Rothamsted Research
Harpenden, Herts, AL5 2JQ

Telephone: +44 (0)1582 763133
Web: http://www.rothamsted.ac.uk/

Rothamsted Repository Download

A - Papers appearing in refereed journals

Prout, J. M., Shepherd, K. D., McGrath, S. P., Kirk, G. J. D. and Haefele, S. M. 2020. What is a good level of soil organic matter? An index based on organic carbon to clay ratio. *European Journal of Soil Science.* pp. 1-11.

What is a good level of soil organic matter? An index based on organic carbon to clay ratio

Running Title: Soil carbon index

J. M. Prout[1,2] | K. D. Shepherd[3] | S. P. McGrath[1] | G. J. D. Kirk[2] | S. M. Haefele[1]

[1] Department of Sustainable Agricultural Sciences, Rothamsted Research, West Common, Harpenden, Hertfordshire, AL5 2JQ, UK

[2] School of Water, Energy & Environment, Cranfield University, Cranfield, Bedford MK43 0AL, UK

[3] World Agroforestry (ICRAF), UN Avenue, P. O. Box 30677-00100, Nairobi, Kenya

Correspondence: Jonah Prout. E-mail: jonah.prout@rothamsted.ac.uk

Abstract

Simple measures of appropriate levels of soil organic matter are needed for soil evaluation, management and monitoring, based on readily-measurable soil properties. We test an index of soil organic matter based on the soil organic carbon (SOC) to clay ratio, defined by thresholds of SOC/clay ratio for specified levels of soil structural quality. The thresholds were originally delineated for a small number of Swiss soils. We assess the index using data from the initial sampling (1978–83) of the National Soil Inventory of England and Wales, covering 3809 sites under arable land, grassland and woodland. Land use, soil type, annual precipitation and soil pH together explained 21% of the variance in SOC/clay ratio in the dataset, with land use the most important variable. Thresholds of SOC/clay ratio of 1/8, 1/10 and 1/13 indicated the boundaries between 'very good', 'good', 'moderate' and 'degraded' levels of structural condition.

On this scale, 38.2, 6.6, and 5.6% of arable, grassland and woodland sites, respectively, were degraded. The index gives a method to assess and monitor soil organic matter at national, regional or sub-regional scales based on two routinely measured soil properties. Given the wide range of soils and land uses across England and Wales in the dataset used to test the index, we suggest it should apply to other European soils in similar climate zones.

Highlights

- We assess the use of SOC/clay ratios as guidelines for soil management in England and Wales.
- We use data from 3809 sites to assess thresholds based on work for Polish, French and Swiss soils.
- SOC/clay threshold values can indicate degraded and good soil structural condition.
- The thresholds show the effect of land use and provide an index for use in England and Wales.

Keywords: land use, soil clay content, soil organic carbon, soil structure.

1 | INTRODUCTION

What is a good level of soil organic matter? Maintaining and if possible increasing the level of soil organic matter is generally a good thing for most functions expected of soils, including carbon sequestration, and increased levels improve soil structure. Farmers, food producers and governments need to know their soil status in relation to a critical value of soil organic matter. However, as soil organic matter varies with land use, soil type, location and other variables an index for gauging the level of soil organic matter under given conditions needs to account for these variables.

Verheijen *et al.* (2005) derived indicative ranges of soil organic carbon (SOC) content for arable soils of England and Wales that are potentially attained under different types of management and environmental conditions, and they found that clay content, precipitation and depth of topsoil could explain 25% of the variation in SOC content. Clay soils under wetter conditions had higher values than more-sandy soils, and grassland soils had higher values than arable soils with similar clay content. Clay content is a key factor because of its effects on SOC protection including adsorption on mineral surfaces and within soil aggregates (Dungait *et al*., 2012; Six *et al.*, 2002). Under constant land management and organic matter inputs, soils tend towards a steady-state SOC content, with a capacity for stabilising SOC modelled as a function of clay content (Hassink, 1997; Hassink & Whitmore, 1997; Six *et al.*, 2002; Stewart et al., 2007).

Dexter *et al*. (2008) found that soil physical properties (bulk density, water retention characteristics and clay dispersibility) could be better explained by the relative amounts of SOC and clay content to each other than by their total contents. In their analysis of data on French and Polish arable and grassland soils, maxima of correlations between the mass of clay per unit mass of SOC and soil physical properties corresponded to SOC/clay = 1/10, the SOC/clay content ratio was a good indicator of soil physical conditions, and this ratio gave a general separation between the different land uses. These findings were subsequently supported by others in Danish and (de Jonge *et al*., 2009; Jensen *et al*., 2019; Schjønning *et al*., 2012). Johannes *et al*. (2017) developed the approach further, and, in an analysis of Swiss soils, defined SOC/clay thresholds of 1/8, 1/10 and 1/13 as indicating the boundaries between 'very good', 'good', 'suggest improvement' and 'poor' levels of structural condition.

In this paper, we assess the three SOC/clay thresholds of Johannes *et al*. (2017) for soils across different land uses and climates in England and Wales. We use data from the original sampling of the National Soil Inventory (NSI) which contains information on soils at 5662 sites under agricultural and non-agricultural land uses across the two countries (Bellamy *et al.*, 2005). This is a far larger dataset with greater variation in soils, environments and land use than the datasets used by Dexter *et al*. (2008) and Johannes *et al*. (2017), and so provides a more comprehensive test of the SOC/clay ratio. We have three objectives. First, to assess the variation in SOC/clay ratio and its drivers across the NSI dataset. Second, to test its ability to delineate soils of different structural quality. Third, to illustrate the use of the SOC/clay index for mapping soil carbon across England and Wales, and for gauging changes in a long-term experiment with contrasting organic and inorganic fertiliser treatments.

2 | MATERIALS AND METHODS

2.1 | National-scale data

The NSI was first conducted between 1978 and 1983. Topsoil (0–15 cm depth) samples were collected at the intersections of an orthogonal 5 km grid over the entire area. A full description of the survey methods, analytical methods and available data is given in the LandIS database (www.landis.org.uk; Proctor *et al*., 1998). We considered only arable, ley grassland, permanent grassland and woodland sites, and excluded sites without measurements of soil clay content, pH or depth of topsoil, or that were classified as 'peat'. To reduce the impact of sites with very high SOC content relative to clay content, we excluded 290 outliers with SOC/clay > third quartile + 1.5 × interquartile range. This gave 3809 sites. Figure 1 shows the distribution of the sites across the two countries and Table 1 gives summary statistics for SOC and clay contents.

Soils at each site were classified by major soil group (Avery, 1980). Data on soil carbonate content were obtained from field observations of fizzing on addition of HCl to samples on a five-point scale from non-calcareous to very calcareous.

Soil structural quality was characterised using the Agricultural Land Classification of England and Wales (MAFF, 1988), which gave scores of good, moderate or poor structural quality according to the texture and shape, size and development of aggregates, and friability of subsoil. The NSI contains values for each of these except friability, therefore we estimated based on the shape and size criteria (and where possible development of aggregates was taken into account) (Table S1).

Monthly average precipitation was obtained from the UKCP09 dataset (Met Office, 2017). Mean accumulated annual precipitation was calculated for the years 1910–1983, and values at each NSI site were intersected using ArcGIS version 10.4. (ESRI, 2015). Ranges for precipitation classes were taken from Verheijen et al. (2005): < 650, 650–800 and 800–1100 mm year^{-1} with the addition of "very wet" for annual precipitation > 1100 mm year^{-1}.

2.2 | Data analysis

Statistical analyses were performed in R version 3.5.0 (R Core Team, 2017). Random Forest analysis (package: randomForest; Liaw & Wiener, 2002) was used to analyse the variance of SOC/clay with land use, average annual precipitation, major soil group, pH, lower depth of topsoil, calcareous score and risk of flooding. A square-root transformation was applied to SOC/clay to reduce the skewness of the data. Three-quarters of the data ($n = 2857$) was used as a training set, and the RMSE and R^2 values of predictions of the remaining set ($n = 952$) were calculated. Training and sample sets were randomly selected. Spatial or other correlations across training and validation sets were unlikely because only topsoil samples were used and the minimum distance between sites was 5 km.

Chi-square tests were used to compare numbers of sites within SOC/clay ranges under different land uses and precipitation classes and to test the relationship between the SOC/clay thresholds and soil structure. We used the results of statistically significant chi-square tests to interpret interactions between variables, with contributions by specific combinations of variables to the chi-square statistic inferred from the differences between observed frequency and that expected if there was no interaction between the variables.

We tested SOC/clay thresholds of 1/8, 1/10 and 1/13 as indicating the boundaries between ‘very good’, ‘good’, ‘moderate’ and ‘degraded’ levels of structural condition, following Johannes *et al*. (2017).

Figures were produced using R package ggplot2 (Wickham, 2016) and maps were produced using QGIS 3.0.1-Girona (QGIS Development Team, 2020).

2.3 | Field-scale data

We assessed the effects of field-scale soil management on SOC/clay ratios relative to the threshold values using data from a long-term organic manuring experiment at

Woburn, Bedfordshire, UK (Mattingly, 1974). The experiment had eight treatments with four replicates: (1) peat for 6 yr then ley, (2) farmyard manure (FYM), (3) grass ley plus nitrogen, (4) grass-clover ley, (5) green manure (GM) for 6 yr then ley, (6) straw, and (7) and (8) two inorganic fertiliser treatments (details in Mattingly, 1974). Treatments were applied in two cycles (1965 to 1972 and 1979 to 1987), second cycle treatment denoted by ‘then’ above if different from first. We calculated SOC/clay ratios for each plot and then averaged the values for each treatment. The plot-level soil clay content ranged from 78 to 131 g kg^{-1} and the initial SOC content ranged from 5.7 to 8.6 g kg^{-1} (Table S2).

3 | RESULTS

3.1 | Variation in SOC and clay contents and SOC/clay ratio

Mean SOC contents increased in the order arable << ley grass < permanent grass ≈ woodland soils (Table 1). Mean clay contents and their ranges were similar across land uses, except those of woodland soils were smaller. The dominant soil types in all land uses were brown soils and surface-water gleys; the proportions of other soil groups varied (Table S3). Arable sites tended to have smaller average annual precipitation than the other land uses (Tables S4 and S5).

The proportions of sites above and below the three SOC/clay thresholds differed between land uses, particularly for the SOC/clay = 1/13 threshold (Figure 2 and Table 2). A greater proportion of arable sites had SOC/clay < 1/13 (i.e. depleted in SOC for their clay content) and a greater proportion of permanent grassland and woodland sites had SOC/clay > 1/8 (i.e. enriched in SOC for their clay content; $X^2(9) = 681.3$, $p < 0.001$).

Analysis of the influence of land use, soil and other variables on SOC/clay ratio by random forest analysis showed that 21.0% of the variance was explained by the variables examined (Table 3). Land use, average annual precipitation, major soil group and pH were more important than carbonate score, flood risk and depth of topsoil. When the model was run with just the top four variables, the variance explained did not change, however the importance of land use increased relative to the other variables.

3.2 | Effects of land use and precipitation

The effect of land use was clear with lower SOC/clay ratios observed for arable and predominantly higher SOC/clay ratios for grassland and woodland (Table 2). As there was some geographical relationship between the distributions of land use and precipitation, the effects of each on numbers of sites relative to the SOC/clay thresholds were considered. Verheijen *et al*. (2005) suggested that dry sandy soils were more at risk of lower SOC content than wetter clayey soils and that grassland soils would have higher SOC content than (ley-) arable soils. Comparing SOC/clay threshold ranges, land uses, and precipitation classes (< 650, 650 to 800, 800 to 1100 and > 1100 mm yr^{-1}; Table S6), two questions were asked: 1) were arable soils under dry climate conditions more likely to have SOC/clay < 13 than arable soils under wetter climate conditions, and 2) for soils under dry climate conditions (< 650 mm yr^{-1}), were arable soils more likely to have SOC/clay < 13 than other land uses?

In answer to the first question, chi-squared analysis showed that precipitation class was not independent of SOC/clay ratio for arable soils ($X^2(9) = 78.9$, $p < 0.001$). Comparing the contributions of each combination to the chi-square statistic showed that a larger number of soils receiving less than 650 mm yr^{-1} and smaller numbers of soils receiving more than 650 mm yr^{-1} than expected had SOC/clay < 1/13. Also, a smaller number of dry soils and larger number of soils with greater than 800 mm yr^{-1} had

SOC/clay > 1/8. This suggests that lower precipitation conditions were related to SOC/clay < 13 for arable soils.

Chi-squared analysis to answer the second question showed that land use was not independent of SOC/clay ratio for soils under dry climate conditions ($X^2(9) = 94.0$, $p < 0.001$). A larger number of arable soils and smaller number of grassland and woodland soils than expected had SOC/clay < 1/13 than if the land use was independent of SOC/clay ratio range for soils receiving < 650 mm yr^{-1} annual precipitation. For soils with SOC/clay > 1/8, the reverse was true (i.e. arable < grassland or woodland). This suggests that land use was affecting the number of dry climate soils with SOC/clay < 1/13.

The relative effects of land use, precipitation and soil type were evident from the distribution of the 820 sites with SOC/clay ratio < 13 across England and Wales (Figure S1). These sites were predominantly arable, and their distribution across eastern and central England confirmed the lesser statistical effect of precipitation and major soil group observed. Northwest England and Wales had notably few degraded sites though soils sampled there were mostly under non-arable land uses.

3.3 | Effects of soil type and soil pH

The statistical effect of major soil group appeared to be driven by three of the soil groups and some of this might already have been accounted for by land use (Figure 3). Podzolic soils tended to have SOC/clay > 1/8 and were mostly not arable, whereas clay-rich pelosols were more likely to have SOC/clay < 1/13 and a higher proportion were arable. The lower importance of soil group might be linked to the smaller sample sizes of the podzolic and pelosol soils compared with brown and gley soils for which SOC/clay ratios showed similar variation.

As pH decreased below pH = 5, the SOC/clay ratio tended to increase (Figure S2). Above pH = 5 there was less of a trend when considering permanent grass and woodland soils, however, arable and ley grass soils showed decreasing minimum SOC/clay ratio particularly above pH = 7, though sites with SOC/clay > 1/8 were still observed.

3.4 | Relation between structural quality and SOC/clay ratio

Structural quality – classified as good, moderate, moderate-degraded and degraded – tended to improve with increasing SOC/clay ratio as shown by the box plots in Figure 4 and the chi-squared test result for the relation between SOC/clay range between the thresholds and structural quality ($X^2(9) = 129.3$, $p < 0.001$). Most (82%) of the relationship between SOC/clay and structural quality was explained by (a) a larger than expected frequency of sites with SOC/clay < 1/13 and moderate-degraded or degraded structure; (b) a smaller than expected frequency of sites with SOC/clay > 1/8 and degraded structure; and (c) smaller than expected frequency of sites with SOC/clay < 1/13 and good structure.

3.5 | Variation in SOC/clay ratio across England and Wales

Mapping the index across the two countries (Figure 5) showed the effect of land use and geography at the time of survey. For any land use, degraded sites were not limited to a particular region. But, as previously mentioned, there were fewer degraded sites towards the northwest and in Wales. Calculating summary values of SOC/clay by land use (Table 4) showed that the minimum value increased slightly in the order: arable = ley grass < permanent grass < woodland. The median results showed a stronger difference between arable sites and the other land uses, with arable in the moderate category and the other land uses equal to or above the very good threshold. The different land uses have similar upper SOC/clay values as a result of excluding outliers.

3.6 | Changes in SOC/clay ratio with field management

Figure 6 shows changes in SOC/clay ratios over 30 years of the Woburn organic manuring experiment. Leys and treatments with organic matter application (straw, manures) showed similar trends of increasing SOC/clay ratio during the application period and decreasing ratio after the treatment was stopped, but with differing magnitudes. Peat and farmyard manure gave the largest increases, followed by the ley treatments and then straw. Inorganic fertiliser only treatments showed a general trend of decreasing SOC/clay ratio, and consistently occupied the degraded class.

4 | DISCUSSION

4.1 | Variation in SOC/clay ratio with land use and soil type

In agreement with Dexter *et al.* (2008) and Johannes *et al.* (2017), arable soils had a larger proportion of sites with SOC/clay ratios below 1/10, and permanent grassland soils had a larger proportion above 1/10. Dexter *et al*. (2008) did not consider soil group or structural condition of the soils in their study, and Johannes *et al*. (2017) chose only one soil type. Based on the agreement of their results with previous studies on the importance of the SOC/clay = 1/10, Johannes *et al.* (2017) suggested it should apply to a range of soils. Our finding that few grassland and woodland sites had SOC/clay < 1/13 supports the use of SOC/clay = 1/13 as an indicative threshold for degradation, as grassland and woodland soils are not generally subject to major disturbance and are close to semi-natural systems. Our analysis shows that many arable soils were depleted in SOC compared with the more natural systems. Ley grassland soils were intermediate between arable and permanent grassland soils. The NSI survey did not include information on the length of leys nor the time under ley at sampling, but typically this is between 3 and 8 years. Some proportion of arable sites will have been part of ley rotations at the time of sampling.

The large variation of SOC/clay ratio within each land use and soil group demonstrates that clay content is not the only determinant of SOC dynamics, especially considering land use history before the sampling will have big effects too. As discussed above, despite the scatter, the thresholds show differences between soils under different land management.

The variance of the SOC/clay ratio explained by random forest analysis was similar to the variance of SOC content explained by Verheijen *et al.* (2005) with step-wise general regression modelling, using similarly-derived precipitation data, and the same soil dataset (though a different subset). We would expect the variance explained to increase with more specific measures of land management within land use classes (crop type, residue treatment, land-use history and, for grassland systems, grazing management). Interpolated precipitation data is another estimation which could be improved, however this is what is generally available at this scale.

The effect of major soil group on SOC/clay ratio suggests some consideration should be given to soil type, as highlighted by Johannes *et al.* (2017). Comparing the variation in SOC/clay ratio between major soil groups, similar variation and medians were found for lithomorphic, brown, gley and man-made soils. The tendency for higher SOC/clay ratio of podzolic soils might be attributed to concentrated organic horizons in the topsoil. The tendency for lower SOC/clay ratios of Pelosols might be attributed to higher clay contents combined with a higher proportion (62%) being arable than most major soil groups. Whilst acidic soils had a tendency for higher SOC/clay ratios, there appeared to be little relationship between pH and SOC/clay ratio in agriculturally productive pH ranges (circa. pH = 5.5 to 7).

4.2 | Significance of the threshold values

The fact that the empirical threshold values found by Johannes *et al.* (2017) for Swiss soils also hold for the wide range of soils and land uses across England and Wales in our study, suggests they have some fundamental basis, and that they may apply in soils in similar climate zones across Europe. An association of soil structural quality with the SOC/clay = 1/10 ratio was expected from physico-chemical considerations (de Jonge *et al.*, 2009; Jensen *et al.*, 2019). Intuitively there will be some minimum range of SOC/clay ratio below which soil structure is impaired, and some maximum range above which the capacity of soil clays of given mineralogy to bind SOC is exceeded. However, there are no obvious reasons why the precise threshold values indicated by our and the Swiss study should be absolute.

The observed decrease in soil structural quality with decreasing SOC/clay ratio was statistically significant, though there was overlap between the boxplots of SOC/clay ratio between structure classes. Our analysis was limited by the quality of the available data on structure. This was based on the scheme defined for the Agricultural Land Classification of England and Wales, which includes a measure of friability. Since friability was not recorded in the NSI, we had to estimate structural quality without it, introducing error.

The mechanistic link between structural quality and SOC/clay ratio should reduce errors due to cross correlation with spatial and temporal variations in the data. We found, as did Verheijen *et al.* (2005), that SOC content tended to decrease with decreasing precipitation across England and Wales, partly in interaction with land use. However, low SOC/clay ratios were not limited to particular combinations of land use and precipitation; therefore, we would not consider precipitation to limit SOC/clay ratio in this data set and geographical range. Land management was shown to affect proportions of very good and degraded soils under dry (< 650 mm year^{-1}) climate conditions. So, SOC/clay ratios of at least 1/10 should be attainable in such soils.

4.3 | Practical usefulness of the index

The SOC/clay index is a simple measure to evaluate the SOC status of any given soil in England and Wales, independent of the land use. It will therefore be meaningful for experts and non-experts and has consequences for many soil functions beyond agricultural uses. It could allow farmers to identify degraded soils on their farms and adjust their management accordingly. It could also be used to monitor and understand the state of soils at a national scale to inform decision making and policy.

Application of the index to data from the long-term Woburn experiment showed its behaviour was consistent with expectations, with an improving index in treatments favouring organic matter accumulation, and a deteriorating index in soil-degrading treatments. This illustrates the magnitude and time taken for the various contrasting managements to change SOC and the index. The soil in the Woburn experiment is a sandy loam; the results show that the index can be used for soils with low clay content, despite the narrowing of the SOC/clay thresholds with decreasing clay content, and the relatively small changes in SOC content between the treatments. It should be noted that, to be useful for monitoring purposes, measurements of SOC and clay over time and between sites need to be consistent.

It would be interesting to look at other longer-term studies to explore a wider range of clay contents, treatments and time periods. Saturation concepts suggest that a soil closer to steady state or saturation limit should accumulate carbon more slowly than one further from saturation (Six *et al.*, 2002; Stewart *et al.*, 2007). Hence, whether sites with lower index values (higher degradation) improve more quickly could be tested.

5 | CONCLUSIONS

An index of soil organic matter with threshold SOC/clay ratios of 1/8, 1/10 and 1/13 satisfactorily separates the soils of England and Wales into very good, good, moderate and degraded classes of SOC content and physical structure condition. In agreement with previous publications, grassland and woodland soils mostly had SOC/clay ratio > 1/10, indicating that their SOC contents are close to or above the capacity for protection of SOC by interaction with clay particles. That these more natural systems tend to have SOC/clay ratio > 1/10 supports this as a suitable threshold for good condition.

Arable soils and soils receiving less annual rainfall were most likely to be physically degraded, though rainfall was a less important factor determining SOC/clay ratio. Very good status soils (SOC/clay > 1/8) occurred in low rainfall areas, even under arable management, suggesting that rainfall does not fundamentally limit SOC content in this climate.

The SOC/clay ratio index allows the evaluation of soils on a scale from degraded to good soil conditions. It therefore gives a ready metric for communication to experts and non-experts, enabling users to adjust their practices and decision makers to develop adequate policies. A SOC/clay ratio greater than 1/10 should be achievable for all managed soils of different textures. Many arable soils in England and Wales evidently have a substantial SOC deficit, suggesting a significant opportunity to increase SOC storage to both improve soil conditions and sequester carbon.

Being based on two routinely measured soil properties, the index provides a suitable means of monitoring SOC at national, regional or sub-regional scales. Given the wide range of soils and land uses across England and Wales in the dataset used to test the index and agreement with literature using French, Polish and Swiss soils, it should apply to other European soils in similar climate zones.

ACKNOWLEDGEMENTS

We thank John Hollis, formerly of Cranfield University, for his advice on the structural quality assessment using the NSI dataset. We thank the e-RA data managers and Rothamsted Research for access to data and information on the Rothamsted Long-term Experiments. We also thank Paul Poulton and Andy Macdonald for their helpful comments and additional information. Mapping derived from soils data © Cranfield University (NSRI) and for the Controller of HMSO 2019.

Funding Information

This research was funded by a PhD studentship for JP through the Rothamsted-Cranfield Soil Agriculture Research and Innovation Accelerator. Rothamsted Research receives grant aided support from the Biotechnology and Biological Sciences Research Council (BBSRC) of the United Kingdom. SPM and SMH are funded by the Institute Strategic Programme (ISP) grant "Soils to Nutrition" (BBS/E/C/000I0310). The Rothamsted Long-term Experiments National Capability (LTE-NC) is supported by BBSRC (BBS/E/C/000J0300) and the Lawes Agricultural Trust.

Data Sharing and Data Accessibility statement

The NSI dataset is held by Cranfield University and accessed via LandIS (www.landis.org.uk). UKCP09 data are available from the Centre for Environmental Data Analysis (CEDA) archive (http://catalogue.ceda.ac.uk/uuid/94f757d9b28846b5ac810a277a916fa7). Data for the Woburn Organic Manuring experiment can be obtained via the Electronic Rothamsted Archive (era.rothamsted.ac.uk).

Conflict of interest statement

The authors have no conflicts of interest related to the work presented in this manuscript.

Authorship

Study concept and design: all. Analysis and interpretation of data: Prout, Shepherd, Haefele. Drafting of the manuscript: Prout. Critical revision of the manuscript for important intellectual content: Kirk, McGrath, Haefele, Shepherd. Statistical analysis: Prout and Shepherd.

REFERENCES

Avery, B.W. (1980). Soil Classification for England and Wales (Higher Categories). Soil Survey Technical Monograph No. 14. Harpenden.

Bellamy, P.H., Loveland, P.J., Bradley, R.I., Lark, R.M. & Kirk, G.J.D. (2005). Carbon losses from all soils across England and Wales 1978–2003. *Nature*, *437,* 245–248. DOI:10.1038/nature04038

de Jonge, L.W., Moldrup, P. & Schjønning, P. (2009). Soil infrastructure, interfaces & translocation processes in inner space (“Soil-it-is”): Towards a road map for the constraints and crossroads of soil architecture and biophysical processes. *Hydrology and Earth System Sciences*, *13,* 1485–1502. DOI:10.5194/hess-13-1485-2009

Dexter, A. R., Richard, G., Arrouays, D., Czyz, E. A., Jolivet, C., & Duval, O. (2008). Complexed organic matter controls soil physical properties. *Geoderma*, *144*(3–4), 620–627. DOI:10.1016/j.geoderma.2008.01.022

Dungait, J.A.J., Hopkins, D.W., Gregory, A.S. & Whitmore, A.P. (2012). Soil organic matter turnover is governed by accessibility not recalcitrance. *Global Change*

Biology, *18*, 1781–1796. DOI:10.1111/j.1365-2486.2012.02665.x

Environmental Systems Research Institute (ESRI). (2015). ArcGIS Version 10.4. Redlands, CA: Environmental Systems Research Institute, Inc..

Feller, C. & Beare, M.H. (1997). Physical control of soil organic matter dynamics in the tropics. *Geoderma*, *79,* 69–116. DOI:10.1016/S0016-7061(97)00039-6

Greenland, D. J., Rimmer, D., & Payne, D. (1975). Determination of the structural stability class of English and Welsh soils, using a water coherence test. *Journal of Soil Science*, *26*(3), 294–303. DOI:10.1111/j.1365-2389.1975.tb01953.x

Hassink, J. & Whitmore, A.P. (1997). A model of the physical protection of organic matter in soils. *Soil Science Society of America Journal, 61,* 131. DOI: 10.2136/sssaj1997.03615995006100010020x

Hassink, J. (1997). The capacity of soils to preserve organic C and N by their association with clay and silt particles. *Plant and Soil, 191,* 77–87. DOI:10.1023/A:1004213929699

Jensen, J. L., Schjønning, P., Watts, C. W., Christensen, B. T., Peltre, C., & Munkholm, L. J. (2019). Relating soil C and organic matter fractions to soil structural stability. *Geoderma*, *337*, 834–843. DOI:10.1016/j.geoderma.2018.10.034

Johannes, A., Matter, A., Schulin, R., Weisskopf, P., Baveye, P. C., & Boivin, P. (2017). Optimal organic carbon values for soil structure quality of arable soils. Does clay content matter? *Geoderma, 302,* 111. DOI:10.1016/j.geoderma.2017.04.021

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News 2*(3), 18-22.

Loveland, P. & Webb, J. (2003). Is there a critical level of organic matter in the agricultural soils of temperate regions: A review. *Soil and Tillage Research*, *70*, 1–18. DOI:10.1016/S0167-1987(02)00139-3

Mattingly, G.E.G. (1974). The Woburn Organic Manuring Experiment: I. Design, Crop Yields and Nutrient Balance, 1964-72. Rothamsted Experimental Station, Report for 1973, Part 2, 98-133. DOI:10.23637/ERADOC-1-8

Ministry of Agriculture, Fisheries and Food (MAFF). (1988). Agricultural Land Classification of England and Wales: Revised criteria for grading the quality of agricultural land (ALC011).

Met Office (2017). UKCP09: Met Office gridded land surface climate observations - monthly climate variables at 5km resolution. Centre for Environmental Data Analysis. URL: http://catalogue.ceda.ac.uk/uuid/94f757d9b28846b5ac810a277a916fa7 [accessed on 29 November 2017]

Proctor, M.E., Siddons, P.A., Jones, R.J.A., Bellamy, P.H., Keay, C.A. (1998). Land information systems for sustainable development in the UK. In: Heineke, H.J., Ecklemann, W., Thomasson, A.J., Jones, R.J.A, Montanarella, L., Buckley, B. (Eds.), EUR 17729 EN (pp. 219 – 233). Office for Official Publications of the European Communities, Luxembourg.

QGIS Development Team. (2020). QGIS Geographic Information System. Open Source Geospatial Foundation Project. URL http://qgis.osgeo.org

R Core Team. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Schjønning, P., de Jonge, L.W., Munkholm, L. J., Moldrup, P., Christensen, B. T., & Olesen, J. E. (2012). Clay dispersibility and soil friability-testing the soil clay-to-carbon saturation concept. *Vadose Zone Journal*, *11*(1).

DOI:10.2136/vzj2011.0067

Six, J., Conant, R.T., Paul, E.A. & Paustian, K. (2002). Stabilization mechanisms of soil organic matter: Implications for C-saturation of soils. *Plant and Soil*, *241*, 155–176. DOI:10.1023/A:1016125726789

Stewart, C.E., Paustian, K., Conant, R.T., Plante, A.F. & Six, J. (2007). Soil carbon saturation: Concept, evidence and evaluation. *Biogeochemistry*, *86*, 19–31. DOI:10.1007/s10533-007-9140-0

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 and Management*, *21*, 2–9. DOI:10.1111/j.1475-2743.2005.tb00099.x.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. DOI:10.1007/978-3-319-24277-4.

Table 1 Soil organic carbon (SOC) and clay contents by land use class in the National Soil Inventory

		SOC content (g kg^{-1})				Clay content (g kg^{-1})			
	n	Mean	Med.	Min.	Max.	Mean	Med.	Min.	Max.
Arable	1661	25	22	4	126	262	247	26	879
Ley grass	602	34	31	7	109	267	257	60	756
Permanent grass	1277	42	39	6	138	281	260	47	795
Woodland	269	40	37	1	158	251	242	10	606
All land uses	3809	34	30	1	158	268	252	10	879

Table 2 Percentages of sites above, below and between SOC/clay thresholds of 1/8, 1/10 and 1/13 for each land use and land use precipitation class combination.

		Percentage of sites with indicated SOC/clay ratio			
	n	≥ 1/8	<1/8 ≥1/10	<1/10 ≥1/13	<1/13
Arable	1661	28.8	14.0	19.0	38.2
Ley grass	602	50.2	20.3	14.6	15.0
Permanent grass	1277	66.9	15.4	11.1	6.6
Woodland	269	67.7	16.0	10.8	5.6

Table 3 Contributions of indicated variables to variance in SOC/clay ratio analysed using random forests. Training data was a random selection of 75% of the data (*n* = 2857). With all seven explanatory variables, root mean square error (RMSE) for training data = 0.06, $R^2 = 0.21$; RMSE for remaining data = 0.07; $R^2 = 0.21$. With only top four variables, RMSE for training data = 0.06, $R^2 = 0.21$; RMSE for remaining data = 0.06; $R^2 = 0.22$.

	Increase of mean square error (%)	
Land use	32.7	39.8
Annual precipitation	28.0	26.0
Major soil group	26.4	20.3
pH	22.5	20.3
Depth of topsoil	10.4	
Carbonate score	10.0	
Risk of flooding	5.2	

Table 4 Summary of SOC/clay ratio decimal values calculated for each land use in the NSI subset

	SOC/clay ratio			
	Mean	Median	Min.	Max.
Arable	0.109	0.090	0.018	0.357
Ley grass	0.139	0.125	0.018	0.359
Permanent grass	0.165	0.154	0.022	0.360
Woodland	0.174	0.160	0.025	0.355

FIGURE CAPTIONS

FIGURE 1 Map of arable, ley grass, permanent grass and woodland sites in the National Soil Inventory sampled between 1978 and 1983 (n = 3809).

FIGURE 2 Soil organic carbon content as a function of clay content for different land uses. Lines are SOC/clay thresholds: solid = 1/8, dashed = 1/10, dot-dash = 1/13.

FIGURE 3 Box plots of SOC/clay ratio for each major soil group. Horizontal lines are SOC/clay thresholds: solid = 1/8, dashed = 1/10, dotted = 1/13. Abbreviated major soil groups: Terr. raw = terrestrial raw, Lith. = lithomorphic, SW gley = surface-water gley, GW gley = ground-water gley.

FIGURE 4 Box plots of SOC/clay ratio for each structural quality score. Horizontal lines are SOC/clay thresholds: solid = 1/8, dashed = 1/10, dotted = 1/13. Numbers of samples in each group were n = 2250, 1111, 229 and 208 for Good, Moderate, Moderate-Degraded and Degraded, respectively.

FIGURE 5 Maps of SOC/clay ratio across England and Wales under (a) arable, (b) ley grass, (c) permanent grass, and (d) woodland coloured by SOC/clay index.

FIGURE 6 Changes over time in SOC/clay ratio in the Woburn long-term manuring experiment. Points are treatment means. Horizontal lines are thresholds separating degraded (SOC/clay < 1/13), moderate (SOC/clay = 1/13–1/10), good (SOC/clay = 1/10–1/8) and very good (SOC/clay > 1/8) soil conditions. Treatments were applied in two cycles (1965 to 1972 and 1979 to 1987); peat and green manure (GM) treatments were replaced by grass ley for the second cycle. Fert. 1 = (PKMg) ≡ Straw plus P, Fert. 2 = (PKMg) ≡ FYM.