Science of the Total Environment 886 (2023) 163973

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Development of soil health benchmarks for managed and semi-natural landscapes

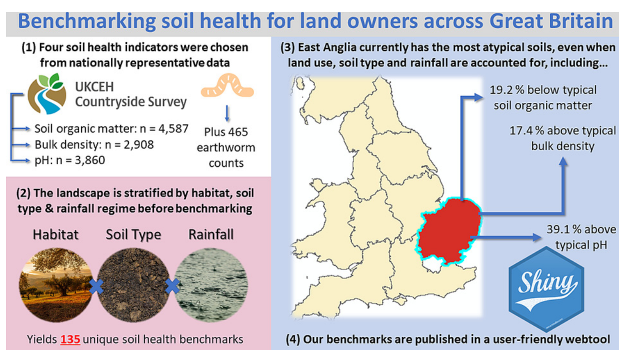
Christopher J. Feeney [a,*], David A. Robinson [a], Aidan M. Keith [b], Audric Vigier [c], Laura Bentley [a], Richard P. Smith [d], Angus Garbutt [a], Lindsay C. Maskell [b], Lisa Norton [b], Claire M. Wood [b], B. Jack Cosby [a], Bridget A. Emmett [a]

[a] UK Centre for Ecology and Hydrology, Environment Centre Wales, Deiniol Road, Bangor, Gwynedd LL57 2UW, UK
[b] UK Centre for Ecology and Hydrology, Lancaster Environment Centre, Library Avenue, Bailrigg, Lancaster LA1 4AP, UK
[c] UK Centre for Ecology and Hydrology, Bush Estate, Penicuik, Midlothian EH26 0QB, UK
[d] Environment Agency, Manley House, Kestrel Way, Exeter EX2 7LQ, UK

HIGHLIGHTS

- Soil health is benchmarked for landscapes defined by habitat, soil type and rainfall.
- Middle 80 % of measurements of 4 indicators define 135 soil health benchmarks.
- Generally, BD and pH decrease with land management intensity, SOM and EA increase.
- East Anglian soils are the most atypical compared to similar landscapes nationwide.
- Benchmarks feature on an app for landholders to assess their soil health condition.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Paulo Pereira

Keywords:
Soil organic matter
pH
Bulk density
Earthworm abundance
Ecosystem services
Land use

ABSTRACT

Efforts to improve soil health require that target values of key soil properties are established. No agreed targets exist but providing population data as benchmarks is a useful step to standardise soil health comparison between landscapes. We exploited nationally representative topsoil (0–15 cm) measurements to derive soil health benchmarks for managed and semi-natural environments across Great Britain. In total, 4587 soil organic matter (SOM), 3860 pH, 2908 bulk density (BD), and 465 earthworm abundance (EA) datapoints were used. As soil properties are sensitive to site-specific characteristics, data were stratified by habitat, soil type, and mean annual precipitation, with benchmarks defined as the middle 80 % of values in each distribution – yielding 135 benchmarks. BD and pH decreased with land management intensity (agriculture > semi-natural grasslands > woodlands > heathlands > wetlands), and vice versa for SOM and EA. Normalising benchmark ranges by medians revealed soil health indicator benchmark widths increased in the order: pH < BD < SOM < EA, while width increased with decreasing land management intensity. Arable and horticulture and improved grassland exhibited narrow benchmarks for SOM, pH and BD, yet the widest EA benchmark, suggesting additional drivers impact EA patterns. Upland wetlands had the widest BD benchmarks, important when determining carbon stocks. East Anglia currently possesses the largest proportions of atypical soils, including below typical SOM (19.2 %), above typical BD (17.4 %) and pH (39.1 %), and the smallest proportions of above typical SOM (2.4 %), and below typical BD (5.8 %) and pH (2.3 %). This is found even after land use, soil type and rainfall have been considered, underscoring how urgently soil health should be addressed here. Our benchmarking framework allows landowners to compare where their measured soil health indicators fall within expected ranges and is applicable to other biomes, national and multinational contexts.

* Corresponding author.
E-mail address: chrfee@ceh.ac.uk (C.J. Feeney).

http://dx.doi.org/10.1016/j.scitotenv.2023.163973
Received 22 February 2023; Received in revised form 28 April 2023; Accepted 2 May 2023
Available online xxxx

1. Introduction

Soils underpin land use, and with that, the provision of myriad ecosystem goods and services upon which civilisation ultimately depends (Haygarth and Ritz, 2009). Increasing demands for food, fibre and fuel production, coupled with global environmental challenges such as climate change are placing soils under unprecedented threat (Amundson et al., 2015; Lewis, 2020).

There has been an increasing push from scientists to assess soil health. Generally, soil health has been characterised as promoting the continued capacity of the soil to function as a vital living ecosystem that sustains plants and animals, environmental quality and human needs (Doran and Zeiss, 2000; Lal, 2016; Lehmann et al., 2020; Moebius-Clune et al., 2017). The European Union (EU) Mission Board for Soil Health and Food defines soil heath as: "the continued capacity of soils to support ecosystem services, in line with the [United Nations] Sustainable Development Goals and the [EU] Green Deal" (Veerman et al., 2020, p. 5). Soil health therefore focuses on soils being fit for purpose, and this should encapsulate the multi-functionality of soils; in particular, the capacity of soils to deliver regulating, supporting, cultural and provisioning ecosystem functions and services (Haygarth and Ritz, 2009).

Concerningly, 60–70 % of Europe's soils have been rated as unhealthy (Veerman et al., 2020), with a third of soils worldwide estimated to be degraded to the point of losing organic carbon rapidly (FAO and ITPS, 2015). In response to the perceived global soil crisis, many governments are now responding and setting targets for restoration and providing greater legal protection. The UK Department for Environment, Food and Rural Affairs (Defra) announced in their 25-year environment plan their intentions to develop a soil health index for soils in England (Defra, 2018). Meanwhile, the EU is in consultation with the public to establish a new EU Soil Health Law in 2023 (Directorate-General for Environment, 2022).

Underpinning these policy aims is a need to develop benchmarks of proxy indicators of soil health. Benchmarks are generated from representative datasets which allow for an indicative comparison with regionally representative measured values, but do not allow for a direct evaluation of specific soil functions (Bünemann et al., 2018; Verheijen et al., 2005). While the use of benchmarks has been characterised as "reductionist" (Kibblewhite et al., 2008) for failing to assess healthy soil function holistically (Harris et al., 2022), the alternative "integrated" approach of measuring the flow of energy and carbon between soil functions (Kibblewhite et al., 2008) is difficult to achieve given the multitude of processes in question. A further complication is that soil properties show high spatial variability. Therefore, a single reference value for a large polity is inappropriate, and one must instead consider key environmental controls such as soil type, land-use and climate.

Soil health assessment should ideally encompass physical structure, biological condition, and chemical composition (Egan and Bay, 2021). To cover each of these overarching characteristics comprehensively would require several dozen soil properties to be measured. This is unlikely to be workable, especially if benchmarking soil health is to be undertaken to encourage land managers to monitor their own soils regularly. While clustering, correlation, or principal components analysis can distil several soil properties into a small number of key soil health indicators (e.g. Rinot et al., 2019), these approaches can conflict with priority considerations such as ease of sampling and interpretation, and sensitivity to land use and management practices (Bünemann et al., 2018). The 5 most popularly proposed soil quality and health indicators are soil organic matter (SOM), pH, available P, water storage and bulk density (BD) (Bünemann et al., 2018; each indicator appeared in >50 % of reviewed literature therein). These indicators are useful metrics for key soil functions including soil fertility, nutrient cycling, carbon storage, habitat for biological activity, and water storage and filtration (Vogel et al., 2019).

Several publications have proposed standardised approaches to benchmarking soil health, including estimating "indicative soil organic carbon (SOC) management ranges" for different "physiotopes" (landscape units defined by land-use, soil type and precipitation) (Verheijen et al., 2005; Drexler et al., 2022); and using SOC/clay ratios (Prout et al., 2021). Indicative soil health benchmarks have also been established for several other soil properties in addition to SOC for physiotopes across the Netherlands (Rutgers et al., 2008, 2009). Griffiths et al. (2018) developed soil health scorecards, using a traffic light system colour-coded according to management risk, which follows similar efforts from the USA (e.g. Moebius-Clune et al., 2017). Most of these studies focus on soil health for agriculture, neglecting the wider semi-natural environment. The lack of consideration of semi-natural habitats, such as wetlands, grasslands and woodlands, represents both a major research gap and a lack of policy focus.

The aim of this study, therefore, was to establish benchmarks for multiple soil health indicators across the breadth of soils and land uses across Great Britain (GB), including semi-natural habitats. Core objectives included (1) identifying the range of physiotopes that soil health benchmarks could be generated for; (2) defining benchmarks from easily calculable statistics for multiple soil properties, and (3) contextualising our benchmarks by referring to known trends in soil health indicators over space and time in GB. We chose to assess soil health across GB because it has a huge diversity of soils, with over 700 unique soil series in England and Wales alone (Avery, 1973; National Soil Resources Institute, 2001). Additionally, GB benefits from a methodologically coherent state and change database of several soil properties (the UKCEH Countryside Survey, CS) covering >40 years of repeat nationwide monitoring (Reynolds et al., 2013). Finally, these benchmarks represent the first soil health metrics of their kind to have been established for all of GB.

2. Data and methods

2.1. Datasets

To derive unbiased benchmarks of soil health, nationally representative records of physical, chemical, and biological properties of soils need to be sourced. In addition, the landscape needs to be partitioned into smaller units, or "physiotopes" that reflect key soil formation factors (Jenny, 1941). The current status of national topsoil health will likely vary at least as much between different habitats and land uses as it does between different soil types (e.g. Simfukwe et al., 2010). Thresholds may also be derived for commonly measured climatic variables (e.g. temperature and rainfall) to separate parts of the landscape from one another. While several other variables may significantly influence soil properties, we wanted to restrict our analysis to environmental variables that can be easily determined by non-specialists, and ensure we had sufficient sample sizes to generate benchmark distributions for a defined physiotope.

CS is an integrated national monitoring program that regularly assesses the condition of the vegetation, land use, water quality and topsoil (0–15 cm) of GB (Carey et al., 2008). The CS dataset contains thousands of measurements of physical, chemical, and biological topsoil properties from surveys conducted in 1978, 1998, 2007, and an ongoing reduced annual survey from 2019 to 23. Further, CS is built on a stratified random sampling design based on an underlying land classification of GB that ensures all major land classes are represented proportionately (Bunce et al., 2007). As new soils data for the upcoming Northern Ireland Countryside Survey become available, it should become possible to establish UK-wide soil health benchmarks. Here, we use measures of soil organic matter (SOM; recorded as the % loss on ignition), pH (measured in suspension of deionised water) and bulk density (BD; recorded as the density of the oven dry fine earth fraction in g cm^{-3}) from all years where data are available to create benchmarks. For more information on CS and how topsoil properties were measured, see Text S1 (Supplementary Materials).

Quantifying soil health has tended to ignore the importance of soil biodiversity, owing to limited functional knowledge and a lack of effective methods (Lehmann et al., 2020). Earthworm abundance (EA) represents an exception to most other soil biology metrics: it is an easy indicator to measure, and the benefits earthworms bring to physical soil structure (Keith and Robinson, 2012) and plant nutrition (van Groenigen et al., 2014) make this metric simple to convey to non-specialists. Earthworm abundance (EA; recorded as the number of worms per 20 × 20 cm spadeful of soil) data were extracted from relevant sources given in a compendium of datasets and peer reviewed publications from across the British Isles (Mason et al., 2022) plus additional data from the #WorldWormWeek farmland earthworm survey at Rothamsted Research Farm in 2019 (Stroud and Goulding, 2022). Records used for benchmarks here include 465 datapoints of mean EA, with means derived from differing numbers of replicates depending on the study/data source (see Text S1; Supplementary Materials).

The UK Biodiversity Action Plan (BAP) Broad Habitat (Jackson, 2000) is recorded for each CS topsoil sample to indicate the local vegetation and land use at the time of sampling. Here, we classified the habitats of each topsoil sample by defining 11 habitat groups based on aggregations of some of the BAP Broad Habitats (Fig. 1a; see also Table S1; Supplementary Materials). A key advantage of this habitat classification is that relatively uncommon BAP Broad Habitats with similar soil characteristics are aggregated together. This allows habitats with relatively little topsoil data to be included.

The National Soil Map of England and Wales (NATMAP) represents nearly 300 soil associations across England and Wales (National Soil Resources Institute, 2001). These soil associations form part of a hierarchy of soil subgroups, soil groups and major soil groups (Avery, 1980). Scottish soils are classified in a similar way for the National Soil Map of Scotland (Soil Survey of Scotland Staff, 1984). Here, we aggregated soil subgroups into 12 new soil types covering the whole of GB (Fig. 1b). These new soil types are aggregated based on organic matter content, depth, flood risk, clay content, drainage characteristics, and degree of modification from industrial activity (see Table S2; Supplementary Materials). With these new soil types, we created two new maps: a 1:250,000 scale map of GB, with soil type defined at the association level (i.e. by the soil type of the dominant soil series); and a 1:10,000 scale map at the level of 1 km^2 CS squares where soil series was mapped in the year 1990. Further details on the creation of these two maps can be found in the Supplementary Materials (Text S2). CS sample points from all years were intersected with the 1:10,000 scale map to derive soil types in R using the "st_intersection" function from the "sf" package (Pebesma et al., 2023). As the number of 1 km^2 squares surveyed by CS has changed over the years (see Text S2; Supplementary Materials), it was not possible to assign a soil type to all CS samples. Therefore, some records were discarded at this point, leaving 4857, 3860 and 2908 SOM, pH and BD records, respectively. Soil types at EA sample locations were derived using information in data sources or associated publications; where this was not possible intersecting EA point locations with the 1:250 k soil map was performed to derive the best possible estimate of soil type.

Mean annual rainfall rates (mm yr^{-1}) and mean daily temperatures (°C) recorded in the UKCEH Climate, Hydrology and Ecology research Support System (CHESS) dataset (Robinson et al., 2020) were used to represent climatic controls. CHESS data are represented by 1 km^2 gridded rasters at daily resolution, covering the period 01/01/1961-31/12/2017. Here, we summarised the CHESS precipitation and temperature data to obtain

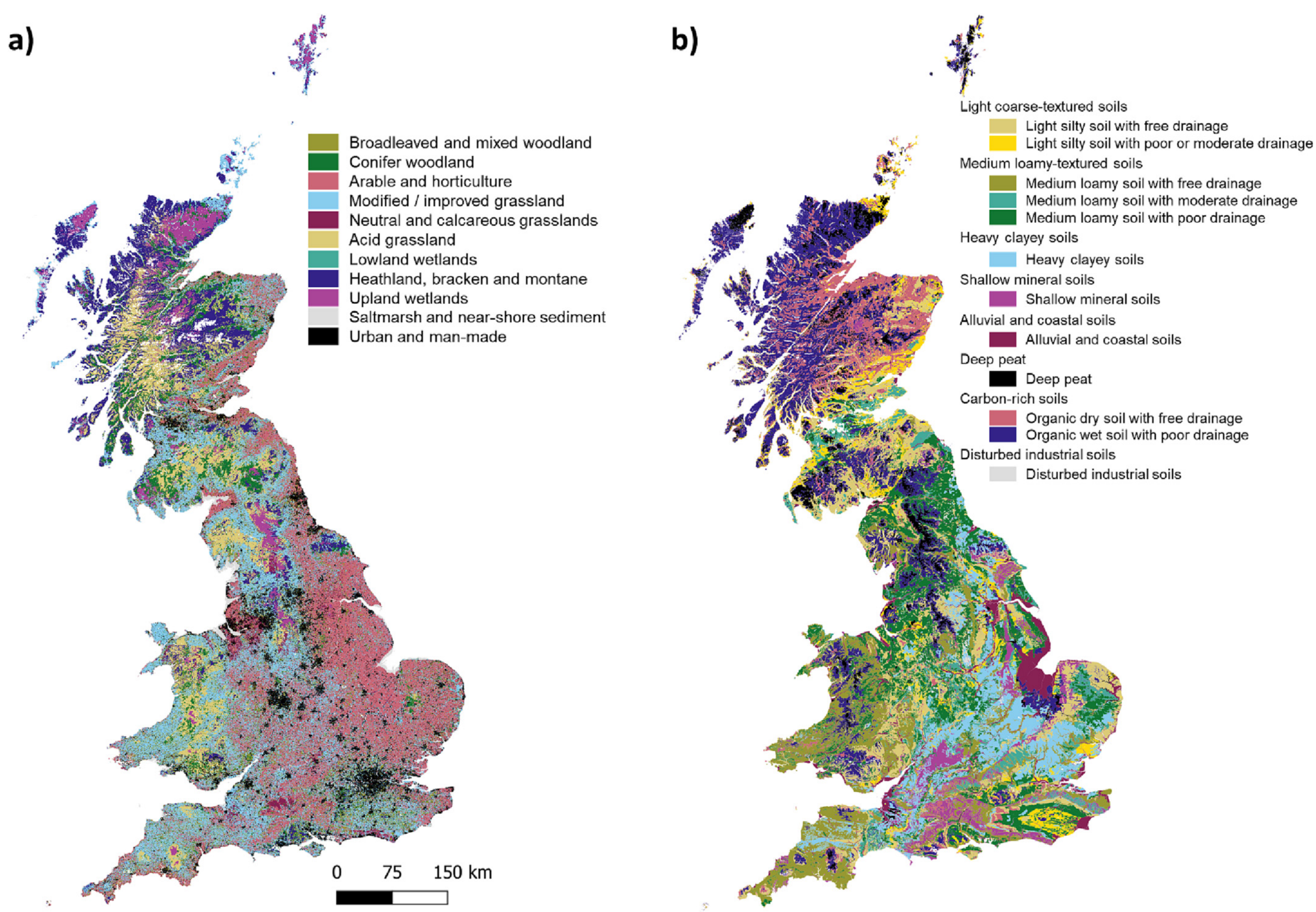


Fig. 1. a: Habitat groups of GB derived from the 2020 UKCEH Land Cover Map 10 m raster layer (Morton et al., 2021; see Table S1; Supplementary Materials for more details); b: Soil types of GB based on a bespoke classification scheme focussed on key structural attributes, including SOM content, soil texture, depth, flood risk, anthropogenic disturbance, and drainage (see Table S2 and Text S2; Supplementary Materials for more details). The legend consists of 8 high-order categories, with light, medium and carbon-rich soils split into subgroups defined by drainage to make 12 low-order classes. 1:250,000 scale vector map created from NATMAP (National Soil Resources Institute, 2001) and the National Soil Map of Scotland (Soil Survey of Scotland Staff, 1981).

mean annual rainfall (converted from kg m^{-2} s^{-1}) and mean daily temperature (converted from degrees Kelvin) for the whole period CHESS covers. Averaged climate data, along with elevation (m) from the NEXTMap 5 m resolution British Digital Terrain Model, were then extracted to each CS and earthworm sample point.

2.2. Stratification approach

To produce benchmarks with distinct distributions of SOM, pH, BD and EA, the CS and earthworm datasets would be progressively partitioned into finer physiotopes according to the dominant factors influencing the soil health indicators. However, we first needed to evaluate the suitability of the habitat classes and soil types for generating distinct benchmarks. Therefore, multivariate tests of differences in distributions of the soil health indicators between habitats and between soil types were calculated. The Kruskall-Wallis rank sum test was run, using the “kruskal.test” function in R's “stats” package (R Core Team, 2022) to test whether samples of the soil health indicator in question (SOM, pH, BD or EA) originated from the same habitat, low-order soil type or high-order soil type distribution. If the result of this test returned a statistically significant difference ($p < 0.05$), we next performed multiple pairwise comparisons between unique pairs of distributions using the “pairwise.wilcox.test” function in the “stats” package (R Core Team, 2022). The purpose of this second step was to determine how many unique pairs of habitats, low-order soil types and high-order soil types exhibited statistically significant differences (p < 0.05) from one another. If the pairwise comparisons generally resulted in significant differences, we could infer that our habitat and soil type categories were satisfactorily distinct. The Kruskall-Wallis rank sum test and pairwise Wilcox tests were run one at a time for each soil health indicator as the dependent variable, and these tests were chosen as the soil health indicator data do not conform to all of the assumptions required for the parametric one-way ANOVA and pairwise *t*-tests. Linear discriminant analysis (LDA) was implemented using the “lda” function of the “MASS” package (Venables and Ripley, 2002) to investigate further the discrimination among habitats and soil types for all CS samples with coincident measurements of SOM, pH, and BD. The degree of overlap between the different 95 % confidence intervals was used to visually judge the distinctiveness of topsoil properties predicted by habitats, and low-order and high-order soil types (following Simfukwe et al., 2010). The LDA results were used to judge whether to use the high- or low-order soil classes for benchmarking.

Following this, regression tree modelling was performed with the “rpart” function of the “rpart” package (Therneau et al., 2022) to determine the importance of environmental factors for predicting each soil health indicator. Regression tree modelling allows many potential predictor variables, including a mix of continuous and categorical variables, to be included in the modelling (Breiman et al., 1984). It also allows for the rank-ordering of importance of predictors with results visualised using the “ggplot2” package (Wickham et al., 2023). Additionally, SOM, pH, BD and EA were compared with co-located information on mean annual precipitation, mean daily temperature, and elevation using Pearson's correlation, implemented with the “cor” function of the “stats” package (R Core Team, 2022), to determine which of these 3 continuous predictor variables was most strongly associated with each of the 4 soil health indicators. The results of the regression tree modelling and correlation analysis were used to set the order in which distributions of soil health indicators would be stratified into smaller populations. Here, the number of stratification factors was capped at the point at which benchmarks would be sufficiently differentiated and based on large enough sample sizes (outlined in Section 2.3) to ensure benchmarks were statistically robust.

Our regression tree modelling and correlation (see Section 3.1) suggested that habitat was the most important control on the variability of soil health indicators, followed by soil type. Rainfall was the most important climate variable for pH, BD and for EA. Thus, after stratifying the soil property data by habitat and soil type, the data were sub-divided further into high rainfall and low rainfall regimes. The details on defining rainfall regime splits are available in the Supplementary Materials (Text S3).

2.3. Calculation of benchmarks

Prior to the calculation of benchmarks, we needed to know the minimum sample size required to ensure benchmarks would be statistically robust. To estimate this, we adapted a method outlined by Drexler et al. (2022), which demonstrated how the number of samples influenced SOC benchmarks. For each soil health indicator and subset of observations after stratification by habitat and soil type, bootstrap modelling was used to select a decreasing number of samples; adjusted benchmarks were calculated and then compared with the original benchmarks. For each interval of reducing sample size, the mean of 10,000 bootstrapped replicates of non-overlap between adjusted and original benchmarks was calculated. These results were plotted, and the minimum sample size of each stratum was identified as the inflection or “elbow point” of the curves using the “elbow” package in R (Casajus, 2020). However, because some combinations of habitat and soil type consist of very few data points, we stipulated that the minimum sample size should either be equal to 20 or the elbow point if the latter was larger. The averages of these results were summarised for each habitat and reported as the minimum required sample sizes.

For each soil health indicator of a defined physiotope, benchmarks were defined as the middle 80 % of values – i.e. between the 10th and 90th percentiles, similar to previous benchmarking approaches (e.g. Drexler et al., 2022; Griffiths et al., 2018). Although this choice is ultimately arbitrary, most of the observed values of soil health indicators would be described as “typical”, with the 10 % of values either side considered to be “above typical” and “below typical”. Medians were also calculated to show the midpoint of each benchmark distribution. All analyses were performed in R 4.2.2 (R Core Team, 2022).

3. Results

3.1. Relative importance of environmental factors on soil properties

Kruskall-Wallis rank sum tests showed statistically significant differences ($p < 0.05$) between habitats, and between both low-order and high-order soil types for all soil health indicators (Table S3; Supplementary Materials). Pairwise Wilcox tests of differences revealed statistically significant differences (p < 0.05) between most pairs of habitats for each soil health indicator. This demonstrates that our habitat classes are suitable for separating the distributions of soil health indicators (Table S4; Supplementary Materials). By contrast, among the low-order soil types, there were frequently no statistically significant differences between the medium soil subgroups, and between the light soil subgroups. Further, it was particularly difficult to distinguish the SOM distributions of the light subgroups from the medium loamy-textured soil subgroups and heavy clayey soils, with p values >0.05 (Table S5; Supplementary Materials). For the high-order soil types, BD and pH differed between light, medium and heavy soils (Table S6; Supplementary Materials). The LDA of habitats suggested a good level of discrimination in the 95 % confidence intervals around the group means of upland wetlands and the arable and horticulture habitats from one another, as indicated by a large degree of non-overlap between the confidence interval ellipses (Fig. S1; Supplementary Materials). Acid grassland appeared to strongly overlap with the heathland, bracken and montane class, and to a lesser extent, modified/improved grassland overlapped with neutral and calcareous grasslands (Fig. S1; Supplementary Materials). Between soil types, the greatest level of discrimination (non-overlap between ellipses) occurred between shallow mineral soils and the subgroups of carbon-rich soils (Figs. S2 and S3; Supplementary Materials). The other soil types by contrast overlapped strongly with one another, especially when light coarse-textured and medium loamy-textured soils were fragmented into their subgroups (Fig. S2; Supplementary Materials). Thus, for benchmarking, we used the high-order soil types to define physiotopes.

Through regression tree modelling, habitat was found to have the highest variable importance for all soil health indicators (Fig. 2). Soil type was the second most important for BD and SOM, and in the case of the

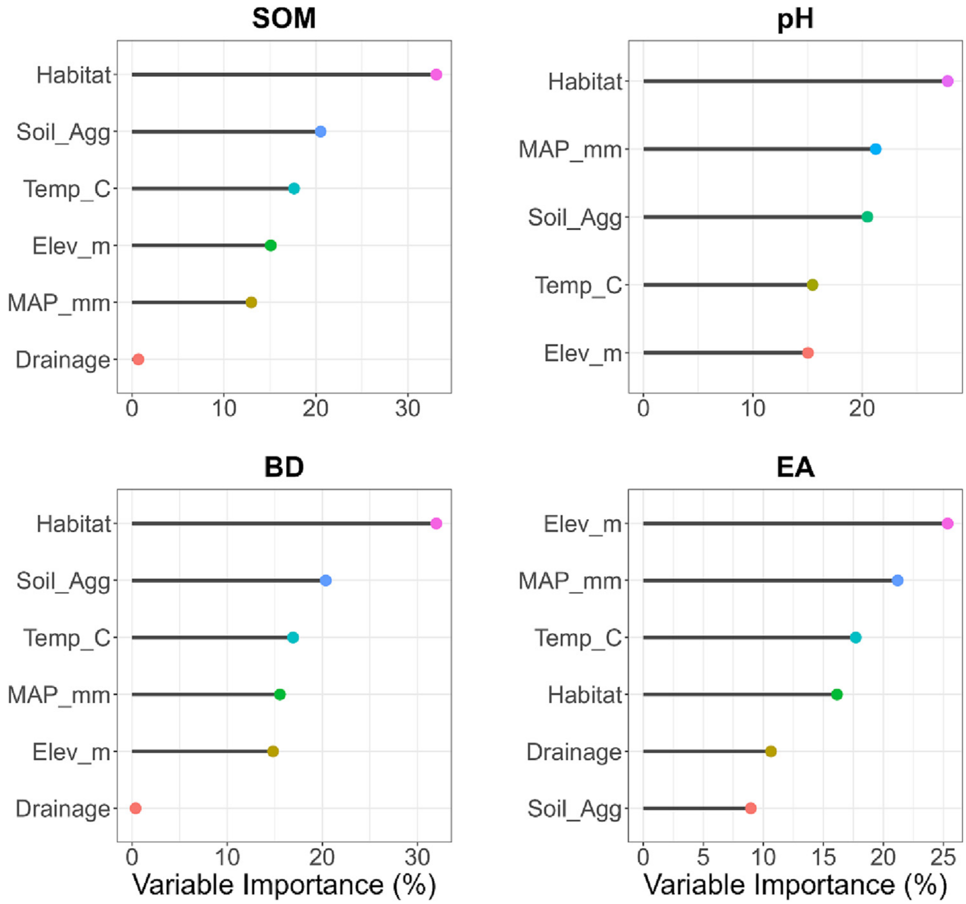


Fig. 2. Variable importance (%) calculated from regression modelling plotted for each soil health indicator. Note the unique order of habitat, high-order soil types (Soil_Agg), rainfall (MAP_mm), temperature (Temp_C), elevation (Elev_m), and drainage per plot.

latter, was twice as important a predictor as rainfall. Rainfall was the second most important variable for pH (after habitat) and EA (after elevation) but ranked third most important after habitat and soil type for BD, and behind temperature and elevation for SOM (Fig. 2). Rainfall, however, was found to have the strongest correlation with each soil health indicator apart from EA (Table 1). Thus, for each soil health indicator, we stratified the data by habitat, then high-order soil type and mean annual rainfall. While the data could have been stratified further, the lack of samples in most cases after dividing by 3 factors made this impractical.

3.2. Stratification and benchmarks of soil health indicators

The stratification of the soil property data into nine habitat groups led to significantly different ($p < 0.05$) distributions of soil health indicators in most cases (Fig. 3; Table S4). The differences in the mean SOM ranged from 1.95 % (neutral and calcareous grasslands cf. modified/improved grassland) to 68.23 % (upland wetlands cf. arable and horticulture). For pH, mean differences ranged from 0.06 (neutral and calcareous grasslands cf. modified/improved grassland) to 2.67 (upland wetlands cf. arable and horticulture), and for BD, from 0.03 g cm^{-3} (acid grassland cf. heathland, bracken and montane) to 1.07 g cm^{-3} (upland wetlands cf. arable and horticulture). EA data provided robust distributions for only 4 habitats (arable and horticulture, broadleaved and mixed woodland, modified/improved grassland, and neutral and calcareous grasslands). Here, differences in habitat-level means ranged from 3 (broadleaved and mixed woodland cf. neutral and calcareous grassland) to 14 (modified/improved grassland cf. neutral and calcareous grassland) earthworms per 20 × 20 cm spadeful.

Table 1
Pearson's R values for each correlation between a soil health indicator and one of the continuous environmental predictors (rainfall, temperature, and elevation). Note, SOM and rainfall were first log-transformed as these variables were found to be strongly right-skewed (skewness >2). Values are statistically significant ($p < 0.05$) and are reported to 2 decimal places.

	$\log_{10}$ rainfall	Temperature	Elevation
$\log_{10}$ SOM	0.64	−0.58	0.51
pH	−0.58	0.53	−0.48
BD	−0.69	0.60	−0.53
EA	−0.17	0.16	−0.21

Having determined the minimum sample sizes required for each soil health indicator (Table S7 and Figs. S4-S7; Supplementary Materials), we found that 6 high-order soil types were benchmarkable on 3 habitats (arable and horticulture, modified/improved grassland, and neutral and calcareous grasslands). Benchmarks for heavy clayey soils and shallow mineral soils exist for broadleaved and mixed woodland; alluvial and coastal soils could not be benchmarked at all for most habitats. By contrast, carbon-rich soils could be benchmarked across all habitats, including the relatively rare lowland wetlands habitat and the normally mineral soil dominated agricultural habitats (Fig. 4).

A total of 37 physiotopes could be split a step further into low and high rainfall regimes, including 13 splits for SOM, 13 for pH and 11 for BD (Fig. 4); there were no rainfall splits for EA (Fig. 5). For each of these soil health indicators, most of the rainfall splits were found in agricultural habitats (i.e. arable and horticulture and modified/improved grassland). This likely reflected that these 2 habitats cover roughly half of all land area across GB (Table S1; Supplementary Materials) and consequently, were represented by a much larger number of records within CS than other habitats.

The stratification by habitats, high-order soil types, and rainfall regimes resulted in 130 benchmarks of SOM, pH and BD (Fig. 4), and 5 benchmarks of EA (Fig. 5), yielding a total of 135 unique benchmarks. These

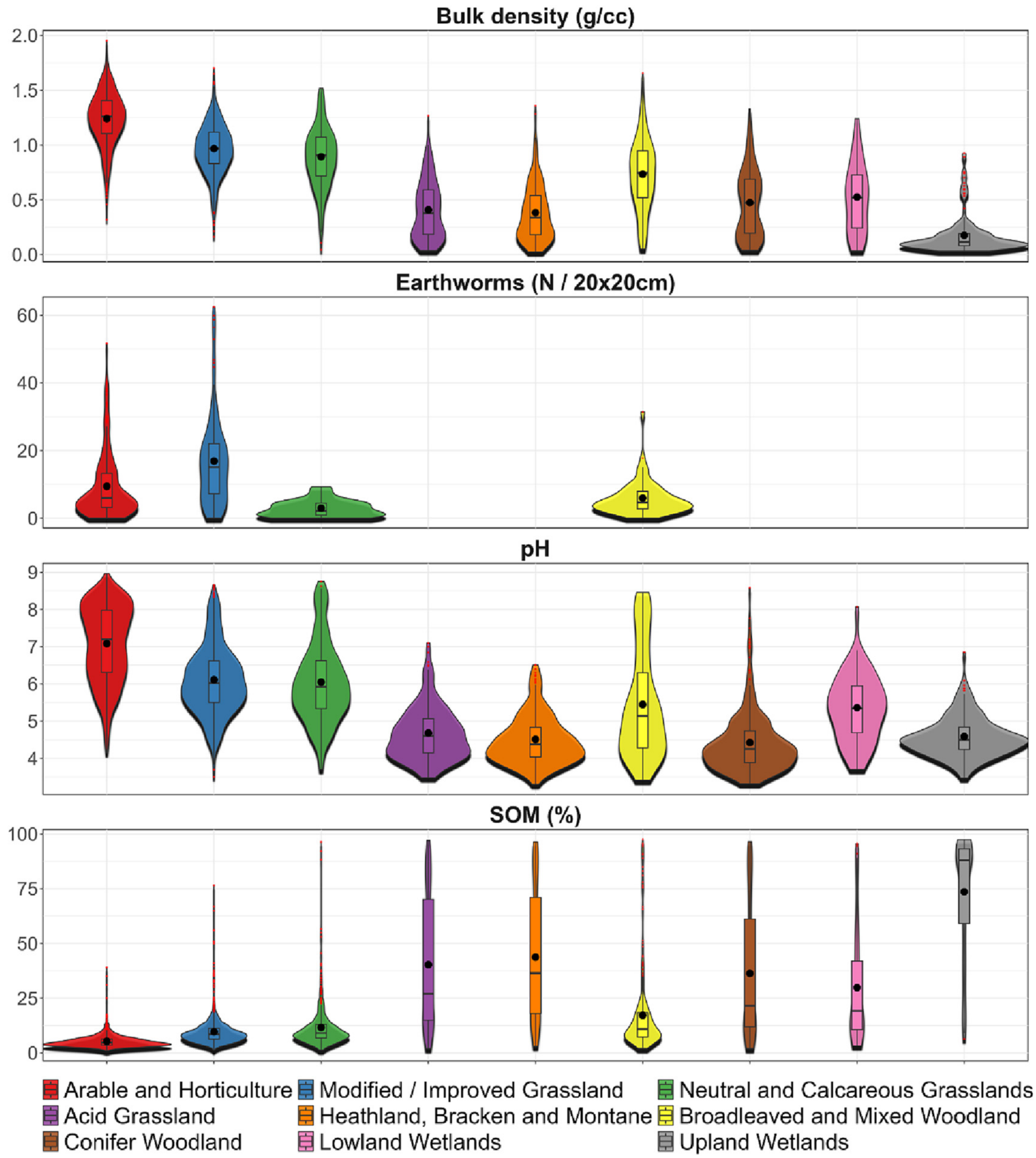


Fig. 3. Distributions of soil health indicators within each habitat. Units in panel headings. Note: EA has been recorded in all habitat classes; but there were too few data points to display distributions for five of these habitats after deep peats and disturbed industrial soils have been removed.

benchmarks show clear gradients in soil health indicators across habitats and among different soil types and climates. Disparities in SOM, pH and BD benchmarks on carbon-rich soils between the arable and horticulture (4.5–11.2 % SOM; pH 5.2–7.7; 0.85–1.35 g cm^{-3}) and upland wetlands (20–96 % SOM; pH 4–5.3; 0.06–0.3 g cm^{-3}) habitats can be seen clearly. Carbon-rich soils exhibit the highest SOM contents (medians ranging 7–89 %) and lowest BD values (medians ranging 0.11–1.13 g cm^{-3}) across all habitats. Shallow mineral soils tended to be the most alkaline (medians ranging pH 7.3–8.2), which likely reflects the dominance of calcareous soils over chalk and limestone (rendzinas) in this group. Soils under high-rainfall regimes showed higher SOM, and lower pH and BD than low-rainfall regimes. Medium loamy-textured soils appeared to be the most common soil type for which EA data were reported, and the greatest average (17) and range (3–36) of recorded counts occurred on modified/improved grassland with this soil type (Fig. 5). The lowest EA values occurred in broadleaved and mixed woodland habitats with heavy clayey soils (2–10; Fig. 5). For a full table of benchmark statistics, see Table S8 (Supplementary Materials).

Normalising benchmarks by medians revealed several patterns on how constrained the soil health benchmarks were. First, normalised pH ranges were the narrowest (0.28–0.55), followed by BD (0.44–2.18), SOM (0.85–3.42) and EA (1.75–3.8) when averaged by habitat (Fig. 6). Second, normalised widths generally increased along a gradient of decreasing land management intensity. Arable and horticulture and improved grassland exhibited among the narrowest benchmarks for SOM (0.95 and 1.22, respectively), pH (0.3 and 0.32, respectively) and BD (0.44 and 0.6, respectively). Paradoxically, upland wetlands exhibited the very narrowest normalised SOM and pH benchmarks (Fig. 6) and despite showing the narrowest normalised SOM ranges, upland wetlands had the widest BD benchmarks (Fig. 6). Woodland habitats exhibited among the widest pH and SOM benchmarks, reflecting the diversity of woodland types and management intensities. Arable and horticulture exhibited the widest EA benchmark (3.8), more than double that of modified/improved grassland (1.75), with broadleaved and mixed woodland in the middle (2.4).

Fig. 7 visualises the relative proportions of above and below typical SOM, pH and BD soils according to the most recent years of CS data

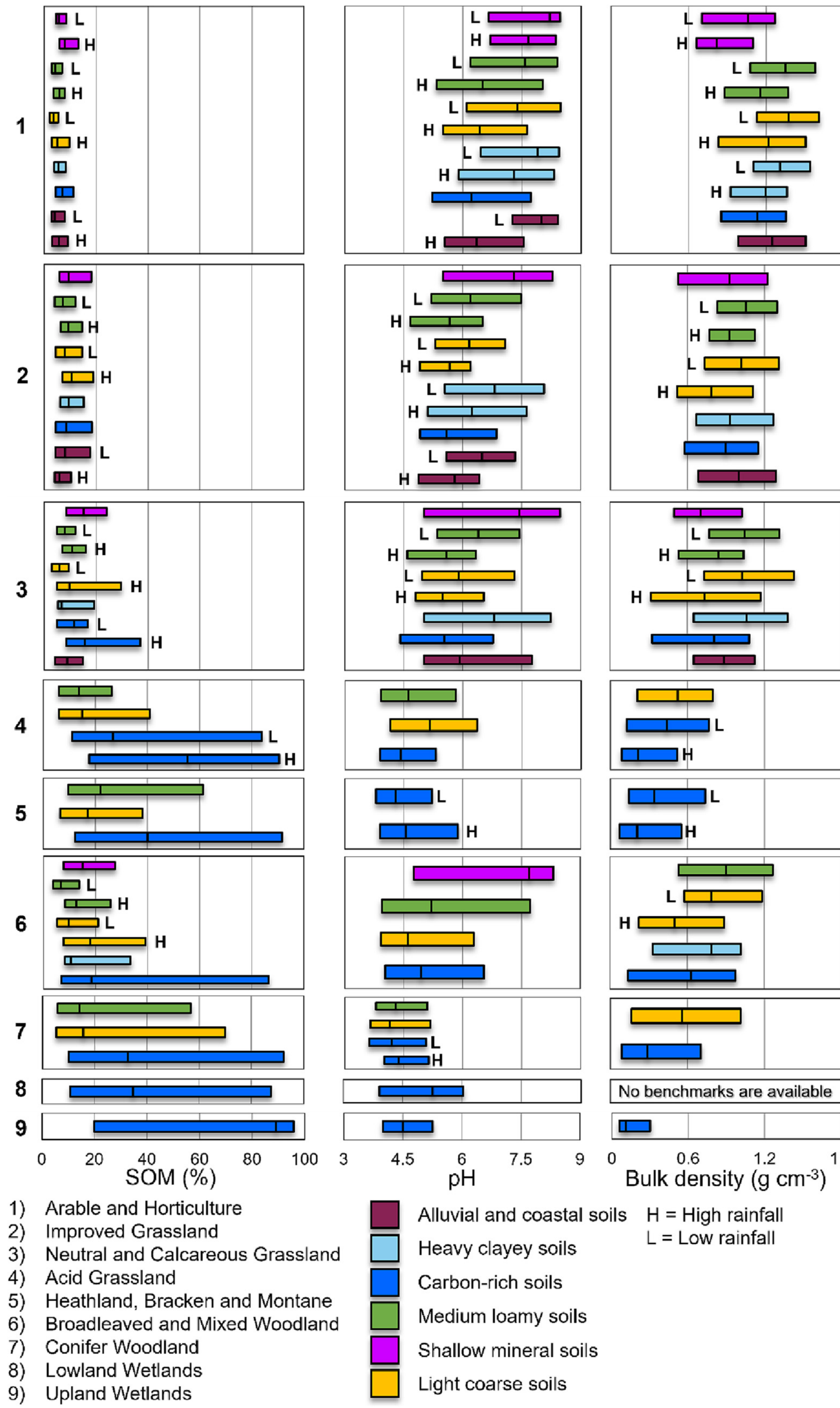


Fig. 4. SOM, pH and BD benchmarks for British soils under every major habitat. Horizontal bars denote the benchmarks (ranges between the 10th and 90th percentiles) and are colour-coded by high-order soil type. Numbered rows of panels correspond to different habitats. The letters, "H" and "L" mark where there is an identified split by rainfall and stand for "high rainfall" and "low rainfall", respectively.

Fig. 5. EA benchmarks for English soils under every habitat with sufficient data available. Each bar is colour-coded by soil type. Note: Much fewer benchmarks were calculable for earthworms compared to other soil health indicators and no rainfall splits were identified. Therefore, all earthworm benchmarks are plotted together on a single panel, covering only 3/9 habitat types and 3/6 soil types.

(2007 and the ongoing 2019–23 surveys) by Nomenclature of Terrestrial Units for Statistics level 1 (NUTS 1) region. As EA data are not derived from a structured survey, it is inappropriate to map spatial trends for this indicator. East Anglia has the lowest level of sites with above typical SOM (2.4 %; Fig. 7a), and below typical pH (2.3 %; Fig. 7e) and BD (5.8 %; Fig. 7f); meanwhile, the same region has the highest proportion of sites with below typical SOM (19.2 %; Fig. 7d), and above typical pH (39.1 %; Fig. 7b) and BD (17.4 %; Fig. 7c). Thus, our results suggest that soils in East Anglia are more frequently deficient in SOM, alkaline and more compacted compared to the rest of the country.

4. Discussion

4.1. Contextualising benchmarks considering known spatio-temporal trends

In our effort to benchmark soil health, we selected properties that are representative of physical soil structure, nutrition, and biological support and function. To that end, we opted for SOM, pH, BD and EA, which all have the additional advantage of having been extensively studied in the past, allowing for contextualisation of our benchmarks against previously reported trends in the literature (e.g. Emmett et al., 2010; Reynolds et al., 2013; Seaton et al., 2021, 2023). This contextualisation is important because differences between benchmarks are driven by additional factors to those that we have incorporated in our stratification approach.

Our analysis revealed that generally, habitat was the single most important control on soil properties, and this was followed by soil type (reflecting soil texture, organic matter content, depth and inundation risk) and then mean annual rainfall. This matches earlier SOC benchmarking of agricultural topsoils in England and Wales (Verheijen et al., 2005) and Germany (Drexler et al., 2022; Vos et al., 2019). It is possible that the inclusion of a much wider range of habitats beyond agriculture has accentuated the role habitat plays in governing the soil health indicator benchmarks. For instance, SOC tends to increase along a decreasing management intensity gradient from agricultural to semi-natural habitats (Guo and Gifford, 2002; Seaton et al., 2021), with wetlands possessing the greatest fraction (20–30 %) of global SOC stocks (Lal, 2008).

The soil health indicators selected from CS (SOM, pH and BD) are highly correlated with each other (Fig. S8; Supplementary Materials), which could imply that a change in one soil health indicator will impact on the others. In particular, SOM and BD show the distinctive negative curvilinear relationship reported previously for CS (Emmett et al., 2010; Reynolds et al., 2013) and a similar stratified random survey for Wales (Seaton et al., 2021), and the Spearman's correlation coefficient (−0.92) is the most negative of all indicator relationships (Fig. S8; Supplementary Materials). This might support the notion that increasing SOM can induce a reduction in BD (increased porosity); but, concomitant changes in SOM and BD over time have not been identified in CS before (Emmett et al., 2010). A related complication to this is that despite showing the narrowest SOM benchmarks, wetlands had the widest BD benchmarks (Fig. 6). This possibly reflects the variety of vegetation conditions in this habitat, with BD increasing on the order of sphagnum-dominated < woody-dominated < sedge-dominated soils (Liu and Lennartz, 2019), and the wide BD variability will be important to consider when determining wetland soil organic carbon stocks.

Soils with the highest SOM content correspond with the lowest pH levels and a distinct upper limit (see Reynolds et al., 2013) is visible (Fig. S8; Supplementary Materials). Upland wetlands typify one end of this relationship (low pH and high SOM) and the observation of the narrowest benchmark ranges for both of these soil health indicators in this habitat mirrors the high signal to noise ratios calculated for SOM and pH from CS data previously (Emmett et al., 2010). Whether this upper limit represents a threshold in the chemical behaviour of SOM remains unclear, but it has been suggested for instance, that liming carbon-rich soils to raise their pH on arable land might shift the soil system from an anaerobic state towards higher productivity and SOM decomposition rates under aerobic conditions (Alison et al., 2019). There are relatively broad pH benchmarks for neutral and calcareous grasslands and broadleaved and mixed woodland (Fig. 6). Here, pH benchmarks likely reflect the gradual recovery of these habitats from acidification since 1978, driven largely by reductions in dry sulphur deposition (RoTAP, 2012). Additionally, the pH benchmarks for shallow mineral soils under arable and horticulture land use may be influenced by a shift over time towards deeper cultivation that brings relatively unweathered chalk parent material to the soil surface (Reynolds et al., 2013). The narrower, more acidic pH benchmarks in coniferous woodland and wetland habitats may be driven by attenuation responses to decreasing rates of dry sulphur deposition within these ecosystems (Kirk et al., 2010; Reynolds et al., 2013). Although pH has generally

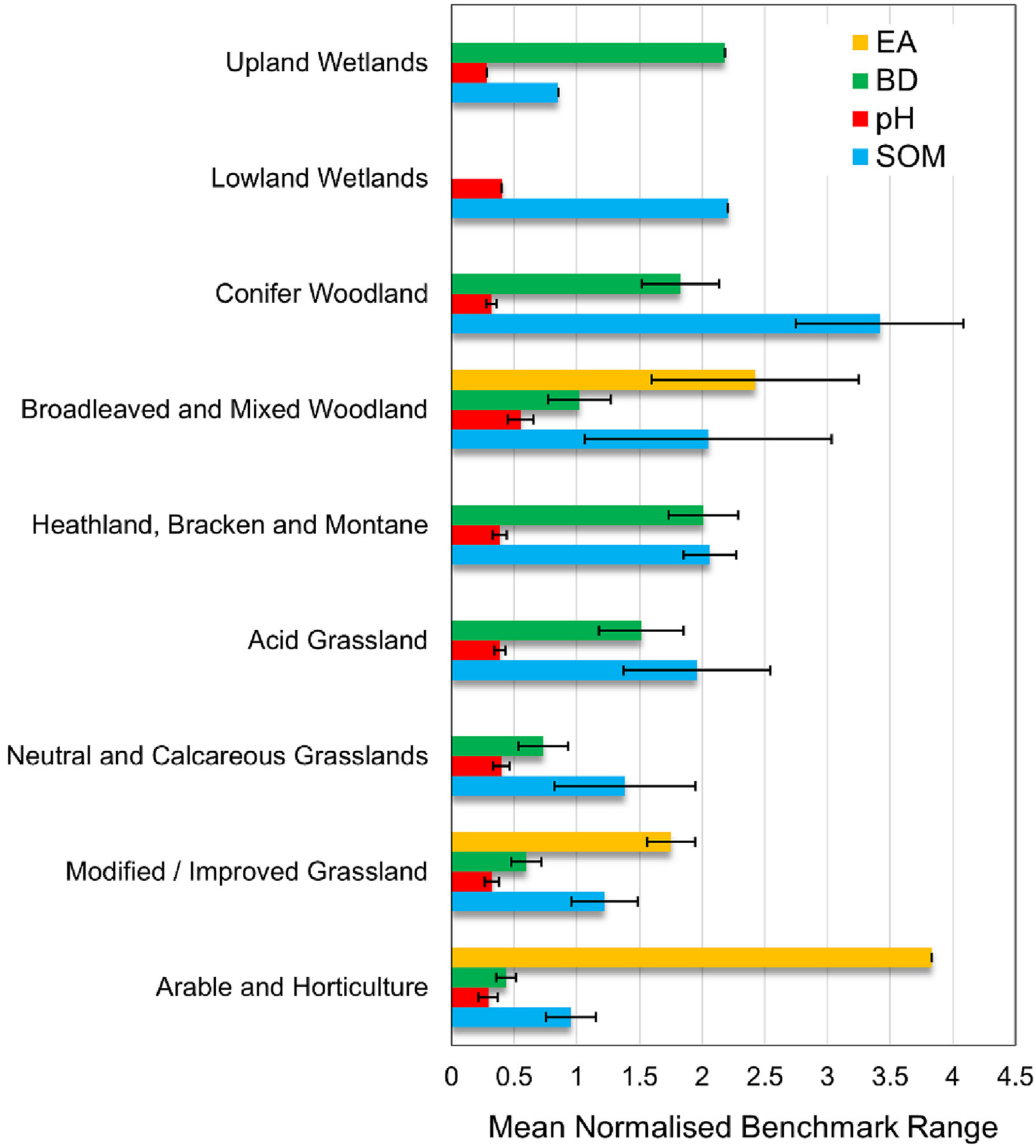


Fig. 6. Normalised benchmark ranges averaged for each soil health indicator and habitat class based on the data presented in Table S8 (Supplementary Materials). Benchmark ranges are normalised by dividing by the medians. Error bars represent ±1 standard deviation of the mean.

increased over time over 1978–2007, recent analysis of CS data suggests some soils have undergone (re-)acidification since 2007, with changing fertiliser use and climate change effects among the speculated drivers of this trend (Seaton et al., 2023).

The topsoil pH-BD plot (Fig. S8; Supplementary Materials) appears to confirm that generally the more acidic soils tend to have lower BD (Spearman's $r = 0.68$). This most likely reflects that soils under agricultural land use tend to be the most compacted and managed for pH, and the least compacted carbon-rich soils occur exclusively in acidic pH ranges. Paradoxically, shallow mineral soils are some of the most alkaline soils yet show the lowest BD in agricultural and grassland habitats (Fig. 4). This may occur because rendzinas drain rapidly and can be worked without compaction soon after rainfall; by contrast heavy clayey and medium loamy-textured soils are the most susceptible to compaction (Batey, 2009). Clay-poor coarse-textured soils are also highly susceptible to compaction despite being workable across a breadth of moisture conditions (Batey, 2009). This is especially true for the more unstructured light soils as the soil aggregates typically form a compact layer that inhibits water and air flow (Morris et al., 2010).

Relationships between EA and SOM, pH and BD have not been plotted here. This is because, unlike CS soils data, the physicochemical data associated with the earthworm records are not always exactly co-located in time and space; SOM, BD and pH are often measured from separate cores rather than the same individual sample, and laboratory methods most likely vary. Earthworms tend to favour drier clayey and loamy soils over sandy soils (van Groenigen et al., 2014), which might explain why most of our benchmarks are for medium loamy soils (though this could also reflect that this is the most common soil type). Our benchmarks show higher EA for modified/improved grassland than arable and horticulture (Fig. 5). This follows analysis that suggests EA increases when land-use shifts to less intense forms (Spurgeon et al., 2013). A global meta-analysis also revealed that EA increases with reduced land-use intensity were observable across light (<18 % clay), medium (18–35 % clay) and most prominently, heavy (>35 %) soils (Briones and Schmidt, 2017). Similar patterns were visible to some extent in our results, though it was not possible here to benchmark EA for arable land with heavy clayey soils (Fig. 5). EA benchmarks were the widest compared to the other soil health indicators (Fig. 6). This perhaps reflects a large proportion of unexplained variance and may support the

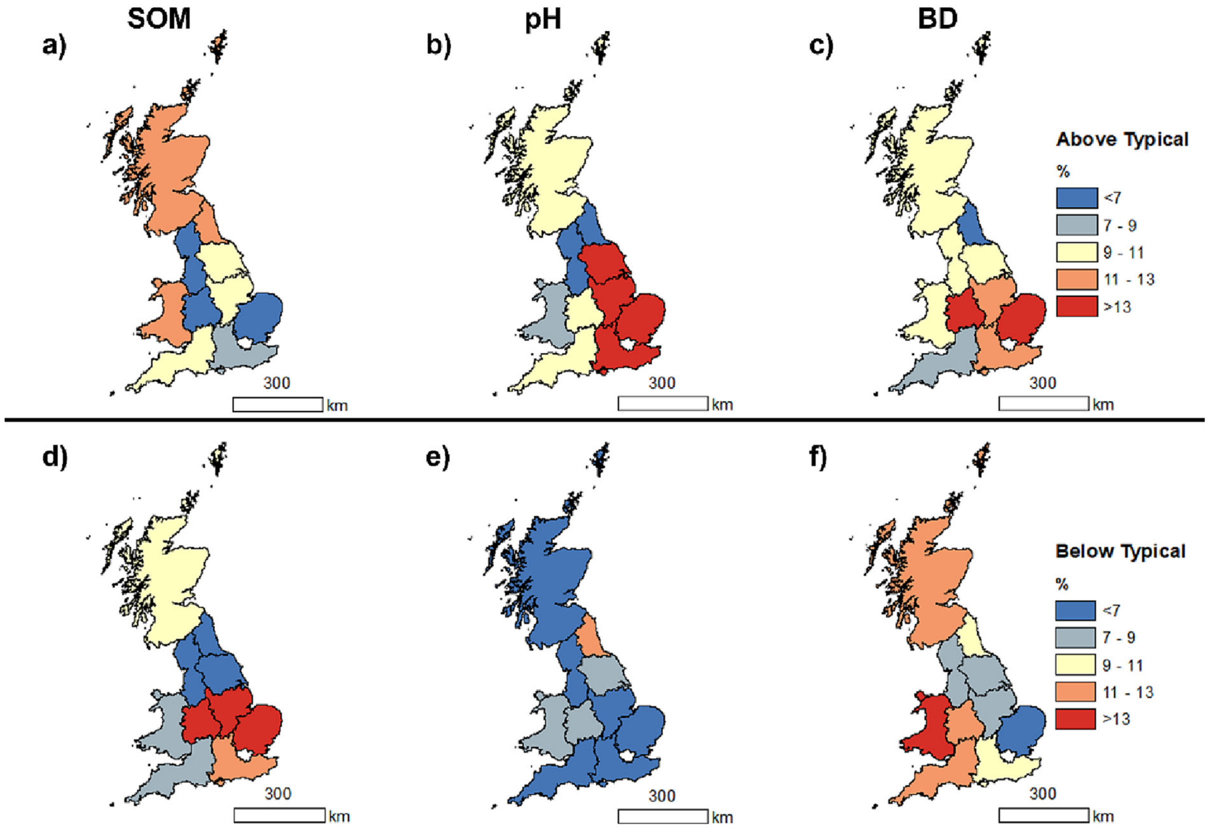


Fig. 7. Breakdown by NUTS 1 regions of above typical SOM (a), pH (b) and BD (c); and below typical SOM (d), pH (e) and BD (f). The proportion of CS datapoints from the most recent two survey years (2007 and the ongoing 2019–23 surveys) are calculated as percentages within each NUTS 1 region. London is excluded due to limited data.

notion that other factors such as soil nitrate levels and moisture content likely drive national-scale EA patterns to a more significant degree than our stratification factors (as argued by Hodson et al., 2021).

Hitherto, the discussion has focussed on explaining differences in the soil health benchmarks between physiotopes. However, consideration also needs to be given to where the atypical soils are located, with "atypical" referring to soils with soil health indicator values in either the bottom 10 % or top 10 % of values outside of the benchmark ranges. If there are regions with a disproportionate presence of atypical soils compared to the expected nationwide average (10 % each for below and above typical), these regions may need to be prioritised over others for targeted land management improvements.

Spatial analysis of where soils with atypical SOM, pH and BD occur (Fig. 7) reveal East Anglia is particularly rich in soils with below typical SOM and above typical pH and BD compared to the rest of the country. These patterns are consistent with the findings of others; for example, East Anglia is known to possess some of the lowest SOC stocks in Europe (Lugato et al., 2014), and is one of the most vulnerable regions in the continent to wind erosion (Borrelli et al., 2017) and soil losses due to crop harvesting (Panagos et al., 2019). Crucially, regional inequalities in our soil health data have emerged even after land use, soil type and rainfall have been accounted for. In other words, East Anglian soils do not simply stand out because they are dominated by arable land cover and drained organic soils; rather, they stand out because relative to the same soil types and land uses across GB, SOM is notably low, and pH and BD are notably high. This does not imply necessarily that existing land management practices are worse in East Anglia than other regions, as other unexplained environmental factors may be responsible. Ultimately, it raises the questions of, "to what extent are these particular soils vulnerable to degradation?", and "should existing land management practices be modified in this part of the country to improve soil health?".

4.2. Limitations and a new soil health webtool

Our establishment of national-scale benchmarks for soils presents new opportunities for land management stakeholders to assess their soil health. However, it is important to highlight the limitations of our approach.

The benchmarks for SOM, pH and BD were calculated from every soil monitoring year on record in the CS database. This results in benchmarks that may be wider than if data from one year were used, because soil properties may have changed significantly over time. However, the key advantage of using all years is that the maximum number of combinations of physiotopes can be considered. Benchmarking using all years means that if additional data from future surveys is incorporated, the benchmarks will evolve. Balancing the limitations of using a single survey year with those stemming from the use of multiple surveys will require additional standardisation of benchmarking in the future. These same issues apply to the EA data, which, additionally, were not recorded as part of a structured national survey but were derived from multiple datasets employing differing methodologies. Consequently, EA benchmarks are less likely to be as robust as the other soil health indicators.

We favoured EA as the primary biological indicator because, although other biological data are collected as part of CS such as mesofauna (see Black et al., 2008), their interpretation might require expertise that non-specialists lack (Bünemann et al., 2018). The EA dataset will continue to be expanded as collaboration with a wider range of data holders increases, meaning a greater number of landscapes should be represented in future than those in Fig. 5. Additional quantitative properties like total nitrogen and heavy metal concentrations could also be considered for future soil health benchmarking. Qualitative metrics such as visual soil assessments of physical structure are also advantageous, especially given their straightforward interpretation, potential to consider the subsoil, and recognised importance in yield gap analysis and land management programs (McKenzie et al., 2015). However, as visual soil assessments are not generally recorded as part of national-scale datasets (but see Newell-Price et al., 2013 for a survey of grasslands in England and Wales), we were unable to produce applicable benchmarks here.

Despite efforts to assess more landscapes, urban and built-up areas, coastal ecosystems, heavily modified industrial soils and deep peats could not be included. In all cases, apart from deep peat, this was because we did not have enough data points to provide reliable benchmark statistics. Deep peat soils meanwhile have highly complex issues that require a different approach for soil health assessment. Additionally, our original intention was to calculate benchmarks using the low-order soil types, with greater consideration of differing drainage characteristics. However, because it was not possible to distinguish distributions of SOC, pH and BD among low-order soil types, we ultimately used the high-order soil types instead. Had we considered soil health at greater depths, differences in drainage characteristics may have become significant. Our categorisation of agricultural land use might be considered too broad for defining soil health benchmarks. A multitude of land management considerations including fertiliser use, tillage practices, cover crop use, and machinery traffic will affect all our soil health indicators. However, such detailed management information was not recorded by the CS field surveyors and in any case, would reflect a brief, likely short-lived snapshot.

As the top 15 cm of soil were sampled for CS, our benchmarks reflect topsoil conditions only. This has the advantage of fulfilling many of the considerations for selecting soil health indicators, particularly criteria such as practicality and sensitivity to land use and management. However, assessing soil health for functions such as favourable soil structure for exploitable root depths or water cycling, necessarily requires information from greater depths into the subsoil (The Royal Society, 2020). Regrettably, contemporary soil health measurements at depths much >0–15 cm are lacking at GB-scale. Sampling larger depths of soil for national monitoring programs like CS is an important future ambition. Incorporating subsoil benchmarks might, however, require modifications to our land stratification approach, given that the best predictors of a soil property in the topsoil may not apply for the subsoil (e.g. Vos et al., 2019).

The benchmarks presented here should not be interpreted as optimum values for each indicator, because the very definition of "optimum" is context specific and subjective. One needs to consider which soil functions and ecosystem services ought to be prioritised to optimise soil health for their land use and management practices (Andrews et al., 2004). This will also be a challenge for policy makers to steer soil health improvements towards achieving the UN sustainability goals (Lehmann et al., 2020). However, by establishing the first GB-wide national soil health benchmarks, our work should offer an intermediate step for policy makers to set targets for soil health.

As an additional step towards soil health targets, some have proposed aggregating information from multiple soil health benchmarks into an overall score or index (e.g. Andrews et al., 2004; Griffiths et al., 2018). However, while an index might conveniently express "good" or "bad" soil health, it may not be very meaningful to some users, especially if the presentation of results is too opaque for land managers to determine what can be done to maintain or enhance soil health. Furthermore, soil health is best assessed in relation to specific soil functions or ecosystem services that may rank differently in importance; thus, some ability to select specific indicators and apply weighting to each will be necessary (Bünemann et al., 2018).

To facilitate land manager engagement, we have incorporated our benchmark statistics in an app developed in R Shiny (Chang et al., 2021): https://connect-apps.ceh.ac.uk/soilhealth/ (Fig. S9; Supplementary Materials). In the app, users are asked to select which habitat, soil type, and mean annual rainfall rate characterise their location. Colour-coded distribution plots for the selected landscape characteristics are then displayed, with the option to enter any soil health indicator values the user might happen to have. In this way, users can assess the typical ranges of soil health indicator values for lands with common characteristics to theirs and use these to directly compare any soil health measurements of their own. If the user finds that their soil's health indicator measurements fall outside of the typical range benchmarked for their landscape type, it may indicate that they need to modify their land management practices to improve soil health. Given the sheer multitude of landscapes we have benchmarked, it is not practical to outline potential strategies to modify land management to benefit soil health here; that will require a separate decision support tool.

5. Conclusion

Soil health benchmarks have been established for semi-natural landscapes, in addition to more intensively managed agricultural environments across GB. This represents one of the few studies of its kind to consider the wider semi-natural environment. Here, we presented state of the art national benchmarks of topsoil (0–15 cm) health indicators (SOM, pH, BD and EA) for GB. These benchmarks represent physical, chemical and biological aspects of soil health, and are conceptual, practical, sensitive and interpretable (sensu Bünemann et al., 2018).

Our analysis highlights that soil health indicators vary markedly by habitat, soil type and rainfall. BD and pH tend to decrease in proportion with management intensity (agriculture > semi-natural grasslands > woodlands > heathlands > wetlands), with the reverse being true for SOM and to a limited extent, EA. Normalising benchmarks by median values revealed that pH benchmarks were the most constrained, followed by BD and SOM, with EA the least constrained when averaged by habitat. Regional comparisons revealed that East Anglia currently possesses the most disproportionate numbers of CS sites with below or above typical soil health indicator values across GB. The fact that this is the case even after land use, soil type and rainfall have been considered underscores how urgently land management may need addressing to promote soil health in this region.

Future efforts in our soil health assessment might include expanding our benchmarking to feature additional soil properties and developing benchmarks for subsoils. Our benchmarking approach could be applied to other countries and at multinational scales. We have created a webtool to communicate our results with the public, which will also allow land managers to assess their soil health against our national benchmarks and judge whether they need to modify their land management practices to benefit soil health.

CRediT authorship contribution statement

Christopher J. Feeney: Conceptualisation, Methodology, Investigation, Formal Analysis, Data Curation, Writing – Original draft preparation, Writing - Review & Editing. **David A. Robinson:** Conceptualisation, Methodology, Supervision, Writing - Review & Editing. **Aidan M. Keith:** Data Curation (Earthworm Data), Formal Analysis, Writing - Review & Editing, **Audric Vigier:** Formal Analysis (Webtool Creation), Writing - Review & Editing. **Laura Bentley:** Data Curation, Writing - Review & Editing. **Richard P. Smith:** Methodology (design of soil classification scheme), Writing – Review & Editing. **Angus Garbutt:** Data Curation (Countryside Survey), Writing - Review & Editing. **Lindsay C. Maskell:** Data Curation (Countryside Survey), Writing - Review & Editing. **Lisa Norton:** Data Curation (Countryside Survey), Writing - Review & Editing. **Claire M. Wood:** Data Curation (Countryside Survey), Writing - Review & Editing. **B. Jack Cosby:** Project Administration, Writing – Review & Editing. **Bridget A. Emmett:** Project Administration, Conceptualisation, Methodology, Supervision, Writing - Review & Editing.

Data availability

CS data for all but 2019-23, as well as CHESS data are available on the EIDC site. CS location data were needed for much of the analysis & (along with R code) cannot be shared due to confidentiality.

Declaration of competing interest

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.

Acknowledgements

This work was supported by the Natural Environment Research Council (NERC) award number NE/R016429/1 as part of the UK-SCAPE programme delivering National Capability. The Countryside Survey is supported by NERC (as part of UK-SCAPE) and was previously funded by a partnership of government funded bodies led by NERC and the Department for Environment, Food and Rural Affairs (Defra), which includes the UK Centre for Ecology and Hydrology, Countryside Council for Wales, Forestry Commission, Natural England, the Northern Ireland Environment Agency, the Scottish Government, Scottish Natural Heritage, and the Welsh Assembly Government. We would like to thank Ellen Fay for contributing her thoughts to this study in previous discussions. Some of the earthworm records stem from a 2019 Rothamsted survey and were supplied by Jacqueline Stroud. Land Cover Map, CHESS, NATMAP and The National Soil Map of Scotland datasets were also used in the creation of our benchmarks, and we gratefully acknowledge the UKCEH staff who developed Land Cover Map and CHESS, and Cranfield University and The James Hutton Institute for providing their soil maps. Elevation data are licensed to UKCEH in perpetuity (NEXTMap Britain digital terrain model by Intermap. © 2009 Intermap Technologies Inc. All rights reserved). We would like to thank Allan Lily for providing data tables that link Scottish soil series names to their England and Wales equivalents, which facilitated the mapping of our soil classes across all of GB. We would also like to thank Sophie Drexler and Axel Don for clarifying to us the methods in their 2022 soil carbon benchmarking study, and to David Cooper and Peter Henrys for assistance with statistical analysis. We would also like to thank Cristina Martin Hernandez for assistance with developing the new webtool in R Shiny, which allows land managers to measure their soil health and compare results against our benchmarks: https://connect-apps.ceh.ac.uk/soilhealth/ Chris Lowe is also gratefully acknowledged for his assistance with editing the abstract of this manuscript. Finally, we would like to thank the handling editor (Paulo Pereira) and two anonymous reviewers for comments that greatly strengthened the quality of this manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2023.163973.

References

Alison, J., Thomas, A., Evans, C.D., Keith, A.M., Robinson, D.A., Thomson, A., Dickie, I., Griffiths, R.I., Williams, J., Newell-Price, J.P., Williams, A.G., Williams, A.P., Martineau, A.H., Gunn, I.D.M., Emmett, B.A., 2019. Technical annex 3: soil carbon management. Environment and Rural Affairs Monitoring and Modelling Programme (ERAMMP): Sustainable Farming Scheme Evidence Review. Report to Welsh Government (Contract C210/2016/2017). Centre for Ecology and Hydrology Project . https://erammp.wales/en/r-sfs-evidence-pack.

Amundson, R., Berhe, A.A., Hopmans, J.W., Olson, C., Sztein, A.E., Sparks, D.L., 2015. Soil and human security in the 21st century. Science 348 (6235), 1261071. https://doi.org/10.1126/science.1261071.

Andrews, S.S., Karlen, D.L., Cambardella, C.A., 2004. The soil management assessment framework. Soil Sci. Soc. Am. J. 68 (6), 1945–1962. https://doi.org/10.2136/sssaj2004.1945.

Avery, B.W., 1973. Soil classification in the soil survey of England and Wales. Eur. J. Soil Sci. 24 (3), 324–338.

Avery, B.W., 1980. Soil Classification for England and Wales: Higher Categories. Rothamsted Experimental Station.

Batey, T., 2009. Soil compaction and soil management - a review. Soil Use Manag. 25 (4), 335–345. https://doi.org/10.1111/j.1475-2743.2009.00236.x.

Black, H., Bellamy, P., Creamer, R., Elston, D., Emmett, B., Frogbrook, Z., Hudson, G., Jordan, C., Lark, M., Lilly, A., Marchant, B., Plum, S., Potts, J., Reynolds, B., Thompson, R., Booth, P., 2008. Design and Operation of a UK Soil Monitoring Network Science Report – SC060073.

Borrelli, P., Lugato, E., Montanarella, L., Panagos, P., 2017. A new assessment of soil loss due to wind erosion in European agricultural soils using a quantitative spatially distributed modelling approach. Land Degrad. Dev. 28 (1), 335–344. https://doi.org/10.1002/ldr.2588.

Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. Classification and Regression Trees. CRC Press.

Briones, M.J.I., Schmidt, O., 2017. Conventional tillage decreases the abundance and biomass of earthworms and alters their community structure in a global meta-analysis. Glob. Chang. Biol. 23 (10), 4396–4419. https://doi.org/10.1111/gcb.13744.

Bunce, R.G.H., Barr, C.J., Clarke, R.T., Howard, D.C., Scott, W.A., 2007. ITE Land Classification of Great Britain 2007. NERC Environmental Information Data Centre. https://doi.org/10.5285/5f0605e4-aa2a-48ab-b47c-bf5510823e8f.

Bünemann, E.K., Bongiorno, G., Bai, Z., Creamer, R.E., De Deyn, G., de Goede, R., Fleskens, L., Geissen, V., Kuyper, T.W., Mäder, P., Pulleman, M., Sukkel, W., van Groenigen, J.W., Brussaard, L., 2018. Soil quality – a critical review. Soil Biol. Biochem. 120, 105–125. https://doi.org/10.1016/j.soilbio.2018.01.030.

Carey, P.D., Wallis, S., Chamberlain, P.M., Cooper, A., Emmett, B.A., Maskell, L.C., McCann, T., Murphy, J., Norton, L.R., Reynolds, B., Scott, W.A., Simpson, I.C., Smart, S.M., Ullyett, J.M., 2008. Countryside Survey: UK Results From 2007.

Casajus, N., 2020. The elbow package (0.0.0.9000). GitHub. https://nicolascasajus.fr/elbow/index.html.

Chang, W., Cheng, J., Allaire, J.J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., Borges, B., 2021. shiny: web application framework for R (1.7.1). CRAN. https://cran.r-project.org/package=shiny.

Defra, 2018. 25 Year Environment Plan.

Directorate-General for Environment, 2022. Commission consults citizens and stakeholders on possible EU Soil Health Law. European commision. https://environment.ec.europa.eu/news/commission-consults-citizens-and-stakeholders-possible-eu-soil-health-law-2022-08-02_en.

Doran, J.W., Zeiss, M.R., 2000. Parent material and soil physical properties. Appl. Soil Ecol. 15 (1), 3–11. https://doi.org/10.1016/S0929-1393(00)00067-6.

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.

Egan, F., Bay, N.S., 2021. Soil health benchmarks. https://pasafarming.org/resources/soil-health-benchmarks-2021-report/.

Emmett, B.A., Reynolds, B., Chamberlain, P.M., Rowe, E., Spurgeon, D., Brittain, S.A., Frogbrook, Z., Hughes, S., Lawlor, A.J., Poskitt, J., Potter, E., Robinson, D.A., Scott, A., Wood, C., Woods, C., 2010. Countryside survey technical report no. 9/07 soils report from 2007. http://nora.nerc.ac.uk/9354/.

FAO, ITPS, 2015. Status of the World's Soil Resources (SWSR) - Main Report.

Griffiths, B., Hargreaves, P., Bhogal, A., Stockdale, E., 2018. Soil biology and soil health partnership project 2: selecting methods to measure soil health and soil biology and the development of a soil health scorecard (Issue Final Report No. 91140002-02). https://ahdb.org.uk/soil-biology-and-soil-health-partnership.

Guo, L.B., Gifford, R.M., 2002. Soil carbon stocks and land use change: a meta analysis. Glob. Chang. Biol. 8 (4), 345–360. https://doi.org/10.1046/J.1354-1013.2002.00486.X.

Harris, J.A., Evans, D.L., Mooney, S.J., 2022. A new theory for soil health. Eur. J. Soil Sci. 73 (4), 1–7. https://doi.org/10.1111/ejss.13292.

Haygarth, P.M., Ritz, K., 2009. The future of soils and land use in the UK: soil systems for the provision of land-based ecosystem services. Land Use Policy 26 (Suppl. 1), 187–197. https://doi.org/10.1016/j.landusepol.2009.09.016.

Hodson, M.E., Corstanjeb, R., Jones, D.T., Witton, J., Burton, V.J., Sloan, T., Eggleton, P., 2021. Earthworm distributions are not driven by measurable soil properties. Do they really indicate soil quality? PLoS One 16 (8), e0241945. https://doi.org/10.1371/journal.pone.0241945.

Jackson, D.L., 2000. Guidance on the interpretation of the biodiversity broad habitat classification (terrestrial and freshwater types): definitions and the relationship with other classifications, JNCC Report No. 307. https://hub.jncc.gov.uk/assets/0b7943ea-2eee-47a9-bd13-76d1d66d471f.

Jenny, H., 1941. Factors of Soil Formation: A System of Quantitative Pedology. Dover Publications.

Keith, A.M., Robinson, D.A., 2012. Earthworms as natural capital: ecosystem service providers in agricultural soils. Econ. J. 2, 91–99.

Kibblewhite, M.G., Ritz, K., Swift, M.J., 2008. Soil health in agricultural systems. Philos. Trans. R. Soc. B Biol. Sci. 363 (1492), 685–701. https://doi.org/10.1098/rstb.2007.2178.

Kirk, G.J.D., Bellamy, P.H., Lark, R.M., 2010. Changes in soil pH across England and Wales in response to decreased acid deposition. Glob. Chang. Biol. 16 (11), 3111–3119. https://doi.org/10.1111/j.1365-2486.2009.02135.x.

Lal, R., 2008. Carbon sequestration. Philos. Trans. R. Soc. B Biol. Sci. 363 (1492), 815–830. https://doi.org/10.1098/rstb.2007.2185.

Lal, R., 2016. Soil health and carbon management. Food Energy Secur. 5 (4), 212–222. https://doi.org/10.1002/fes3.96.

Lehmann, J., Bossio, D.A., Kögel-Knabner, I., Rillig, M.C., 2020. The concept and future prospects of soil health. Nat. Rev. Earth Environ. 1 (10), 544–553. https://doi.org/10.1038/s43017-020-0080-8.

Lewis, J., 2020. Sustainable, Healthy, and Resilient: Practice-based Approaches to Land and Soil Management.

Liu, H., Lennartz, B., 2019. Hydraulic properties of peat soils along a bulk density gradient—a meta study. Hydrol. Process. 33 (1), 101–114. https://doi.org/10.1002/hyp.13314.

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. Chang. Biol. 20 (1), 313–326. https://doi.org/10.1111/gcb.12292.

Mason, K.E., Ashwood, F., Schmidt, O., Cosby, B.J., Keith, A.M., 2022. A Compendium of Earthworm Data Sources and Associated Information From the UK and Ireland, 1891-2021. NERC EDS Environmental Information Data Centre. https://doi.org/10.5285/1a1000a8-4e7e-4851-8784-94c7ba3e164f.

McKenzie, M.C., Moncada, M.A.P., Ball, B.C., 2015. Reductions of yield gaps and improvement of ecological function through local-to-global application of visual soil assessment. In: Ball, B.C., Munkholm, L.J. (Eds.), Visual Soil Evaluation: Realizing Potential Crop Production with Minimum Environmental Impact. CABI, pp. 31–48.

Moebius-Clune, B.N., Moebius-Clune, D.J., Gugino, B.K., Idowu, O.J., Schindelbeck, R.R., Ristow, A.J., van Es, H.M., Thies, J.E., Shayler, H.A., McBride, M.B., Kurtz, K.S.M., Wolfe, D.W., Abawi, G.S., 2017. Comprehensive Assessment of Soil Health - The Cornell Framework (3.2). Cornell University.

Morris, N.L., Miller, P.C.H., Orson, J.H., Froud-Williams, R.J., 2010. The adoption of non-inversion tillage systems in the United Kingdom and the agronomic impact on soil, crops and the environment - a review. Soil Tillage Res. 108 (1–2), 1–15. https://doi.org/10.1016/j.still.2010.03.004.

Morton, R.D., Marston, C.G., O'Neil, A.W., Rowland, C.S., 2021. Land Cover Map 2020 (10m Classified Pixels, GB). NERC EDS Environmental Information Data Centre https://doi.org/10.5285/35c7d0e5-1121-4381-9940-75f7673c98f7.

National Soil Resources Institute, 2001. NATMAP Vector. LandIS Land Information System. Cranfield University. https://www.landis.org.uk/data/nmvector.cfm.

Newell-Price, J.P., Whittingham, M.J., Chambers, B.J., Peel, S., 2013. Visual soil evaluation in relation to measured soil physical properties in a survey of grassland soil compaction in England and Wales. Soil Tillage Res. 127, 65–73. https://doi.org/10.1016/j.still.2012.03.003.

Panagos, P., Borrelli, P., Poesen, J., 2019. Soil loss due to crop harvesting in the European Union: a first estimation of an underrated geomorphic process. Sci. Total Environ. 664, 487–498. https://doi.org/10.1016/j.scitotenv.2019.02.009.

Pebesma, E., Bivand, R., Racine, E., Summer, M., Cook, I., Keitt, T., Lovelace, R., Wickham, H., Ooms, J., Müller, K., Pederson, T.L., Baston, D., Dunnington, D., 2023. Package "sf" (1.0-12). CRAN. https://cran.r-project.org/package=sf.

Prout, J.M., Shepherd, K.D., McGrath, S.P., Kirk, G.J.D., Haefele, S.M., 2021. What is a good level of soil organic matter? An index based on organic carbon to clay ratio. Eur. J. Soil Sci. 72 (6), 2493–2503. https://doi.org/10.1111/ejss.13012.

R Core Team, 2022. R: A Language and Environment for Statistical Computing. CRAN. https://www.r-project.org/.

Reynolds, B., Chamberlain, P.M., Poskitt, J., Woods, C., Scott, W.A., Rowe, E.C., Robinson, D.A., Frogbrook, Z.L., Keith, A.M., Henrys, P.A., Black, H.I.J., Emmett, B.A., 2013. Countryside survey: national "soil change" 1978-2007 for Topsoils in Great Britain-acidity, carbon, and total nitrogen status. Vadose Zone J. 12 (2). https://doi.org/10.2136/vzj2012.0114 vzj2012.0114.

Rinot, O., Levy, G.J., Steinberger, Y., Svoray, T., Eshel, G., 2019. Soil health assessment: a critical review of current methodologies and a proposed new approach. Sci. Total Environ. 648, 1484–1491. https://doi.org/10.1016/j.scitotenv.2018.08.259.

Robinson, E.L., Blyth, E.M., Clark, D.B., Cornyn-Platt, E., Rudd, A.C., 2020. Climate Hydrology and Ecology Research Support System Meteorology Dataset for Great Britain (1961-2017) [CHESS-met]. NERC Environmental Information Data Centre. https://doi.org/10.5285/2ab15bf0-ad08-415c-ba64-831168be7293.

RoTAP, 2012. Review of transboundary air pollution: acidification, acidification, eutrophication, ground level ozone and heavy metals in the UK. Centre for Ecology and Hydrology on behalf of Defra and the Devolved Administrations.

Rutgers, M., Mulder, C., Schouten, A.J.J., 2008. Soil ecosystem profiling in the Netherlands with ten references for biological soil quality. https://rivm.openrepository.com/handle/10029/260810.

Rutgers, M., Schouten, A.J., Bloem, J., Van Eekeren, N., De Goede, R.G.M., Op Akkerhuis, Jagers, G. A. J. M., Van Der Wal, A., Mulder, C., Brussaard, L., Breure, A.M., 2009. Biological measurements in a nationwide soil monitoring network. Eur. J. Soil Sci. 60 (5), 820–832. https://doi.org/10.1111/j.1365-2389.2009.01163.x.

Seaton, F.M., Barrett, G., Burden, A., Creer, S., Fitos, E., Garbutt, A., Griffiths, R.I., Henrys, P., Jones, D.L., Keenan, P., Keith, A., Lebron, I., Maskell, L., Pereira, M.G., Reinsch, S., Smart, S.M., Williams, B., Emmett, B.A., Robinson, D.A., 2021. Soil health cluster analysis based on national monitoring of soil indicators. Eur. J. Soil Sci. 72 (6), 2414–2429. https://doi.org/10.1111/ejss.12958.

Seaton, F.M., Robinson, D.A., Monteith, D., Lebron, I., Bürkner, P., Tomlinson, S., Emmett, B.A., Smart, S.M., 2023. Fifty years of reduction in sulphur deposition drives recovery in soil pH and plant communities. J. Ecol. 111 (2), 464–478. https://doi.org/10.1111/1365-2745.14039.

Simfukwe, P., Griffiths, R.I., Emmett, B.A., Jones, D.L., Hill, P.W., Cooper, D.M., Mills, R.T.E., Rowe, E., Spurgeon, D., Reynolds, B., 2010. Is Current Soil Classification Relevant to Soil Function and Soil Diversity?

Soil Survey of Scotland Staff, 1981. Soil Maps of Scotland at a Scale of 1:250,000. Macaulay Institute for Soil Research, Aberdeen. https://soils.environment.gov.scot/maps/soil-maps/national-soil-map-of-scotland/.

Soil Survey of Scotland Staff, 1984. Organisation and methods of the 1:250 000 soil survey of Scotland. Handbook 8.

Spurgeon, D.J., Keith, A.M., Schmidt, O., Lammertsma, D.R., Faber, J.H., 2013. Land-use and land-management change: relationships with earthworm and fungi communities and soil structural properties. BMC Ecol. 13 (46), 1–13. https://doi.org/10.1186/1472-6785-13-46.

Stroud, J.L., Goulding, K.W.T., 2022. Science and user-based co-development of a farmland earthworm survey facilitated using digital media: insights and policy implications. Ann. Appl. Biol. 181 (1), 70–79. https://doi.org/10.1111/aab.12766.

The Royal Society, 2020. Soil Structure and its Benefits: An Evidence Synthesis.

Therneau, T., Atkinson, B., Ripley, B., 2022. Package "rpart" (4.1.19). CRAN. https://cran.r-project.org/package=rpart.

van Groenigen, J.W., Lubbers, I.M., Vos, H.M.J., Brown, G.G., De Deyn, G.B., van Groenigen, K.J., 2014. Earthworms increase plant production: a meta-analysis. Sci. Rep. 4 (2), 1–7. https://doi.org/10.1038/srep06365.

Veerman, C., Correia, T.P., Bastioli, C., Biro, B., Bouma, J., Cienciala, E., Emmett, B., Frison, E.A., Grand, A., Filchew, L.H., Kriaučiūnienė, Z., Pogrzeba, M., Soussana, J.F., Olmo, C.V., Wittkowski, R., 2020. Caring for Soil Is Caring for Life – Ensure 75% of Soils Are Healthy by 2030 for Food, People, Nature and Climate. https://doi.org/10.2777/611303.

Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S. 4th ed. Springer.

Verheijen, F.G.A., Bellamy, P.H., Kibblewhite, M.G., Gaunt, J.L., 2005. Organic carbon ranges in arable soils of England and Wales. Soil Use Manag. 21 (1), 2–9. https://doi.org/10.1079/sum2005288.

Vogel, H.J., Eberhardt, E., Franko, U., Lang, B., Ließ, M., Weller, U., Wiesmeier, M., Wollschläger, U., 2019. Quantitative evaluation of soil functions: potential and state. Front. Environ. Sci. 7 (164), 1–15. https://doi.org/10.3389/fenvs.2019.00164.

Vos, C., Don, A., Hobley, E.U., Prietz, R., Heidkamp, A., Freibauer, A., 2019. Factors controlling the variation in organic carbon stocks in agricultural soils of Germany. Eur. J. Soil Sci. 70 (3), 550–564. https://doi.org/10.1111/ejss.12787.

Wickham, H., Chang, W., Henry, L., Pederson, T.L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D., 2023. Package "ggplot2" (3.4.1). CRAN. https://cran.r-project.org/package=ggplot2.