







A High-resolution Soil Organic Carbon Map for Great Britain

Salar Mahmood ^a, Beccy Wilebore ^a, Rhosanna Jenkins ^a, Katherine J. Willis ^{a,b}, Marc Macias-Fauria ^{a,c} and Shibu M. Ebrahim ^a

^aNatcap Research, London, UK; ^bDepartment of Biology, University of Oxford, Oxford, UK; ^cScott Polar Research Institute, Department of Geography, University of Cambridge, Cambridge, UK

ABSTRACT

Soil organic carbon (SOC) is used for soil health, indicating soils' agricultural productivity potential, and correlating with other functions like water capacity and soil biodiversity. SOC stocks are increasingly recognized in climate change mitigation strategies. Soil carbon sequestration represents 25% of all natural climate solutions to carbon. Current SOC maps for Great Britain (GB) are limited due to their coarse resolution (0.5–1 km). High demand exists for fine resolution SOC maps to estimate stock baselines and inform field-scale sampling strategies and land management. We present SOC concentration and stock maps at 5 m resolution for GB, generated using machine learning and accounting for physical and chemical soil properties, weather, topography and land cover (LC). Our model explains 74% of SOC variability in the evaluation dataset, with a RMSE of 9.8 (tC ha⁻¹). Soil pH and LC are the most important SOC predictors. We estimate ~2704Tg SOC for GB in the soil's top 30 cm: 1403Tg in England, 283Tg in Wales, and 1017Tg in Scotland. Neutral grasslands contribute the most SOC in England and Wales (37.2% and 50.7%). Dwarf heath shrubs, bogs and natural grasslands have higher contributions in Scotland (22–25%). Our SOC stock estimation compares with previous studies and our map reflects expected SOC distribution for different parts of GB under different LCs. Its high spatial resolution and accuracy enable SOC assessment at small scales (a single farm or a field) and can help to develop sustainable land management and guide soil sampling by optimizing sampling size and cost.

ARTICLE HISTORY

Received 27 May 2024
Accepted 07 October 2024

KEYWORDS



Agricultural soils; soil organic carbon (SOC); Great Britain; carbon map; Random Forest; land management; carbon stock

1. Introduction

Soil organic carbon (SOC) is the largest pool of terrestrial carbon, accounting for approximately three times more carbon than vegetation (Batjes, 1996; Scharlemann et al., 2014). Measures to increase soil carbon alone are estimated to represent 25% of all potential natural climate solutions (Bossio et al., 2020) and could reduce net annual GHG emissions within the Agriculture, Forestry, and Other Land Use (AFOLU) sector by 3–71% (IPCC). SOC is also an (2014) excellent metric for soil health in agricultural soils: high SOC brings additional benefits such as sustaining agricultural productivity, reducing soil degradation, and improving available water capacity which can reduce flood risks. Realizing the soil's potential to tackle climate change has led to the development of policies, nationally and globally, that provide economic incentives to farmers and land managers to encourage soil and land management practices that can enhance soil carbon storage. In Great Britain (GB), with the Department for Environment, Food & Rural Affairs (DEFRA) new policies on payments for public goods within Environmental Land management schemes (ELMS), such as Sustainable Farming Incentive

(SFI), and with the emergence of new private sector markets for carbon credits, many farmers and land managers are looking for additional income through carbon sequestration by regenerative agricultural practices. Since soil sampling will be an absolute minimum criterion to take advantage of such funding schemes, a sampling approach informed by a priori knowledge of soil organic carbon variability within a farm will be cost-effective, especially in the absence of previous sampling. A digital SOC map developed nationally at spatial resolutions relevant to land managers would provide a baseline representing the current status of SOC, which is fundamental in devising land management measures to improve SOC build-up in the future. Such a baseline SOC map would in addition help to identify the variables that control SOC locally, guide soil sampling, and be used for further modelling to estimate SOC offsetting opportunities.

Soils may gain or lose organic carbon depending on plant carbon input, soil type, land management, and climatic conditions. Spatial and temporal patterns of soil carbon stock and changes in carbon stock are therefore

CONTACT Marc Macias-Fauria  mm2809@cam.ac.uk  Scott Polar Research Institute, Department of Geography, University of Cambridge, Cambridge CB2 1ER, UK

Supplemental data for this article can be accessed online at <https://doi.org/10.1080/27658511.2024.2415166>

© 2024 Natural Capital Research Ltd. Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

a result of both local (e.g. land use, soil, parent material) and regional-to-global (e.g. climate, weather) factors that govern SOC balance in terms of biomass inputs (mainly from vegetation) and losses from decomposition (Muhammed et al., 2018; Schmidt et al., 2011). A number of approaches for spatially estimating SOC exist, including remote sensing (Arthur Endsley et al., 2020; Thaler et al., 2019), semi-mechanistic models (Falloon et al., 2006; Hendriks et al., 2021; Pierson et al., 2022), and digital soil mapping (DSM) ranging from simple statistical models to sophisticated machine learning (ML) approaches. DSM is a powerful tool that can be used to model the spatial distribution of SOC content as a function of the local variability as determined by a set of environmental covariates. DSM has been applied at field (Zhou et al., 2022), regional (L. Yang et al., 2021), national (Guevara et al., 2018; Heuvelink et al., 2021; Tao et al., 2020) and global (FAO, 2018; Hengl et al., 2014, 2017) scales to map SOC. Many different statistical approaches have been used for DSM to map SOC, including simple geostatistical models with a spatially varying mean (Kempen et al., 2019), ordinary kriging (Mishra et al., 2009), Scorpion kriging (Ungaro et al., 2010), ML (Emadi et al., 2020), or a combination of several modelling approaches (FAO, 2018). Usually, ML models are more accurate than simple statistical models, although insight behind the functioning of these models is complex (Wadoux et al., 2023). A number of ML approaches have been used to map SOC, encompassing Random Forest (Guevara et al., 2018), quantile regression forest (Heuvelink et al., 2021), Gradient Boosting (Zhou et al., 2022), deep learning or neural networking (Tao et al., 2020; L. Yang et al., 2021), or a combination of ML approaches (Hengl et al., 2017).

Great Britain has a wealth of broad scale SOC measurements through past national-level soil survey initiatives (Land Information System for the UK (LANDIS) (Proctor et al., 1998), National Soils Inventory of Scotland (NSIS) (Lilly et al., 2010), CEH Countryside Survey (Emmett et al., 2010)). These are in addition to the soil surveys done at the European scale (Land Use and Cover Area frame Statistical survey (LUCAS)). These various surveys provide consistent point-based datasets with spatial resolutions ranging from 1 km to 100 m (Aitkenhead & Coull, 2020) and provide SOC values, associated soil parameters, and land cover (LC) information for the whole or part of GB. Output from these data have been pivotal in the development of maps showing the distribution of SOC in soils for England & Wales, Scotland, and the UK. Comparison amongst these different maps also demonstrates widespread conformity. For example, when Feeney, Cosby (Feeney et al., 2022) compared eight SOC (%) maps of top soil organic carbon with spatial resolutions ranging from

250 m to 1 km for the whole of GB, similar distributions were apparent for mineral soils with relatively low SOC. In contrast, higher uncertainty was apparent in mineral soils with very low SOC and in soils rich in carbon, particularly peatlands in Scotland.

The problem with all of these existing models, however, is their unsuitability for determination of SOC at local field or farm scale. Their low granularity means that when assessing individual fields, they are of limited use because their finest resolution is $250\text{ m} \times 250\text{ m} = 6.25\text{ ha}$. The reason for this lack of fine resolution is due to the limited availability of fine-resolution covariate data such as land cover (LC), which is one of the most important predictors of SOC (Meersmans et al., 2012; Thomas et al., 2020). This lack of fine resolution maps for SOC is becoming a serious knowledge gap. Underestimation of organic soils can have serious ramifications in quantifying SOC inventories nationally or globally and the potential of soil carbon to meet internationally agreed carbon storage targets. Similarly, overestimation of SOC concentrations may over-credit farmers for their land management practices and misguide them in their soil offsetting approaches.

To address this knowledge gap, this paper describes the results of a machine learning modelling approach where we have utilized existing point field data along with other associated open-source soil parameters and environmental data to develop a 5 m resolution SOC map for GB. We then tested the mapped output of SOC using an evaluation field dataset. We demonstrate that our model explains 74% of the variability in SOC in the GB landscape with a RMSE of $9.8\text{ (tC ha}^{-1}\text{)}$ and represents an important forward step in detailing the fine resolution distribution of SOC in GB.

2. Data and methods

To build a SOC model for GB, we implemented a Random Forest machine learning approach (Breiman, 2001; RF package 4.7–1.1 in the R environment) capable of predicting 0–15 cm SOC (%) at 5 m spatial resolution by combining soil and environmental covariates from different open data sources (detailed below) along with point-based SOC (%) measurements to train the model. Upon internal evaluation, the model was projected onto the whole of GB, predicting SOC (%) for 0–15 cm at a 5 m spatial resolution. To enable model results to be compared against multiple outputs from different modelling studies (Feeney et al., 2022) (external evaluation), our modelled SOC (%) values, initially estimated at 0–15 cm, were estimated at 0–30 cm. Subsequently, 0–30 cm soil organic carbon stock (Cs, t/ha) was estimated (Figure 1).

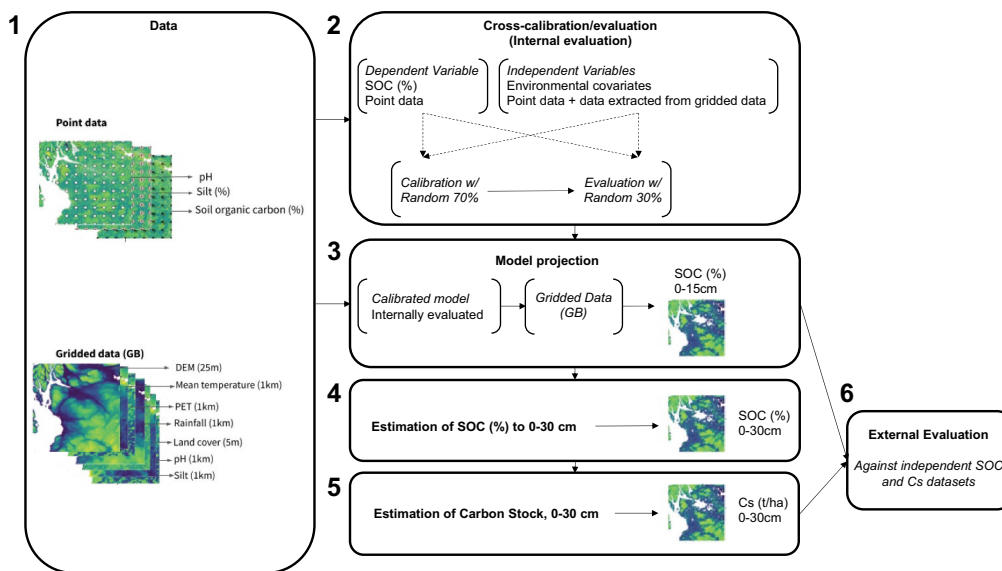


Figure 1. Flowchart showing the data used in this study and the modelling workflow. (1): data (point-based and gridded); (2) cross-calibration/evaluation of the RF model; (3) model projection onto a 5 m GB grid (SOC (%), 0-15 cm); (4): estimation of SOC (%) to 0-30 cm; (5) estimation of carbon stock (cs, t/ha) from SOC (%), 0-30 cm; (6) external evaluation against independent SOC and Cs datasets.

2.1. Data

2.1.1. Model training and evaluation

A suite of soil and environmental variables were initially chosen as potential predictors in the model, based on a priori knowledge of variables known to affect SOC (Ayala Izurieta et al., 2021; Vos et al., 2019; Wang et al., 2022; Zeraatpisheh et al., 2022). These predictors included Land Cover (LC), clay (%), sand (%), silt (%), pH, elevation

(m.a.s.l.), slope (%), aspect ($^{\circ}$), Topographic Wetness Index (TWI), annual mean soil temperature ($^{\circ}$ C), annual mean air temperature ($^{\circ}$ C), average monthly precipitation (rainfall, mm), and Potential Evapotranspiration (mm).

2.1.1.1. Dependent variables. SOC values were obtained from the LUCAS and the National Soils Inventory of Scotland (NSIS) datasets. LUCAS is the most consistent land use and cover database in Europe, managed

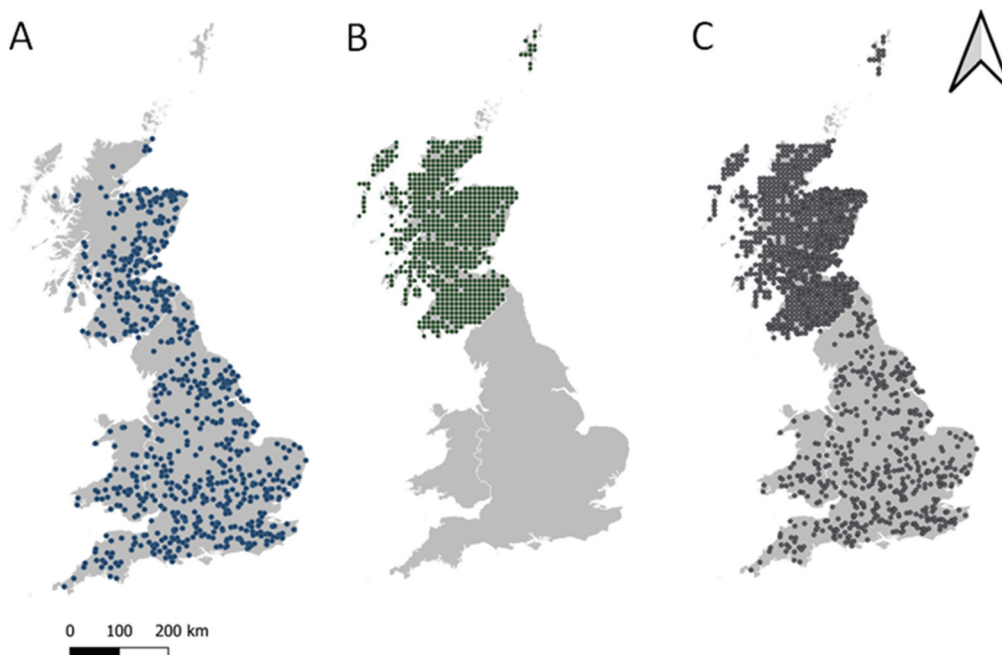


Figure 2. Sampled data points across Great Britain; (A) the Land Use and cover area frame statistical survey (LUCAS) topsoil dataset; (B) the national soils inventory of Scotland (NSIS) dataset, and (C) LUCAS and NSIS datasets together.

by the European Statistical Office of the European Commission (Eurostat). Since 2006, Eurostat has conducted LUCAS surveys every three years (Jones et al., 2020). We used data from a survey conducted in all 28 EU member states in 2015 resulting in 22,631 sampling points, of which 711 were collected in Great Britain (Figure 2A). In each LUCAS point, a composite 500 g soil sample was taken from five subsamples collected with a spade to a depth of 20 cm within a distance of 2 m, once stones, vegetation residues, grass and litter had been removed from the surface. The five subsamples were mixed, vegetation residues and stones removed, and 500 g of the mixed soil used in subsequent analysis. Soil physical and chemical properties were analysed according to standard ISO methods. Organic carbon content was analysed by dry combustion (elementary analysis; ISO 10,694:1995). The NSIS dataset was used to cover the area of Scotland (Figure 2B), which was not comprehensively covered by LUCAS. NSIS includes soil and site data taken over a 10 km point-grid at different depths from multiple horizons at each point (Lilly et al., 2010). NSIS samples were conducted at various times from 1971 to 1981. In each NSIS point, a pit was excavated at each site to reveal the soil profile and the collection of soil samples by horizon to at least 75 cm (100 cm in peat). Samples for all analyses were air-dried at 30°C and sieved (<2 mm). Samples for C content were further dried at 50°C. Elemental carbon (g/kg) was performed on a Hewlett-Packard CHN 185 analyser (Hewlett-Packard, Palo Alto, CA, United States). NSIS data were filtered to only include samples taken at depths within 0–15 cm to match the LUCAS dataset, which resulted in 695

points. The final dataset (LUCAS plus NSIS) consists of 1406 georeferenced sampling points throughout Great Britain (Figure 2C). SOC in the LUCAS dataset is given in g/kg, which was divided by 10 to convert it to SOC (% or g/100 g). SOC in NSIS data was given as Soil Organic Matter (SOM, %): we used the conventional factor of 1.72 to convert SOM to SOC (%) (Leigty & Shorey, 1930).

2.1.1.2. Covariates. In addition to SOC, both LUCAS and NSIS datasets included most of the covariates required for model training, such as soil physical and chemical properties, LC, and topographic indices. Climate variables (rainfall, annual mean air temperature, and potential evapotranspiration) were extracted from the 1 km grid cells overlapping with the point data from HadUK-Grid, a gridded climate dataset produced by the Met Office Hadley Centre for Climate Science and Services (Hollis et al., 2019).

2.1.2. Model projection on a map (gridded datasets)

In this study, we used a LC map at 5 m resolution generated from the best available open-source data layers in England, Wales, and Scotland (Table S1), which were selected based on the most recent data with the highest possible spatial resolution. All input datasets were harmonized using the UK Habitat classification (UK Habitat Classification Working Group, 2018). The UK Habitat classification is a hierarchical classification system with five nested levels (1–5) for categorizing primary habitats. The highest level of detail possible was used when harmonizing the different data

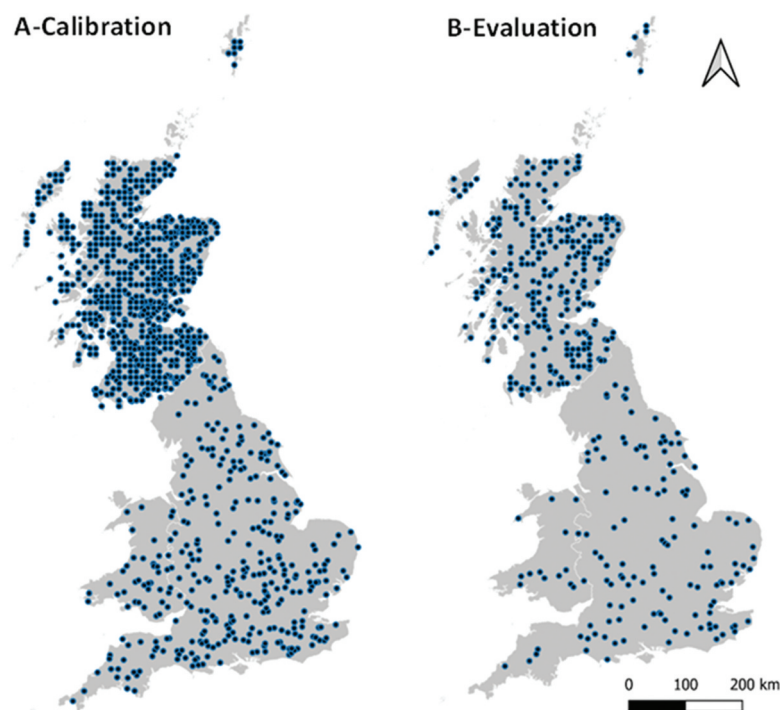


Figure 3. The distribution of (A) calibration and (B) evaluation datasets across Great Britain.

sources. Open-source data were complemented by a map of hedgerows and trees outside of woodlands created using canopy height data from BlueSky Ltd (Zellweger et al., 2022).

Soil pH and silt (%) data were downloaded from European Soil Data Centre (ESDAC) at 500 m spatial resolution. These maps were generated from LUCAS soil surveys in 2009 and 2012 (Ballabio et al., 2016, 2019). Since large parts of Scotland were not covered by LUCAS sampling points (Figure 1A), for some parameters like soil pH, gridded data (1 km) from The James Hutton Institute were used (Carag et al. 2022). Elevation and slope data were derived from the EU-Digital Elevation Model (DEM) at 25 m spatial resolution (EU-DEM - Version 1.1, 2018). Gridded climate data (rainfall, mean temperature, and potential evapotranspiration) were obtained from HadUK-Grid [1 km resolution; 39]. Average monthly precipitation, mean monthly temperature and monthly potential evapotranspiration were calculated from 1981 to 2020. Raster maps of all predictor variables were resampled to 5 m spatial resolution to match the resolution of our LC map.

2.2. Variable importance and variable selection

In the RF regression model, variable importance was estimated by (i) Mean Decrease in Accuracy and (ii) Increase in Node Purity. Mean Decrease in Accuracy (also known as Percent Increase in Mean Square Error, or %IncMSE) quantifies the importance of a variable by measuring the change in model prediction accuracy (% decrease) when the values of the variable are randomly permuted compared with the original observations. Increase in Node Purity (IncNodePurity) uses the Gini coefficient to measure the change in the homogeneity of the groups created by the trees in the Random Forest model when the values of a given variable are randomly permuted.

Variable selection in the RF regression model was done through iteratively running the model evaluation as follows: any covariate that did not increase the model's percent variance explained or reduce Root Mean Squared Error (RMSE) was removed from the analysis; covariates with the lowest %IncMSE were removed. This was done until a set of covariates that resulted in a model with the highest variance explained and lowest RMSE was obtained.

2.3. Estimation of SOC (%) to 0–30 cm

Most of the organic carbon in soil is concentrated on the surface layer, which means that soil measurements are usually done to a depth of 30 cm. SOC in these top layers is considered to be agronomically

more important, and farmers usually only measure soil to a depth of 30 cm. However, since soil covariates in both point and gridded datasets corresponded to 0–15 cm depth, the prediction in this study was initially made for 0–15 cm. We then used the modelled 0–15 cm SOC (%) to predict SOC to 0–30 cm employing an exponential decline function described by Hilinski (Hilinski, 2001) and Ottoy, Elsen (Ottoy et al., 2016) to obtain SOC (%) for the entire 0–30 cm. The function as described by Ottoy, Elsen (Ottoy et al., 2016) is:

$$C(z) = C_b + (C_0 - C_b)e^{-(k_{ED}z)}, \quad (1)$$

where $C(z)$, C_b , and C_0 are SOC (%) at depth z (30 cm in this case), the base of the profile, and the surface (5 cm), respectively; and k_{ED} is a constant that determines the shape of the exponential decline function. A higher value of k_{ED} indicates a strong decline in SOC with depth increase (Ottoy et al., 2016). Since C_0 is unknown, it has to be estimated from the predicted SOC (%) for the known depth of 0–15 cm. Hilinski (Hilinski, 2001) used a factor of 1.5 to calculate SOC at 0–5 cm from estimated SOC at a known depth of 0–20 cm. Accordingly, here we used a factor of 1.375 to calculate C_0 from predicted SOC at 0–15 cm.

Equation 1 was integrated for the entire 0–30 cm depth as follows:

$$\int_0^{30} C = \frac{e^{-30k}((30C_b k - C_b + C_0)e^{-30k} + C_b - C_0)}{k}, \quad (2)$$

2.4. Soil organic carbon stock (Cs)

Estimates of soil C stock for 0–30 cm were generated from SOC (%) and bulk density at 0–30 cm, which enables the comparison of our predicted SOC with other similar maps which are measured and modelled to a depth of 30 cm. Soil organic carbon stock (C_s , t/ha) down to 30 cm depth was calculated as follows:

$$C_s = SOC_{30} * BD * D, \quad (3)$$

where SOC_{30} and Bulk Density (BD) are soil organic carbon and bulk density down to 30 cm depth, respectively, and D is soil depth in cm. This resulted in C_s (mass/area), which were presented as tonnes/hectare (t/ha) to facilitate comparison with other maps and datasets.

Bulk Density for Great Britain usually changes between LC classes, ranging from 0.17 g cm⁻³ for 'Bog & Blanket Bog' to 1.23 g cm⁻³ for 'Arable & Horticulture' LC classes (Emmett et al., 2010). However, most BD maps available at national scales

show relatively high *BD* values (>0.5) for LC classes with higher SOC (%), such as bogs and heathlands, and result in extremely high carbon stock and estimation errors (Aitkenhead & Coull, 2020). For example, the European Soil Data Centre (ESDAC) estimates *BD* from soil packing density and clay content (Ballabio et al., 2016) show high uncertainty, especially in the peatland area of northwest Scotland that is not covered by the LUCAS soil sampling. Therefore, in this study, we used the method described by The Countryside Survey: Soils Report 2007 (Emmett et al., 2010) to calculate *BD* from SOC%, which assumes an exponential decline in *BD* with an increase in SOC as follows:

$$BD = 1.29e^{-0.0206x} + 2.51e^{-0.0003x} - 2.057, \quad (4)$$

where x is SOC g/kg at 0–30 cm depth. We used this approach since (i) pedo-transfer functions with a negative exponential relationship between SOC (%) and *BD* have been shown to perform better than linear ones (Yi et al., 2016); (ii) this empirical function (equation 4) is based on data from British soils; and (iii) pedo-transfer functions using SOC (%) as a predictor have long been shown to have high predictability (Manrique & Jones, 1991).

2.5. Model performance (evaluation)

2.5.1. Internal model evaluation

To evaluate the predictive power of our RF regression model, we randomly split our initial point-dataset in two, where 70% (985 data-points) was used for training the model (calibration) and the remaining 30% (421 data-points) for model evaluation (cross-calibration /evaluation; Figures 1 & 3), whereby the model calibrated on 70% of the data points was projected on the evaluation data points. Percent explained variance (R^2) and RMSE were calculated for the evaluation dataset for all habitats, collectively and separately.

2.5.2. External model evaluation

2.5.2.1. Maps. We compared our predicted maps against existing available national maps for Great Britain. These were:

- SOC (%) 0–15 cm: the Countryside Survey SOC (Emmett et al., 2010), 1 km spatial resolution) and the LUCAS (500 m spatial resolution) maps;
- Cs (t/ha) 0–30 cm: the Global map of C stock (GSOC) (FAO, 2018), 1 km spatial resolution.

2.5.2.2. Summary statistics. We also compared country- and LC-based Cs 0–30 cm summary statistics from our model with those available from the soil carbon and

land use database for the United Kingdom published by Bradley, Milne (Bradley et al., 2005), at 1 km spatial resolution.

2.5.3. LC classes in model evaluation

The LC map at 5 m resolution in our study includes very detailed habitat information, based on the UK Habitat classification (UK Habitat Classification Working Group, 2018). Due to this level of detail, some unique habitat categories were missing in the calibration/evaluation dataset or occurred at numbers so low that could not be used in model calibration/evaluation. In these latter cases, some LC classes were grouped and reclassified to broader LC groups in the calibration/evaluation dataset according to their overall higher-level similarity and similar SOC observed levels. Likewise, we reclassified some LC classes in the external evaluation to match them with those of the maps and datasets from other studies against which we compared our study. Finally, ‘Water Bodies’, ‘Urban Areas’, and ‘Inland Rocks & Sediment’ were removed, as they represented insufficient data for calibration or were not relevant to the study. **Fig. S1** provides a detailed account of how and why LC classes have been managed in this study.

3. Results

3.1. Internal model evaluation

The summary statistics and evaluation results suggest that the predicted SOC (%) 0–15 cm is within the ranges of observed SOC (%) 0–15 cm for different habitats, with highest average SOC (%) 0–15 cm values (44.7%) for ‘Bog & Blanket Bog’ LC classes and lowest average SOC (%) 0–15 cm values (3.8%) for ‘Arable & Horticulture’ LC classes (Table 1). LC-based average predicted SOC (%) 0–15 cm values are the same or very close to the average observed SOC (%) 0–15 cm for the same LC. However, within the same LC class, the standard deviation was higher for observed SOC (%) 0–15 cm compared to predicted SOC (%) 0–15 cm, since our model slightly overestimated the lowest observed SOC (%) 0–15 cm values and slightly underestimated the largest observed SOC (%) 0–15 cm values (Table 1). Figure 4 shows the linear relationships between observed and predicted SOC (%) 0–15 cm in the evaluation dataset (not used for training the model) and for each LC class separately. The model was able to predict the upper and lower limits of SOC (%) 0–15 cm in the evaluation dataset, while the points are more scattered around the middle of the regression line when considering all LC classes together (Figure 4A). Model performance was mostly consistent over all LC

classes, with R^2 ranging from 0.77 for ‘Woodland’ to 0.85 for ‘Bog & Blanket Bog’, except for ‘Dense Scrub, Fen Marsh & Swamp’, which showed R^2 values at 0.55 (Table 1; Figure 4).

We found similar spatial patterns between observed and predicted SOC (%) 0–15 cm (Figure 5A, B), with higher SOC (%) 0–15 cm values predicted in areas with higher observed SOC (%) 0–15 cm and vice versa.

Table 1. Summary statistics of observed and predicted SOC (%) 0–15 cm in evaluation dataset, for different LC classes, determination factor (R^2), and root mean square error (RMSE) of the linear relationship between them. For further information on the LC classes, please see ‘data and methods’ and fig. S1

Habitats		Mean	Min	Max	SD	R^2	RMSE
All habitats in the evaluation dataset	Obs	16.8	0.43	59.92	19.2	0.74	9.8
	Pre	16.4	1.36	53	15.7		
Woodland	Obs	18.7	1.09	58.9	18	0.77	8.76
	Pre	19.6	3.58	52.95	14.9		
Grassland	Obs	10.5	0.84	57.9	13.1	0.78	6.24
	Pre	10.5	2.7	5.4	10.4		
Arable & Horticulture	Obs	3.8	0.3	57.7	6.0	0.81	2.8
	Pre	3.8	1.25	45.4	4.6		
Bog & Blanket Bog	Obs	44.7	1.9	59.9	15.1	0.85	6.61
	Pre	43.8	9.7	55.3	11.2		
Dwarf Shrub Heath	Obs	28.7	0.62	57.9	19.1	0.82	8.32
	Pre	28.4	4.54	54.0	14.5		
Dense Scrub, Fen Marsh & Swamp	Obs	14.5	1.7	47.9	13.9	0.55	10.23
	Pre	11.9	4.9	35.8	6.2		

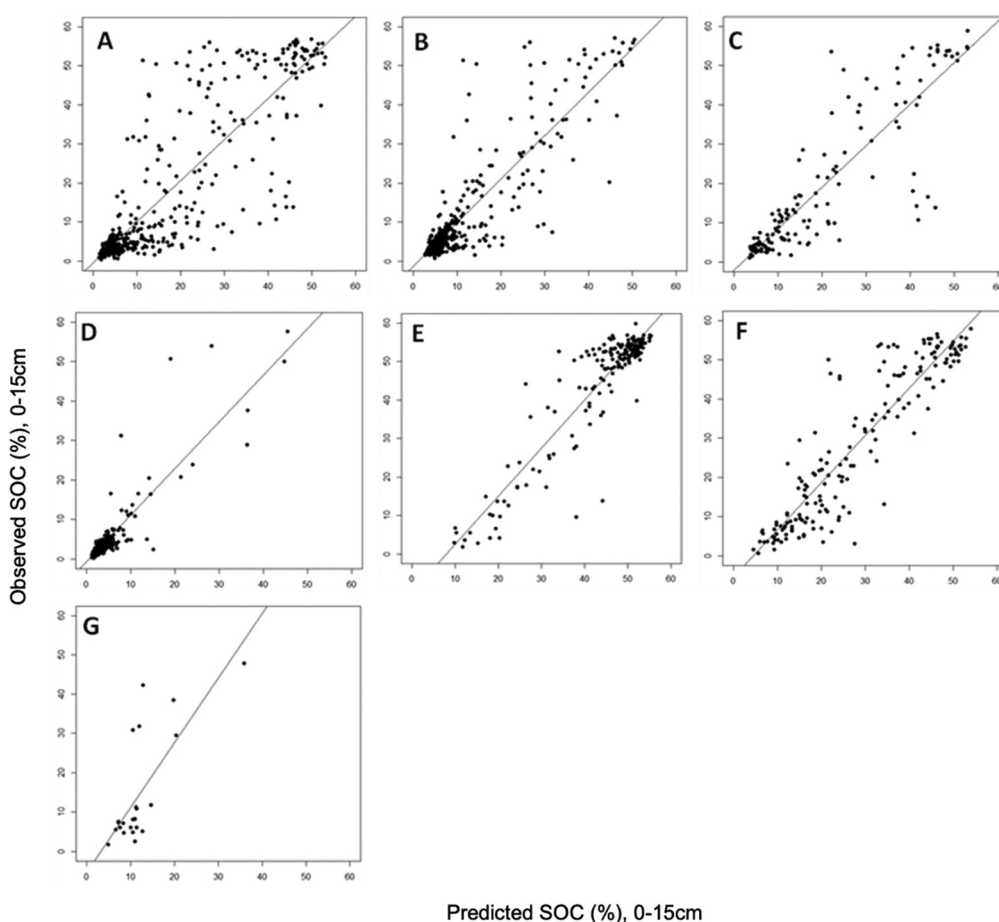


Figure 4. Linear regression between observed (evaluation dataset) and predicted SOC (%) 0–15 cm (evaluation values predicted with the model trained using the calibration dataset). (A) all habitats; (B) ‘grassland’; (C) ‘woodland’; (D) ‘arable & horticulture’; (E) ‘bog & blanket bog’; (F) ‘dwarf shrub heath’; and (G) ‘fen marsh, swamp & dense scrub’. For further information on the LC classes, please see ‘data and methods’ and fig. S1.

Overall, differences between observed and predicted SOC (%) 0–15 cm in the evaluation dataset were small, ranging from –5 to 5% for most points across Great Britain, especially in England and Wales (Figure 5C), while most underestimation and overestimation occurred in Scotland, in sites with high SOC (%) 0–15 cm values.

Finally, the distributions of SOC (%) 0–15 cm values across different LC classes for the data points used in the modelling exercise show a good fit between each of the LC classes' range of observed SOC (%) 0–15 cm values (Figure 6A) and those modelled in this study (Figure 6B). Whereas our model accurately matched the distributions and ranges of SOC (%) 0–15 cm values when compared to observations, Figure 6 also shows that the highest values tended to be underestimated, especially those in the 'Bog & Blanket Bog' LC class.

3.2. Variable importance

Eight predictors were finally selected to calibrate the model. Variable importance was estimated by Mean Decrease in Accuracy (%IncMSE) and Increase in Node Purity (IncNodePurity; see section 2.4). According to both statistics, soil pH was the most important variable, followed by LC, average monthly precipitation, silt (%), annual mean air temperature, slope (%), elevation (8th most important variable according to %IncMSE and 7th according to IncNodePurity), and Potential Evapotranspiration (7th most important variable according to %IncMSE and 8th according to IncNodePurity; Figure 7).

3.3. Cs ($t\ ha^{-1}$) 0–30 cm

SOC (%) 0–15 cm was converted to SOC (%) 0–30 cm and then to Carbon stocks (Cs) ($t\ ha^{-1}$) 0–30 cm as explained in sections 2.2 and 2.3. Results suggest that the total Cs at 0–30 cm is 2704Tg for GB, of which more than half is in England (52%), followed by Scotland (38%) and Wales (10%) (Table 2). Although total area under all the LC classes in Scotland (5.74Mha) is only less than half the total area in England (11.72Mha), total Cs in Scotland represents > 70% of the total Cs in England because of a higher average Cs ($169.1\ t\ ha^{-1}$) in Scotland compared to England ($139.6\ t\ ha^{-1}$) and Wales ($153.3\ t\ ha^{-1}$). The average Cs varies with different LC classes, with the highest values found in 'Bog & Blanket Bog' LC class, which range from $181.3\ t\ ha^{-1}$ in England to $212.3\ t\ ha^{-1}$ in Scotland, and the lowest values found in 'Arable & Horticulture' LC class, ranging from $88.6\ t\ ha^{-1}$ in England to $110.7\ t\ ha^{-1}$ in Wales. 'Neutral Grasslands' LC class has the highest relative contribution of total Cs in England (37.2%) and Wales (50.7%), representing 34% and 52% of total area in England and Wales respectively. In England, 'Arable & Horticulture' is the most frequent LC class (29% of the total area) and has a relative contribution to total Cs of 21.3%, while in Scotland, 'Dwarf Shrub Heath' LC class had the highest relative contribution (25%) followed by 'Bog & Blanket Bog' (23.5%) and 'Neutral Grasslands' (22.4%) LC classes, together representing 67% of total area.

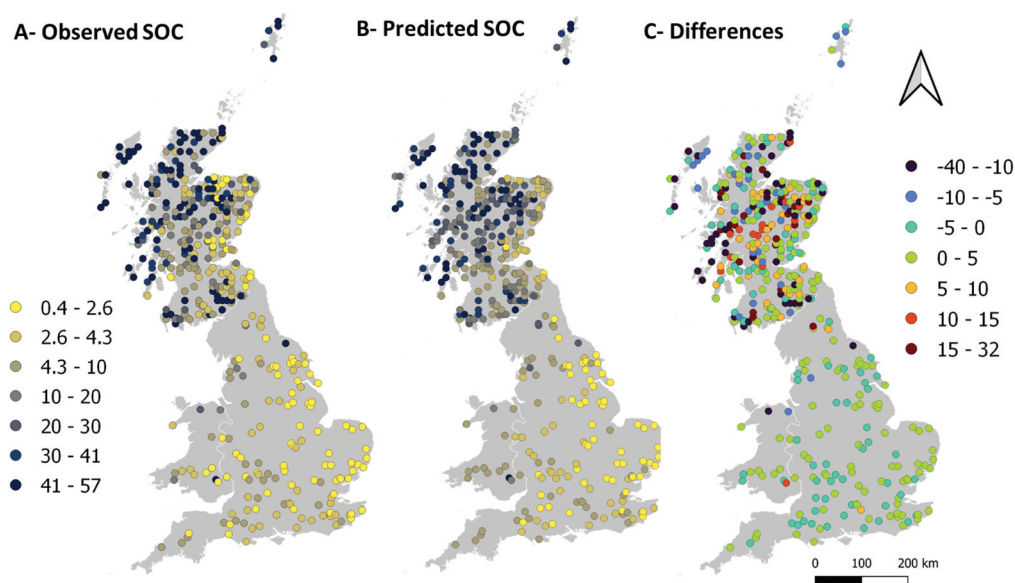


Figure 5. The distribution of SOC (%) evaluation points at 0–15 cm across Great Britain showing (A) observed, (B) predicted and (C) the difference between observed and predicted.

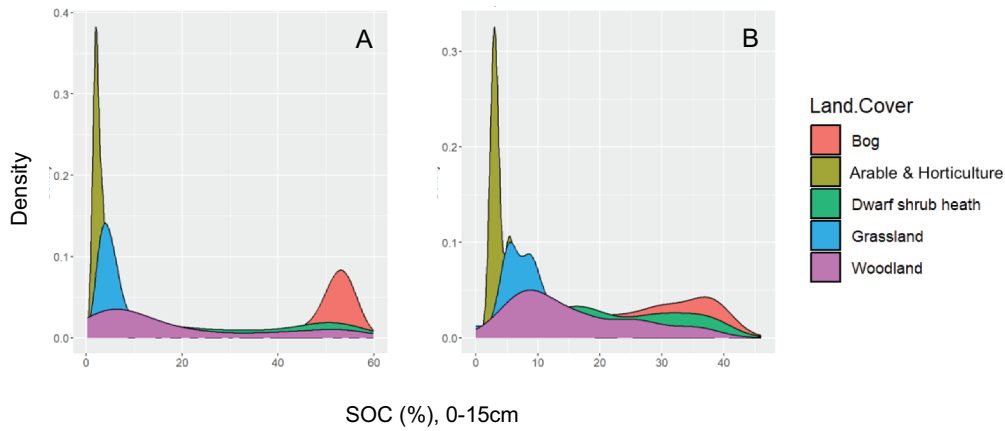


Figure 6. Probability density functions of SOC (%) 0–15 cm values per LC class for the point-based observations used in this study. (A) Observed points; (B) values for the same points predicted by this study. Plots generated with ggplot2 in R. The range of values in x-axes stops at the maximum SOC (%) 0–15 cm in each panel. For further information on the LC classes, please see ‘data and methods’ and fig. S1.

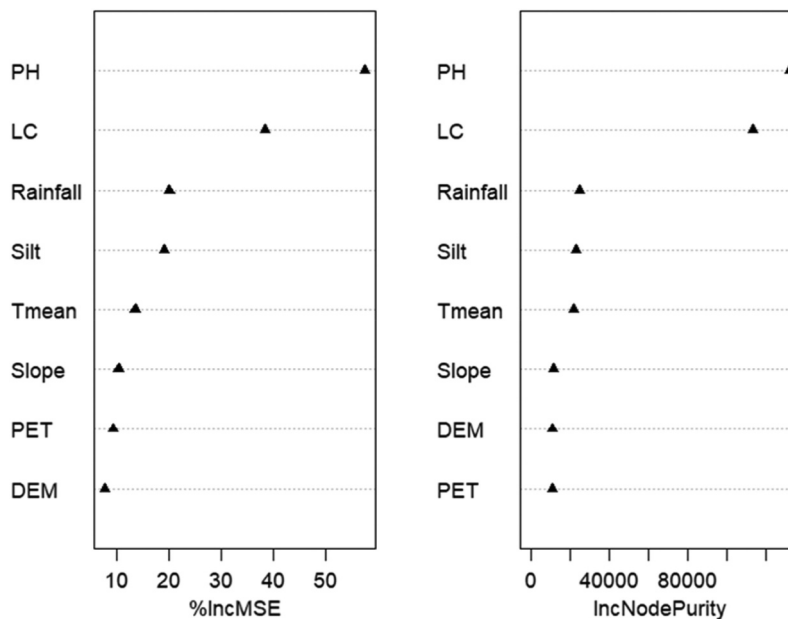


Figure 7. Predictor variable importance of the model (see section 2.2 for details). %IncMSE: mean decrease accuracy (percent increase in mean square Error); IncNodePurity: increase in node purity. PH: soil pH; LC: land cover; rainfall: average monthly precipitation (mm); Silt: silt (%); Tmean: annual mean air temperature ($^{\circ}$ C); slope: slope (%); PET: potential evapotranspiration (mm); DEM: elevation (m.a.s.l.).

3.4. External model evaluation

3.4.1. SOC (%) 0–15 cm: Countryside survey and LUCAS maps

There is an overall agreement in the spatial pattern of our predicted SOC (%) 0–15 cm map when compared to those of Countryside Survey and LUCAS for the same soil depths (Figure 8). For all these maps, high SOC (%) 0–15 cm (>45%) values are found mostly in central to western Scotland, whereas low SOC (%) 0–15 cm (<5%) values are mostly found in south and south-east England. This

high spatial association suggests that our model generates robust predictions. Further, our model was able to predict the upper and lower margins of SOC (%) 0–15 cm at national scales. Some disagreements can be noted in some areas between our predicted SOC (%) 0–15 cm map and the other two maps, especially in western parts of Scotland and northern England (–15 to –25%), with small differences in SOC (%) 0–15 cm ranging from –5 to +5% for most of GB. Importantly, the magnitude of the disagreement between the map in this study and those

Table 2. Summary statistics of predicted carbon stock (cs, t ha⁻¹) 0–30 cm for common LC classes in Great Britain. For further information on the LC classes, please see 'data and methods' and fig. S1

Land Cover Class	Min (t/ha)	Max (t/ha)	Mean (t/ha)	Total (Tg)	% Total	Area (1000 ha)
England						
Dense Scrub	98.9	209.9	137.1	60.8	4.33	443.5
Bog	139	216.1	182.3	45.7	3.25	250.7
Coniferous Woodland	115.1	216.1	147.6	62	4.42	420.3
Sparsely Vegetated Land	40.9	203.4	96.8	44.3	3.15	457.5
Acid Grassland	131.5	215	177.6	27.4	1.95	154.1
Arable & Horticulture	37.7	206.4	88.6	299.6	21.35	3383.4
Broadleaved & Mixed Yew Woodland	71	214.5	126.8	249.8	17.8	1970.3
Fen Marsh & Swamp	109.3	204.7	143.7	36.2	2.58	252
Neutral Grassland	78.7	211.5	129.2	522.1	37.2	4041
Dwarf Shrub Health	101.8	216	160.3	55.3	3.94	345.1
All Land Cover Classes	37.7	216.1	139.6	1403		11718
Wales						
Dense Scrub	107.1	183	151.5	0.8	0.29	5.5
Bog	139.7	216	181.3	7.7	2.73	42.7
Coniferous Woodland	121.2	215.8	160.9	28.4	10.04	176.8
Sparsely Vegetated Land	75.1	192.4	142.7	1.3	0.47	9.4
Acid Grassland	131.4	212.3	161.7	37.7	13.29	232.9
Arable & Horticulture	63.1	179.1	110.7	6.6	2.34	59.9
Broadleaved & Mixed Yew Woodland	86	199.1	145.2	37.2	13.13	256.4
Fen Marsh & Swamp	115.5	207.2	165.9	7.7	2.7	46.2
Neutral Grassland	89.9	196.7	144.6	143.6	50.66	993.2
Dwarf Shrub Health	109.3	215	163.7	12.1	4.28	74
All Land Cover Classes	63.1	216	153.3	283		1897
Scotland						
Dense Scrub	114.1	214	165.5	4.6	0.45	27.9
Bog	136.8	216.1	212.3	238.7	23.48	1124.3
Coniferous Woodland	117	216.1	183.4	134.6	13.24	733.8
Sparsely Vegetated Land	78.3	214.4	148	13.1	1.29	88.5
Acid Grassland	142.9	216.1	190.1	17.8	1.75	93.7
Arable & Horticulture	65.5	214.7	102.8	49.4	4.86	481.1
Broadleaved & Mixed Yew Woodland	91.8	215.8	157.9	57.8	5.68	365.8
Fen Marsh & Swamp	117.8	215	186.6	16.5	1.62	88.3
Neutral Grassland	96.3	216	158.2	228	22.43	1441
Dwarf Shrub Health	119	216.1	198.3	254.7	25.04	1283.9
All Land Cover Classes	6.5	216.1	169.1	1017		5728
Great Britain	37.7	216.1	154	2704		19343

*Tg (Teragram) = 1,000,000 tonnes

from the Countryside Survey and the LUCAS maps (Figure 8D, E) is well within the range of disagreement between the latter two maps (Figure 8F).

3.4.2. Cs (t ha⁻¹) 0–30 cm

3.4.2.1. Global map of C stock (GSOC). Predicted Cs 0–30 cm values were compared for GB from GSOC (FAO, 2018) (Figure 9). According to our study, Cs ranged from 37.7 t ha⁻¹ for the 'Arable & Horticulture' LC class in England to 216.1 t ha⁻¹ for the 'Bog & Blanket Bog' LC class across GB (Table 2; Figure 9A). Our predicted Cs map (Figure 9A) shows some association with the GSOC map (Figure 9B), but there is considerable disagreement between the two (Figure 9C). In general, our predicted Cs tends to be slightly higher in most areas across GB compared to the GSOC map. Our Cs estimates appear to be lower than those of the GSOC map in some areas in the east and south of Scotland, and across small pockets in the east of England.

3.4.2.2. National and LC-based Cs statistics. We estimate that soils of GB hold 2704Tg of organic carbon in the 0–30 cm depth soil, with an average of 154 t ha⁻¹ SOC. Our estimated value of total SOC stock compares well with the value reported by Howard, Loveland (Howard et al., 1995) for GB (2773Tg) and is slightly higher than the total Cs reported by Bradley, Milne (Bradley et al., 2005) (2371Tg; Table 3). At national scales, our predicted values of Cs for England, Wales and Scotland were 1403Tg, 283Tg and 1017Tg, respectively. These values are higher than those reported by Bradley, Milne (Bradley et al., 2005) for England and Wales (1015Tg and 194Tg, respectively), and slightly lower for Scotland (1161Tg). Our mean predicted values of Cs for the most common LC classes ('Coniferous Woodland', 'Arable & Horticulture', and 'Neutral grasslands') are higher for England and Wales compared to mean values reported by Bradley, Milne (Bradley et al., 2005), but the differences are much lower for Scotland. However, the total Cs (Tg) held by each of

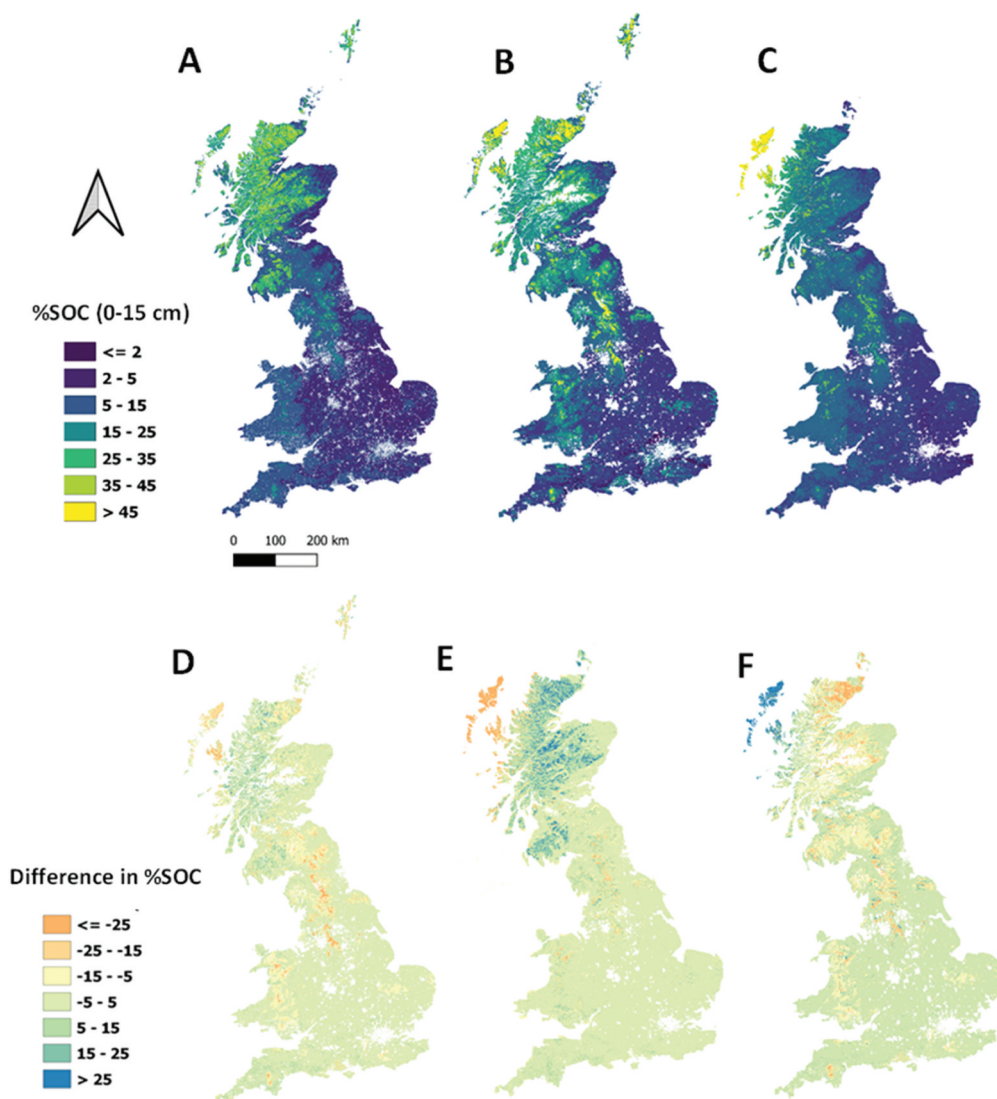


Figure 8. Spatial distribution of topsoil (0–15 cm) SOC concentration (%) across Great Britain: (A) predicted at 5 m spatial resolution—this study; (B) Countryside survey (0–15 cm) SOC (%) map (1 km spatial resolution); (C) LUCAS (0–15 cm) SOC (%) map (500 m spatial resolution). Differences between our predicted map and (D) Countryside survey map and (E) LUCAS map; (F) difference between LUCAS and Countryside maps.

those LC classes, and their relative contributions, in our study are close to the values reported in Bradley, Milne (Bradley et al., 2005).

4. Discussion

4.1. Internal model evaluation

Random Forest with selected predictors performed well in predicting SOC (%) 0–15 cm, explaining 74% of the variability in SOC (%) 0–15 cm in the evaluation dataset, with a RMSE of 9.8 tC ha^{-1} (Table 1, Figure 4). Predicted SOC (%) 0–15 cm reflects the observed spatial pattern of SOC (%) 0–15 cm across GB and observed levels of SOC (%) 0–15 cm under different LC classes (Figure 5) and

captures their overall range of values (Figure 6). Relatively marginal under- or overestimation of SOC (%) 0–15 cm in the evaluation dataset could be partly due to the bias in the spatial distribution of the training dataset used for building the model (R.-M. Yang et al., 2016). The uncertainty in predicted SOC (%) 0–15 cm is especially evident in some parts of Scotland which are not covered by LUCAS sampling points. Indeed, most NSIS sampling points are relatively old and may add additional noise to the SOC dataset (Chapman et al., 2013). Differences in analytic methods also contributed to uncertainty and difference in SOC% between samples collected in 1978–1988 and resampled in 2008–2009 in Scotland (Lilly et al., 2010).

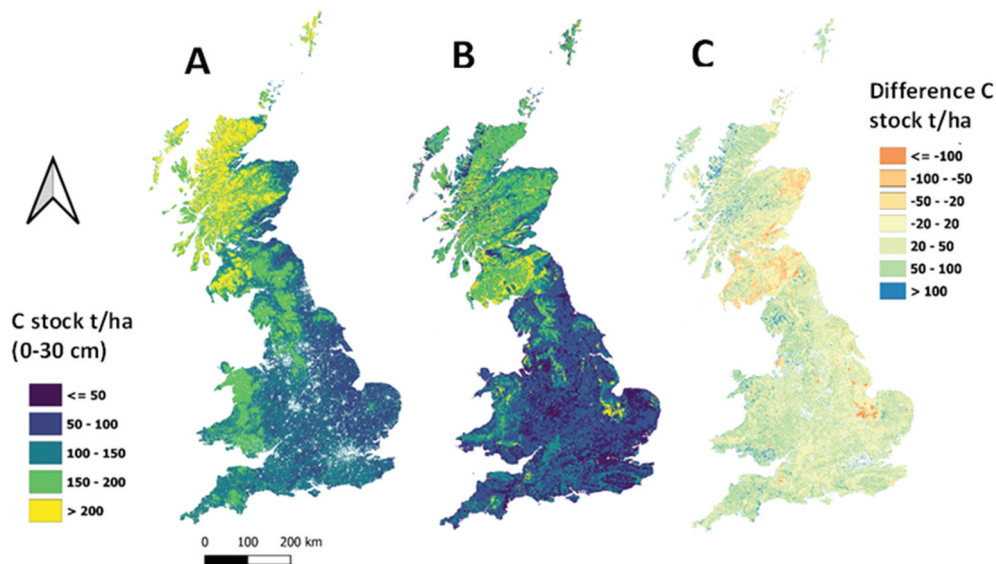


Figure 9. Spatial distribution of 0–30 cm depth SOC stock (C_s , $t\ ha^{-1}$) across Great Britain. (A) predicted at 5 m spatial resolution (this study); (B) map of Global Soil Organic Carbon (GSOC) at 1 km spatial resolution; (C) the difference between both.

Table 3. Summary statistics of predicted carbon stocks (C_s) 0–30 cm for common LC classes in Great Britain compared to those reported by Bradley, Milne (Yi et al., 2016) for the same LC classes. For further information on the LC classes, please see ‘data and methods’ and fig. S1

	LC classes	Predicted C_s 0–30 cm (this study)			Predicted C_s 0–30 cm [48]		
		Mean (t/ha)	Total (Tg)	% of total	Mean (t/ha)	Total (Tg)	% of total
England	Coniferous woodland	147.6	62	4.42	100	62	6.1
	Arable & horticulture	88.6	299.6	21.35	70	356	35.1
	Neutral grassland	129.2	522.1	37.2	120	435	42.9
	Total		1403.5			1015	
Wales	Coniferous woodland	160.9	28.4	10.04	120	26	13.4
	Arable & horticulture	110.7	6.6	2.34	70	5	2.6
	Neutral grassland	144.6	143.6	50.66	110	103	46.4
	Total		283.5			194	
Scotland	Coniferous woodland	183.4	134.6	13.24	170	113	9.7
	Arable & horticulture	102.8	49.4	4.86	120	84	7.2
	Neutral grassland	158.2	228	22.43	160	241	20.76
	Total		1016.8			1161	
Total for Great Britain			2703.8			2371	

4.2. Variable importance

Among the set of environmental variables used to predict SOC (%) 0–15 cm at 5 m spatial resolution, only eight variables were finally selected as predictors (Figure 7). Soil pH was the most important variable, followed by LC and climate variables (rainfall and mean annual temperature), while topographic variables (elevation and slope) were less important. Soil pH has a direct effect on SOC accumulation and is usually used as a proxy to determine the effect of land use changes, land management, climate and parent material on SOC accumulation (Malik et al., 2018). The importance of LC as a predictor for SOC may be due to the fact that it reflects the sharp changes in LC distinct from soil classes, potentially affecting the accumulation of SOC (Siewert, 2018). Other

similar studies (Vos et al., 2019) also show that LC, soil pH, and precipitation are important predictors of SOC.

4.3. External model evaluation

Our modelling approach provides for the first time a continuous map of SOC at 5 m spatial resolution for GB, which constitutes a large improvement in resolution in comparison with previous maps, which were much coarser (1 km to 100 m) (Aitkenhead & Coull, 2020; Bradley et al., 2005; Henrys et al., 2012; Poggio & Gimona, 2014). Whereas we were only able to explicitly compare our SOC (%) 0–15 cm and C_s 0–30 cm maps and summary statistics with those that were available (see section 3.4), it is important to note that other

SOC (%) maps exist in the literature and will also be discussed qualitatively here. The studies employed to produce maps before this study used different approaches to the one we used. For example, the map of the Countryside Survey represents SOC (%) 0–15 cm by habitat, where a single SOC (%) 0–15 cm value is attributed to the same habitat within 1 km square (Emmett et al., 2010). In The National Soil Map of Scotland, SOC (%) values were mapped across a 500 m grid by linking each grid cell to the average SOC (%) calculated from the representative soil profiles. These maps with coarse resolutions are less useful for assessing SOC (%) and stocks at fine scales. Nevertheless, they can be informative in providing a regional overview of where organic or mineral soil can occur (Baggaley et al., 2022).

Overall, our estimates of soil organic carbon (SOC (%)) and Cs) are comparable to those obtained by previous maps at national scales. In addition, our map also shows strong spatial patterns in SOC at finer scales in association with distinct LC classes. Our SOC (%) 0–15 cm map (Figure 8) follows a pattern similar to the spatial distribution and ranges of SOC (%) of the mean of eight GB maps reported by Feeney, Cosby (Feeney et al., 2022). Moreover, our SOC (%) 0–15 cm map compared well against available LUCAS and Countryside Survey SOC (%) 0–15 cm maps (Figure 8). However, our map slightly underestimated SOC (%) 0–15 cm in organic soils, especially in Scotland, and overestimated SOC (%) 0–15 cm in mineral soils, in a similar manner to the LUCAS and Countryside Survey maps (Feeney et al., 2022). In any case, the range of values, their distributions, and the overall spatial patterns of SOC in our maps are comparable to those produced by other studies that were available for direct comparison and provide at a much higher spatial resolution.

An important part of the disagreement between our maps for SOC (%) and Cs and others (GSOC and Bradley, Milne (Bradley et al., 2005)) may also be due to the high spatial resolution of our map, such that sharp changes in SOC (%) 0–15 cm and Cs 0–30 cm linked to our high spatial resolution LC are averaged out or masked in coarser resolution maps (Siewert, 2018). For example, the highest SOC (%) 0–15 cm in our map is ~53%, and values higher than 45% are less common in our map compared to the LUCAS and Countryside Survey maps (Figure 8), especially for large areas of Scotland. Furthermore, water bodies such as lakes, rivers and canals in the Countryside map have the same SOC (%) 0–15 cm value than the neighbouring lands within 1 km, which

resulted in high average values of SOC in this map compared to our map, which assumes no SOC (%) 0–15 cm in water bodies. Figure 10 provides a snapshot of zoomed-in small areas across GB to illustrate this. Since differences in spatial resolution are not visible in national or regional maps, a small area (~1.5×1.5 km) from each country (England, Wales and Scotland) was selected from our predicted Cs 0–30 cm map and compared with the GSOC 0–30 cm map. The maps of our predicted Cs 0–30 cm (Figure 10A–C) clearly show distinct values of Cs for different LC classes, with values being the highest for ‘Woodland’ (including hedgerows and lines of trees; >120 t ha⁻¹) and lowest for ‘Arable & Horticulture’ (<90 t ha⁻¹). On the other hand, GSOC maps (Figure 10D–F) show very low variability (or none) in Cs across the selected areas, especially in England and Wales (Figure 10D–F). The distinct patterns of our map for different LC classes are more evident when compared to the Google Earth very high resolution images (Figure 10G–I), while there is no evidence of any spatial association between GSOC Cs 0–30 cm maps and LC classes as seen in the Google images. These results strongly suggest that the higher spatial resolution of our map compared to the GSOC Cs 0–30 cm map may be an important source of disagreement in Cs. For example, the GSOC Cs 0–30 cm map shows lower values in the selected small areas in England and Wales, which are largely under conventional agriculture. Lower Cs values are expected for mineral soils in England and Wales and are predicted by our map as well (under the ‘Arable & Horticulture’ LC class). Higher Cs values for other land cover classes (shrubs, trees and hedgerows) present in the selected areas in our map were masked in the much coarser GSOC Cs 0–30 cm map. Other sources of disagreement in Cs 0–30 cm between maps, especially in Scotland, may be due to the uncertainty in soil bulk density used to calculate Cs in other studies, as some employed higher bulk density resulting in higher Cs values (Aitkenhead & Coull, 2020). Bulk density was also found to be the main source of uncertainty in Cs in organic soils between England and Scotland (Bradley et al., 2005; Fowler et al., 2023).

Another source of disagreement between the Cs 0–30 cm map in this study and that of Bradley, Milne (Bradley et al., 2005) might be the differences in LC classification, as Bradley, Milne (Bradley et al., 2005) predicted Cs 0–30 cm for 5 LC classes: ‘Seminatural’, ‘Woodland (Coniferous)’, ‘Pasture’, ‘Arable’, and ‘Gardens’, and our prediction is made for 10 habitats (Table 2). Furthermore, Bradley, Milne (Bradley et al., 2005) assumed a zero value of Cs in bare soil and some other LC classes associated with urban areas,

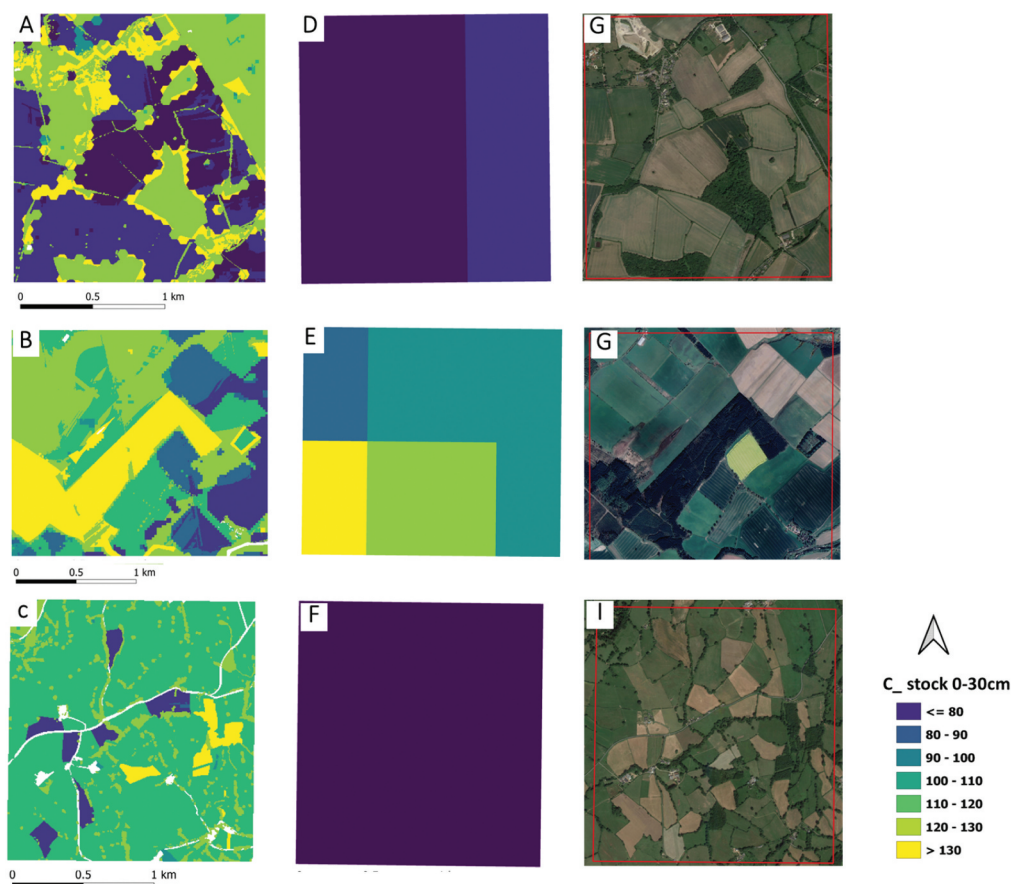


Figure 10. Maps of predicted 0–30 cm depth SOC stock (cs, t ha^{-1}) at 5 m spatial resolution (this study) for randomly selected areas in England, Scotland and Wales (A, B and C, respectively); GSOC maps for the same areas (D, E and F, respectively). G, H and I are google earth high resolution RGB images for these selected areas showing different land covers.

which resulted in underestimating the mean C_s value for arables and grasslands and consequently the total C_s in England and Wales. In addition, the coarse resolution of their map assumes the same land cover within each 1 km^2 . On the one hand, arable and natural grasslands are dominant LC classes in England and Wales and are usually associated with low SOC: these were assigned to 1 km^2 grid cells, ignoring potential SOC hotspots associated with LC occurring at a smaller spatial grain such as hedgerows, small woodlands and bog that might have occurred within the 1 km^2 grid cell. On the other hand, dominant LC classes in Scotland are bog and heathlands, which are usually associated with higher SOC. Therefore, our estimation is higher for England and Wales and lower for Scotland than that of the map produced by Bradley, Milne (Bradley et al., 2005). More generally, the disparity of LC classes used between the different studies makes comparisons difficult (see ‘Data & Methods’ and Fig. S1) and may contribute to disagreements between them.

The unavailability of a truly independent dataset to evaluate the map at national, regional and/or local scales is a limitation faced by this study (Rainford et al., 2021). Secondly, the unavailability of reliable national bulk density data is one of the main sources of uncertainty in C_s estimation (Aitkenhead & Coull, 2020; Bradley et al., 2005; Fowler et al., 2023). Assuming a negative relation between SOC and bulk density increases the uncertainty in estimating C_s (Fowler et al., 2023). Further, our estimation of bulk density is based on fine material only and thus our overall carbon stock estimates should be taken with caution in rocky soils. However, the high resolution of our map and its compatibility with other GB maps, especially with the mean of eight SOC% maps assessed in Feeney, Cosby (Feeney et al., 2022) and the estimated C_s for different LC classes (Bradley et al., 2005; Howard et al., 1995) suggest a greater utility of our map to understand the spatial distribution of C_s at finer scale across GB. Despite these potential limitations, our map provides the first baseline map to assess C_s within small geographical areas, such as a single farm, and even within a field.

Such data are essential to optimize soil sampling (e.g. they can be used to reduce the number of samples based on the spatial variability within a field) and to derive management measures that can increase SOC for sustainable land management. We also believe that as farmers and other stakeholders across Great Britain evaluate the map against their measurements of SOC, the baseline map can be updated accordingly.

5. Conclusions

Using a RF modelling approach, we found good performance in predicting soil organic carbon concentration and stock, and we have provided for the first time a continuous map of SOC at 5 m spatial resolution for Great Britain. Our results are in line with other studies (Ayala Izurieta et al., 2021; Rainford et al., 2021; Siewert, 2018; R.-M. Yang et al., 2016) in finding RF as a robust and reliable algorithm to estimate SOC (McBratney et al., 2003). We found that soil pH and LC class are the most important variables for predicting SOC and our model output allowed us to estimate that GB soils hold approximately 2704Tg of organic carbon in the 0–30 cm depth, with an average of 154 t ha⁻¹ SOC, with 1403Tg in England, 283Tg in Wales and 1017Tg in Scotland. Overall, ‘Natural Grasslands’ make the highest relative contribution of SOC in England and Wales with 37.2% and 50.7% respectively, whereas ‘Dwarf Heath Shrubs’, ‘Bog & Bog Blanket’ and ‘Natural Grasslands’ have relative contributions ranging from 22% to 25% each in Scotland. Our estimation of Cs is comparable with previous studies conducted at national scales, reflecting the expected distribution of Cs across GB and for different LC classes. The high spatial resolution of our map and its compatibility with other maps at national scales suggest its applicability for assessing SOC at smaller scales such as a single farm or field.

Acknowledgements

We would like to thank Tom Ashworth and Simon Eberz for their technical assistance on this paper, and the NatCap team for their thoughtful feedback throughout.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research has been funded by Natural Capital Research Ltd, which holds financial interest from the direct applications of the study.

ORCID

Salar Mahmood  <http://orcid.org/0000-0001-8223-9303>
 Beccy Wilebore  <http://orcid.org/0000-0002-0371-2500>
 Rhosanna Jenkins  <http://orcid.org/0000-0001-7970-5262>
 Katherine J. Willis  <http://orcid.org/0000-0002-6763-2489>
 Marc Macias-Fauria  <http://orcid.org/0000-0002-8438-2223>
 Shibu M. Ebrahim  <http://orcid.org/0000-0001-5065-1950>

Data availability statement

The data that support the findings of this study are available on request from B.W., and will be made available for non-commercial use under CC BY-NC-SA (Attribution-NonCommercial-ShareAlike 4.0 International) licence.

References

- Aitkenhead, M., & Coull, M. (2020). Mapping soil profile depth, bulk density and carbon stock in Scotland using remote sensing and spatial covariates. *European Journal of Soil Science*, 71(4), 553–567. <https://doi.org/10.1111/ejss.12916>
- Arthur Endsley, K., Kimball, J. S., Reichle, R. H., & Watts, J. D. (2020). Satellite monitoring of global surface soil organic carbon dynamics using the SMAP level 4 carbon product. *Journal of Geophysical Research Biogeosciences*, 125(12), e2020JG006100. <https://doi.org/10.1029/2020JG006100>
- Ayala Izurieta, J. E., Márquez, C. O., García, V. J., Jara Santillán, C. A., Sisti, J. M., Pasqualotto, N., Van Wittenberghe, S., & Delegido, J. (2021). Multi-predictor mapping of soil organic carbon in the alpine tundra: A case study for the central Ecuadorian páramo. *Carbon balance and management*, 16(1), 32. <https://doi.org/10.1186/s13021-021-00195-2>
- Baggaley, N., Britton, A., Sandison, F., Lilly, A., Stutter, M., Rees, R.M., Reed, M. & Buckingham, S. (2022). *Understanding carbon sequestration from nature-based solutions*. The James Hutton Institute.
- Ballabio, C., Lugato, E., Fernández-Ugalde, O., Orgiazzi, A., Jones, A., Borrelli, P., Montanarella, L., & Panagos, P. (2019 December 1). Mapping LUCAS topsoil chemical properties at European scale using gaussian process regression. *Geoderma*, 355, 113912. 1. <https://doi.org/10.1016/j.geoderma.2019.113912>
- Ballabio, C., Panagos, P., & Montanarella, L. (2016). Mapping topsoil physical properties at European scale using the LUCAS database. *Geoderma*, 261, 110–123. <https://doi.org/10.1016/j.geoderma.2015.07.006>
- Batjes, N. H. (1996). Total carbon and nitrogen in the soils of the world. *European Journal of Soil Science*, 47(2), 151–163. Reflections by N.H. Batjes. *European Journal of Soil Science*, 2014;65(1):2-3. <https://doi.org/10.1111/j.1365-2389.1996.tb01386.x>
- Bossio, D. A., Cook-Patton, S. C., Ellis, P. W., Fargione, J., Sanderman, J., Smith, P., Wood, S., Zomer, R. J., von Unger, M., Emmer, I. M., & Griscom, B. W. (2020). The role of soil carbon in natural climate solutions. *Nature Sustainability*, 3(5), 391–398. <https://doi.org/10.1038/s41893-020-0491-z>

- Bradley, R. I., Milne, R., Bell, J., Lilly, A., Jordan, C., & Higgins, A. (2005). A soil carbon and land use database for the United Kingdom. *Soil Use and Management*, 21(4), 363–369. <https://doi.org/10.1079/SUM2005351>
- Breiman L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Carag HR, Prieto VG, Yballe LS, Shea C. (2022). Sources TJIHod.
- Chapman, S. J., Bell, J. S., Campbell, C. D., Hudson, G., Lilly, A., Nolan, A. J., Robertson, A. H. J., Potts, J. M., & Towers, W. (2013). Comparison of soil carbon stocks in Scottish soils between 1978 and 2009. *European Journal of Soil Science*, 64(4), 455–465. <https://doi.org/10.1111/ejss.12041>
- Emadi, M., Taghizadeh-Mehrjardi, R., Cherati, A., Danesh, M., Mosavi, A., & Scholten, T. (2020) [cited]. Predicting and mapping of soil organic carbon using machine learning algorithms in northern Iran. *Remote Sensing*, 12(14), 2234. <https://doi.org/10.3390/rs12142234>
- Emmett, B. A., Reynolds, B., Chamberlain, P. M., Rowe, E., Spurgeon, D., Brittain, S.A., Frogbrook, Z., Hughes, S., Lawlor, A.J., Poskitt, J. and Potter, E. (2010). *Countryside survey: Soils report from 2007*. p. 192. Centre for Ecology & Hydrology (Natural Environment Research Council).
- Eu DEM. (2018). Eu-digital elevation model (DEM).
- Falloon, P., Smith, P., Bradley, R. I., Milne, R., Tomlinson, R., Viner, D., Livermore, M., & Brown, T. (2006). RothCUK – a dynamic modelling system for estimating changes in soil C from mineral soils at 1-km resolution in the UK. *Soil Use and Management*, 22(3), 274–288. <https://doi.org/10.1111/j.1475-2743.2006.00028.x>
- FAO. (2018). Global soil organic carbon Map (GSOCmap): Technical report. In FAO (Ed. (p. 167).
- Feeney, C. J., Cosby, B. J., Robinson, D. A., Thomas, A., Emmett, B. A., & Henrys, P. (2022). Multiple soil map comparison highlights challenges for predicting topsoil organic carbon concentration at national scale. *Scientific Reports*, 12(1), 1379. <https://doi.org/10.1038/s41598-022-05476-5>
- Fowler, A. F., Basso, B., Millar, N., & Brinton, W. F. (2023). A simple soil mass correction for a more accurate determination of soil carbon stock changes. *Scientific Reports*, 13(1), 2242. <https://doi.org/10.1038/s41598-023-29289-2>
- Guevara, M., Olmedo, G. F., Stell, E., Yigini, Y., Aguilar Duarte, Y., Arellano Hernández, C., Arévalo, G. E., Arroyo-Cruz, C. E., Bolivar, A., Bunning, S., Bustamante Cañas, N., Cruz-Gaistardo, C. O., Davila, F., Dell Acqua, M., Encina, A., Figueredo Tacona, H., Fontes, F., Hernández Herrera, J. A. . . . Vargas, R. R. (2018). No silver bullet for digital soil mapping: Country-specific soil organic carbon estimates across Latin America. *SOIL*, 4(3), 173–193. <https://doi.org/10.5194/soil-4-173-2018>
- Hendriks, C. M. J., Stoorvogel, J. J., Álvarez-Martínez, J. M., Claessens, L., Pérez-Silos, I., & Barquín, J. (2021). Introducing a mechanistic model in digital soil mapping to predict soil organic matter stocks in the Cantabrian region (Spain). *European Journal of Soil Science*, 72(2), 704–719. <https://doi.org/10.1111/ejss.13011>
- Hengl, T., de Jesus, J. M., MacMillan, R. A., Batjes, N. H., Heuvelink, G. B. M., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J. G. B., Walsh, M. G., & Gonzalez, M. R. (2014). SoilGrids1km — global soil information based on automated mapping. *PLOS ONE*, 9(8), e105992. <https://doi.org/10.1371/journal.pone.0105992>
- Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., & Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, 12(2), e0169748. <https://doi.org/10.1371/journal.pone.0169748>
- Henrys, P. A., Keith, A. M., Robinson, D. A., Thomas, A., Emmett, B.A. & Henrys, P. (2012). *Model estimates of topsoil carbon [countryside survey]*. NERC Environmental Information Data Centre.
- Heuvelink, G. B. M., Angelini, M. E., Poggio, L., Bai, Z., Batjes, N. H., van den Bosch, R., Bossio, D., Estella, S., Lehmann, J., Olmedo, G. F., & Sanderman, J. (2021). Machine learning in space and time for modelling soil organic carbon change. *European Journal of Soil Science*, 72(4), 1607–1623. <https://doi.org/10.1111/ejss.12998>
- Hilinski, T. E. (2001). *Implementation of exponential depth distribution of organic carbon in the CENTURY model. Dep of soil and Crop Sci.* Colorado State Univ, Fort Collins.
- Hollis, D., McCarthy, M., Kendon, M., Legg, T., & Simpson, I. (2019). HadUK-Grid—A new UK dataset of gridded climate observations. *Geoscience Data Journal*, 6(2), 151–159. <https://doi.org/10.1002/gdj3.78>
- Howard, P., Loveland, P., Bradley, R., Dry, F. T., Howard, D. M., & Howard, D. C. (1995). The carbon content of soil and its geographical distribution in Great Britain. *Soil Use and Management*. 1995 11(1), 9–15. <https://doi.org/10.1111/j.1475-2743.1995.tb00488.x>
- IPCC. (2014). Climate change 2014: Impacts, adaptation, and vulnerability. Part A: Global and sectoral aspects. Contribution of working group II to the fifth assessment report of the intergovernmental panel on climate change. In C. Field, V. Barros, D. Dokken, Eds. Cambridge, United Kingdom and New York, NY, USA. (p. 1132).
- Jones, A., Fernandez-Ugalde, O., & Scarpa, S. (2020). *Lucas 2015 topsoil survey. Presentation of dataset and results.* 2020 EUR 30332 EN.
- Kempen, B., Dalsgaard, S., Kaaya, A. K., Chamuya, N., Ruipérez-González, M., Pekkarinen, A., & Walsh, M. G. (2019). Mapping topsoil organic carbon concentrations and stocks for Tanzania. *Geoderma*, 337, 164–180. <https://doi.org/10.1016/j.geoderma.2018.09.011>
- Leigty, W. R., & Shorey, E. C. (1930). Some carbon-nitrogen relations in soils. *Soil Science*. 1930 30(1), 257–266. <https://doi.org/10.1097/00010694-193010000-00002>
- Lilly, A., Bell, J. S., Hudson, G., Nolan, A. J., & Towers, W. (2010). National soil inventory of Scotland 1 (NSIS_1): Site location, sampling and profile description protocols (1978–1988). *Technical Bulletin*.
- Malik, A. A., Puissant, J., Buckeridge, K. M., Goodall, T., Jehmlich, N., Chowdhury, S., Gweon, H. S., Peyton, J. M., Mason, K. E., van Agtmaal, M., Blaud, A., Clark, I. M., Whitaker, J., Pywell, R. F., Ostle, N., Gleixner, G., & Griffiths, R. I. (2018). Land use driven change in soil pH affects microbial carbon cycling processes. *Nature Communications*, 9(1), 3591. <https://doi.org/10.1038/s41467-018-05980-1>
- Manrique, L. A., & Jones, C. A. (1991). Bulk density of soils in relation to soil physical and chemical properties. *Soil*

- Science Society of America Journal*, 55(2), 476–481. <https://doi.org/10.2136/sssaj1991.03615995005500020030x>
- McBratney, A. B., Santos, M., & Minasny, B. (2003). On digital soil mapping. *Geoderma*, 117(1–2), 3–52. [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4)
- Meersmans, J., Martin, M. P., Lacarbe, E., De Baets, S., Jolivet, C., Boulonne, L., Lehmann, S., Saby, N. P. A., Bispo, A., & Arrouays, D. (2012). A high resolution map of French soil organic carbon. *Agronomy for Sustainable Development*, 32(4), 841–851. <https://doi.org/10.1007/s13593-012-0086-9>
- Mishra, U., Lal, R., Slater, B., Calhoun, F., Liu, D., & Van Meirvenne, M. (2009). Predicting soil organic carbon stock using profile depth distribution functions and ordinary Kriging. *Soil Science Society of America Journal*, 73(2), 614–621. <https://doi.org/10.2136/sssaj2007.0410>
- Muhammed, S. E., Coleman, K., Wu, L., Bell, V. A., Davies, J. A. C., Quinton, J. N., Carnell, E. J., Tomlinson, S. J., Dore, A. J., Dragosits, U., Naden, P. S., Glendining, M. J., Tipping, E., & Whitmore, A. P. (2018). Impact of two centuries of intensive agriculture on soil carbon, nitrogen and phosphorus cycling in the UK. *Science of the Total Environment*, 634, 1486–1504. <https://doi.org/10.1016/j.scitotenv.2018.03.378>
- Ottoy, S., Elsen, A., Van De Vreken, P., Gobin, A., Merckx, R., Hermy, M., & Van Orshoven, J. (2016). An exponential change decline function to estimate soil organic carbon stocks and their changes from topsoil measurements. *European Journal of Soil Science*. 2016 67(6), 816–826. <https://doi.org/10.1111/ejss.12394>
- Pierson, D., Lohse, K. A., Wieder, W. R., Patton, N. R., Facer, J., de Graaff, M.-A., Georgiou, K., Seyfried, M. S., Flerchinger, G., & Will, R. (2022). Optimizing process-based models to predict current and future soil organic carbon stocks at high-resolution. *Scientific Reports*, 12(1), 10824. <https://doi.org/10.1038/s41598-022-14224-8>
- Poggio, L., & Gimona, A. (2014). National scale 3D modelling of soil organic carbon stocks with uncertainty propagation — an example from Scotland. *Geoderma*, 232–234, 284–299. <https://doi.org/10.1016/j.geoderma.2014.05.004>
- Proctor, M. E., Siddons, P. A., Jones, R. J. A., Bellamy, P. H., & Keay, C. A. (1998). LandIS—a land information system for the UK. Developments for planning the sustainable use of land resources. *European Soil Bureau Research Report*, 4, 219–233.
- Rainford, S.-K., Martín-López, J. M., & Da Silva, M. (2021). Approximating soil organic carbon stock in the Eastern Plains of Colombia. *Frontiers in Environmental Science*, 9, 685819. <https://doi.org/10.3389/fenvs.2021.685819>
- Scharlemann, J. P. W., Tanner, E. V. J., Hiederer, R., & Kapos, V. (2014). Global soil carbon: Understanding and managing the largest terrestrial carbon pool. *Carbon Management*, 5(1), 81–91. <https://doi.org/10.4155/cmt.13.77>
- Schmidt, M. W. I., Torn, M. S., Abiven, S., Dittmar, T., Guggenberger, G., Janssens, I. A., Kleber, M., Kögel-Knabner, I., Lehmann, J., Manning, D. A. C., Nannipieri, P., Rasse, D. P., Weiner, S., & Trumbore, S. E. (2011). Persistence of soil organic matter as an ecosystem property. *Nature*, 478(7367), 49–56. <https://doi.org/10.1038/nature10386>
- Siewert, M. B. (2018). High-resolution digital mapping of soil organic carbon in permafrost terrain using machine learning: A case study in a sub-arctic peatland environment. *Biogeosciences*, 15(6), 1663–1682. <https://doi.org/10.5194/bg-15-1663-2018>
- Tao, F., Zhou, Z., Huang, Y., Li, Q., Lu, X., Ma, S., Huang, X., Liang, Y., Hugelius, G., Jiang, L., Doughty, R., Ren, Z., & Luo, Y. (2020). Deep learning optimizes data-driven representation of soil organic carbon in Earth system model over the conterminous United States [original research]. *Frontiers in Big Data*, 3, 3. <https://doi.org/10.3389/fdata.2020.00017>
- Thaler, E. A., Larsen, I. J., & Yu, Q. (2019). A new index for remote sensing of soil organic carbon based solely on visible wavelengths. *Soil Science Society of America Journal*, 83(5), 1443–1450. <https://doi.org/10.2136/sssaj2018.09.0318>
- Thomas, A., Cosby, B. J., Henrys, P., & Emmett, B. (2020). Patterns and trends of topsoil carbon in the UK: Complex interactions of land use change, climate and pollution. *Science of the Total Environment*, 729, 138330. <https://doi.org/10.1016/j.scitotenv.2020.138330>
- UK Habitat Classification Working Group. (2018). UK Habitat Classification - Habitat Definitions V1.0. <http://ecountability.co.uk/ukhabworkinggroup-ukhab>
- Ungaro, F., Staffilani, F., & Tarocco, P. (2010). Assessing and mapping topsoil organic carbon stock at regional scale: A scorpan kriging approach conditional on soil map delineations and land use. *Land Degradation & Development*, 21(6), 565–581. <https://doi.org/10.1002/ldr.998>
- Vos, C., Don, A., Hobbey, E. U., Prietz, R., Heidkamp, A., & Freibauer, A. (2019). Factors controlling the variation in organic carbon stocks in agricultural soils of Germany. *European Journal of Soil Science*, 70(3), 550–564. <https://doi.org/10.1111/ejss.12787>
- Wadoux, A. M. J. C., Román Dobarco, M., Malone, B., Minasny, B., McBratney, A. B., & Searle, R. (2023). Baseline high-resolution maps of organic carbon content in Australian soils. *Scientific Data*, 10(1), 181. <https://doi.org/10.1038/s41597-023-02056-8>
- Wang, L., Li, Z., Wang, D., Liao, S., Nie, X., & Liu, Y. (2022). Factors controlling soil organic carbon with depth at the basin scale. *Catena* 217, 106478. <https://doi.org/10.1016/j.catena.2022.106478>
- Yang, L., Cai, Y., Zhang, L., Guo, M., Li, A., & Zhou, C. (2021). A deep learning method to predict soil organic carbon content at a regional scale using satellite-based phenology variables. *International Journal of Applied Earth Observation and Geoinformation*, 102, 102428. <https://doi.org/10.1016/j.jag.2021.102428>
- Yang, R.-M., Zhang, G.-L., Liu, F., Lu, Y.-Y., Yang, F., Yang, F., Yang, M., Zhao, Y.-G., & Li, D.-C. (2016). Comparison of boosted regression tree and random forest models for mapping topsoil organic carbon concentration in an alpine ecosystem. *Ecological Indicators*. 2016 60, 870–878. <https://doi.org/10.1016/j.ecolind.2015.08.036>
- Yi, X., Li, G., & Yin, Y. (2016). Pedotransfer functions for estimating soil bulk density: A case study in the Three-River Headwater Region of Qinghai Province,

- China. *Pedosphere*, 26(3), 362–373. [https://doi.org/10.1016/S1002-0160\(15\)60049-2](https://doi.org/10.1016/S1002-0160(15)60049-2)
- Zellweger, F., Flack-Prain, S., Footring, J., Wilebore, B., & Willis, K. J. (2022). Carbon storage and sequestration rates of trees inside and outside forests in Great Britain. *Environmental Research Letters*, 17(7), 074004. <https://doi.org/10.1088/1748-9326/ac74d5>
- Zeraatpisheh, M., Garosi, Y., Owliaie, H. R., Ayoubi, S., Taghizadeh-Mehrjardi, R., Scholten, T., & Xu, M. (2022). Improving the spatial prediction of soil organic carbon using environmental covariates selection: A comparison of a group of environmental covariates. *Catena*, 208, 105723. <https://doi.org/10.1016/j.catena.2021.105723>
- Zhou, Y., Chartin, C., Van Oost, K., & van Wesemael, B. (2022). High-resolution soil organic carbon mapping at the field scale in southern Belgium (Wallonia). *Geoderma*, 422, 115929. <https://doi.org/10.1016/j.geoderma.2022.115929>