Lorenza Pacini, Pierre Arbelet, Songchao Chen, Aurélie Bacq-Labreuil, Christophe Calvaruso, Florian Schneider, Dominique Arrouays, Nicolas P.A. Saby, Lauric Cécillon and Pierre Berné. A new approach to estimate soil organic carbon content targets in European croplands topsoils. *Science of the Total Environment*, 2023.

Graphics Abstract

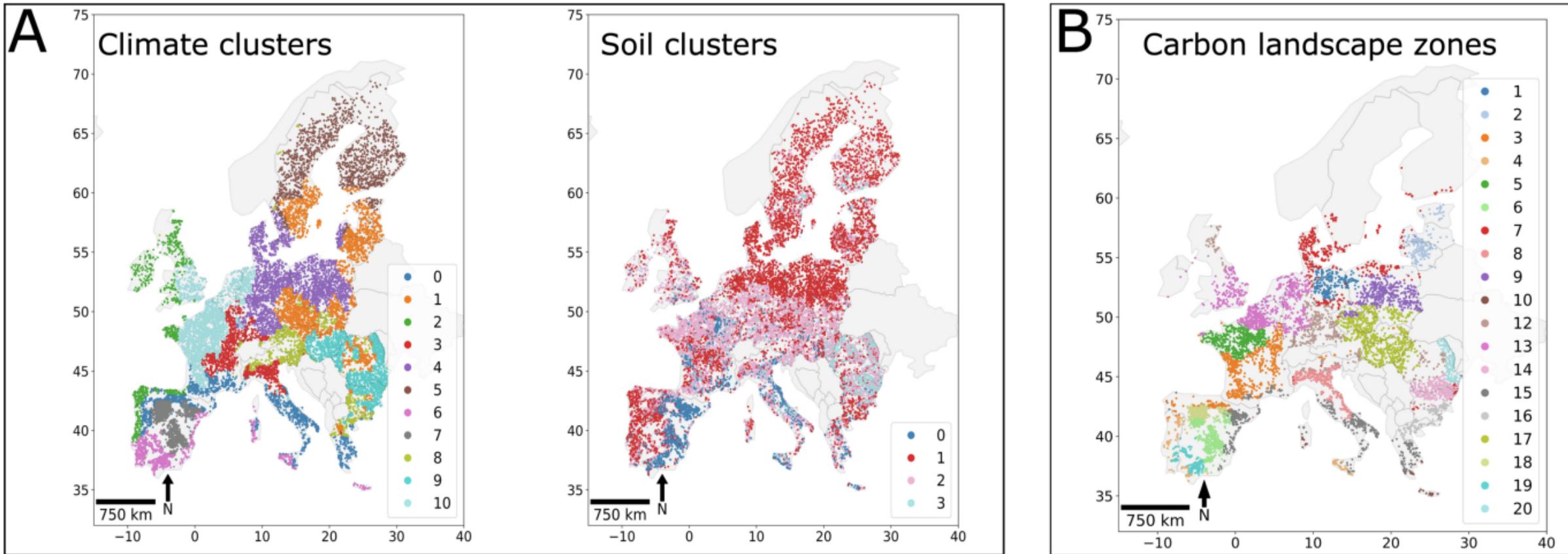
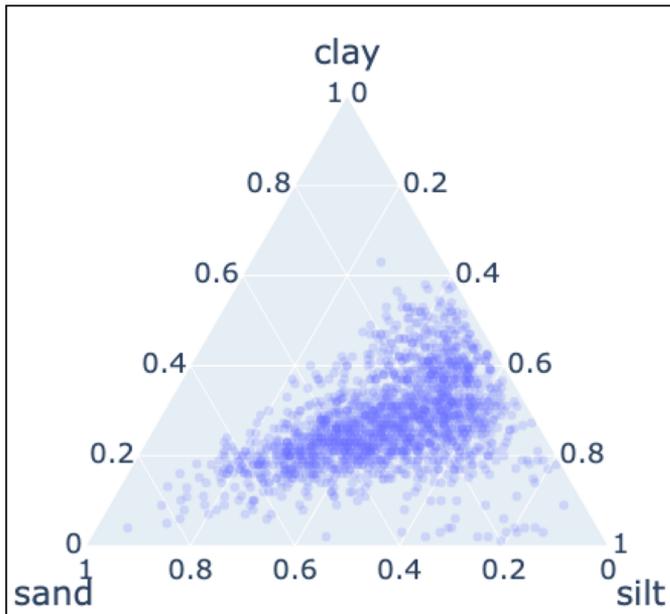
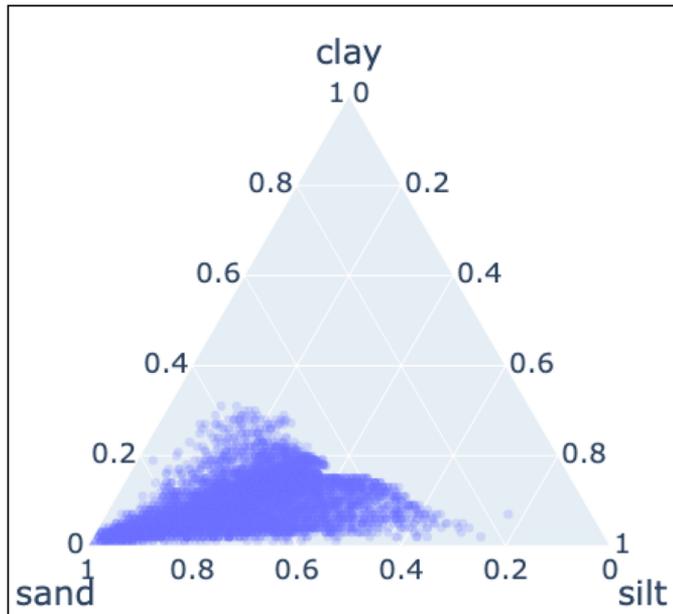


Figure 1

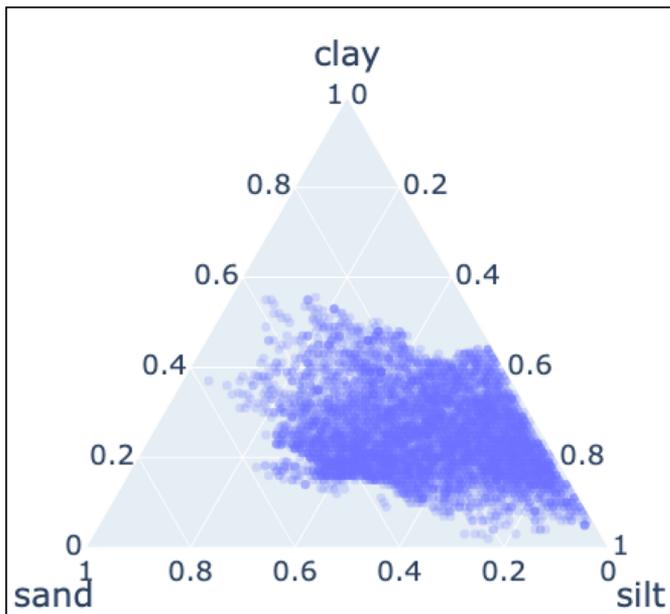
Cluster soil 0



Cluster soil 1



Cluster soil 2



Cluster soil 3

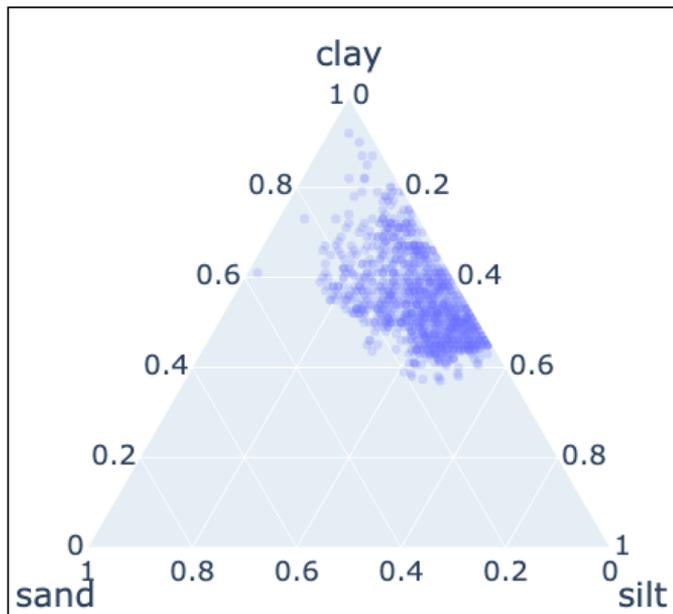


Figure 2

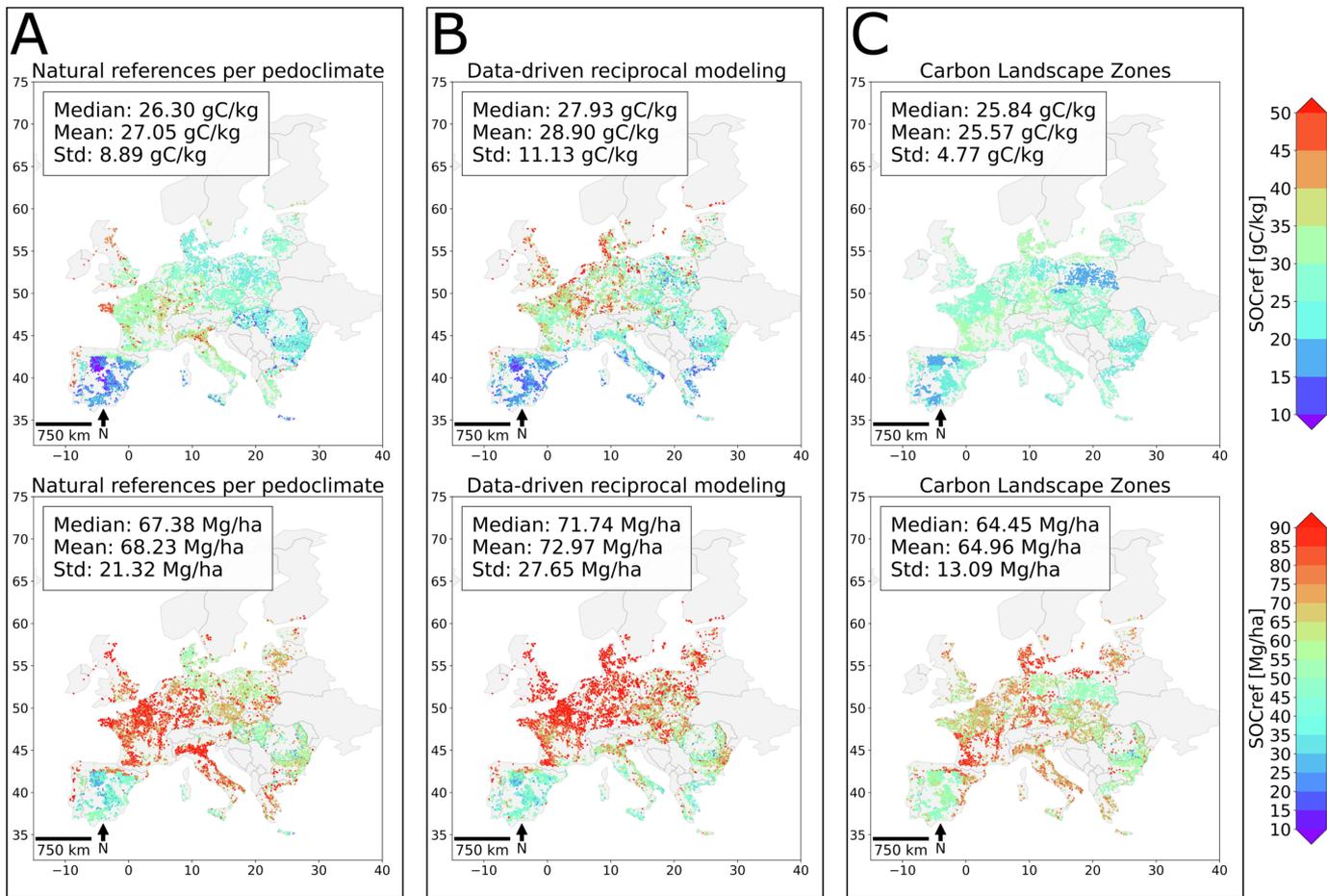


Figure 3

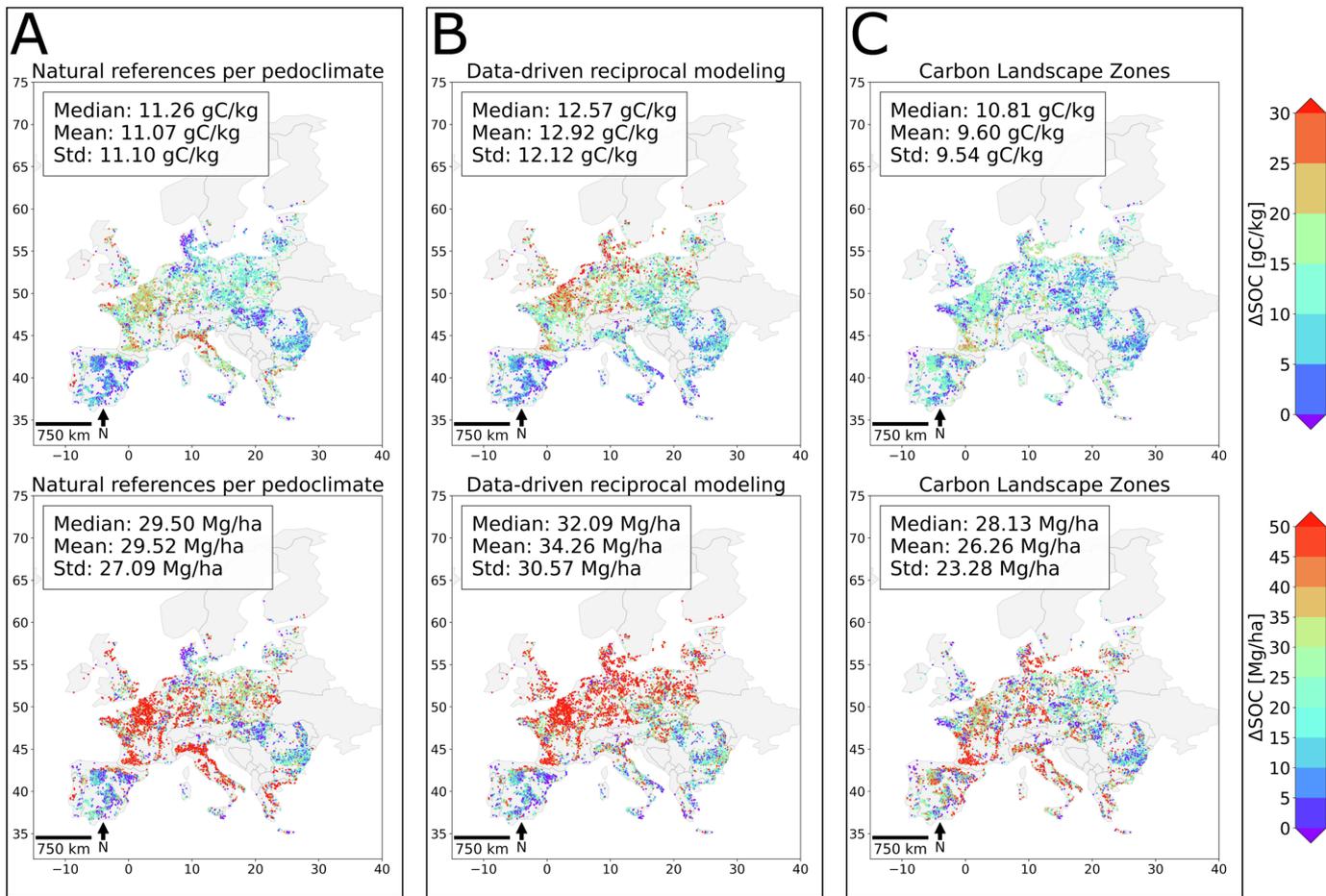
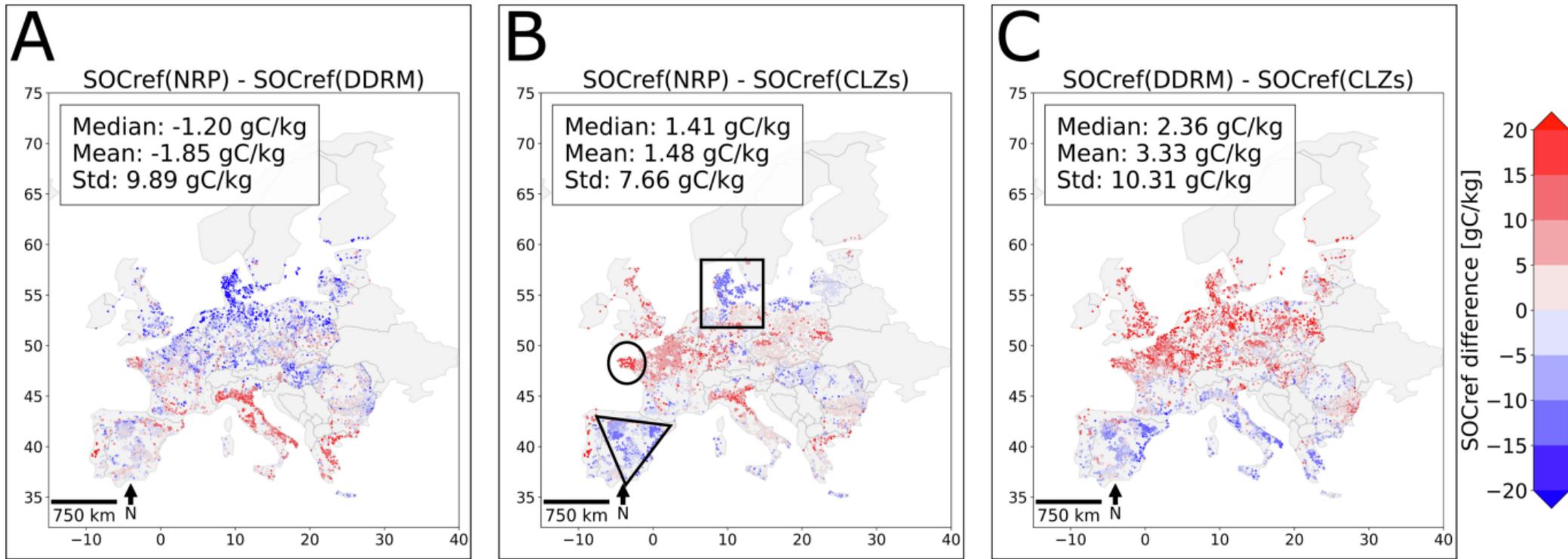


Figure 4



NRP: Natural references per pedoclimate
DDRM: Data-driven reciprocal modelling
CLZs: Carbon landscape zones

Figure 5

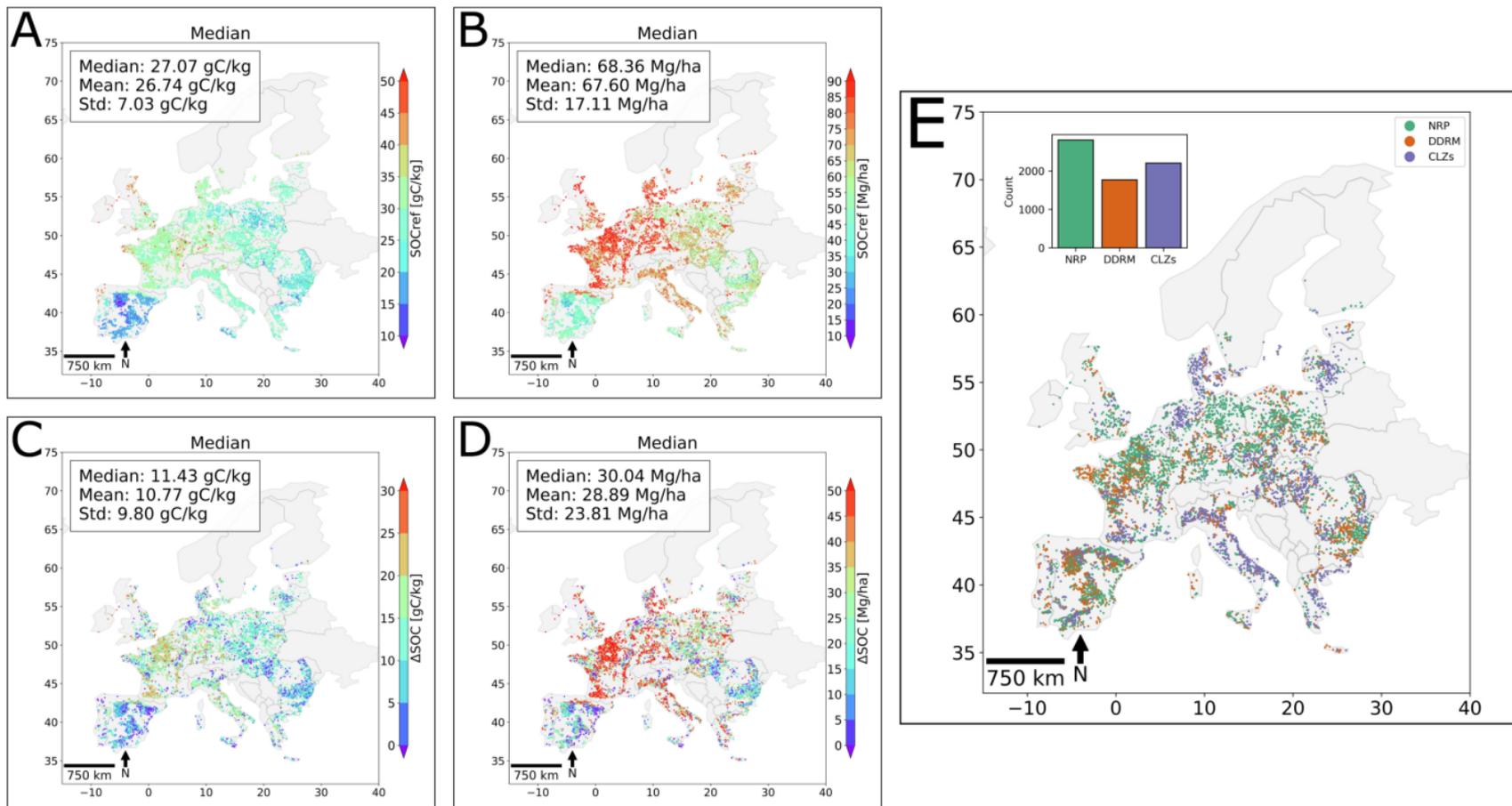


Figure 6