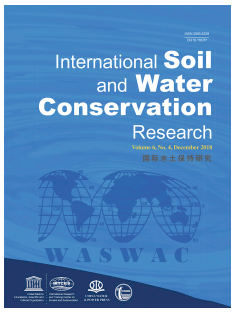
Journal Pre-proof



Soil porosity prediction across Europe with a focus on soil particle density determination

David A. Robinson, Arthur Fendrich, Amy Thomas, Sabine Reinsch, Jens Leifeld, Tim Moore, James Seward, Pasquale Borrelli, Cristiano Ballabio, Panos Panagos

PII: S2095-6339(26)00006-7

DOI: https://doi.org/10.1016/j.iswcr.2026.100614

Reference: ISWCR 100614

To appear in: *International Soil and Water Conservation Research*

Received Date: 1 May 2025

Revised Date: 8 January 2026

Accepted Date: 19 January 2026

Please cite this article as: Robinson D.A., Fendrich A., Thomas A., Reinsch S., Leifeld J., Moore T., Seward J., Borrelli P., Ballabio C. & Panagos P., Soil porosity prediction across Europe with a focus on soil particle density determination, *International Soil and Water Conservation Research*, https://doi.org/10.1016/j.iswcr.2026.100614.

This is a PDF of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability. This version will undergo additional copyediting, typesetting and review before it is published in its final form. As such, this version is no longer the Accepted Manuscript, but it is not yet the definitive Version of Record; we are providing this early version to give early visibility of the article. Please note that Elsevier's sharing policy for the Published Journal Article applies to this version, see: https://www.elsevier.com/about/policies-and-standards/sharing#4-published-journal-article. Please also note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2026 Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd.

Journal Pre-proof

Soil porosity prediction across Europe with a focus on soil particle density determination

David A. Robinson[1], Arthur Fendrich[2], Amy Thomas[1], Sabine Reinsch[1], Jens Leifeld[3], Tim Moore[4], James Seward[4], Pasquale Borrelli[5], Cristiano Ballabio[2], Panos Panagos[2]

Affiliations:

1 UK Centre for Ecology & Hydrology, Bangor, UK

2 European Commission, Joint Research Centre (JRC), Ispra, Italy

3 Agroscope, Reckenholz-Tänikon Research Station ART, Zurich, Switzerland.

4 Department of Geography, McGill University, Montreal, QC, Canada

5 Department of Science, Roma Tre University, Rome, Italy

Journal Pre-proof

Journal Pre-proof

Soil porosity prediction across Europe with a focus on soil particle density determination

ABSTRACT

This study emphasizes the critical role of soil porosity as an environmental variable influencing infiltration, compaction, runoff, and erosion, which are inversely related to bulk density. An analysis of topsoil porosity across Europe (0-20cm) was conducted using data from the LUCAS monitoring program, focusing on the fine earth fraction of soils. The conversion from bulk density to porosity – more intuitive for hydrological studies - requires knowledge of the particle density of both mineral and organic components, which is often lacking. A novel method was developed to estimate the particle density of organic matter using stoichiometric datasets from various land use types, resulting in an EU LUCAS average soil particle density of 2.53 g cm^{-3}. The generated fine earth porosity map aligns with high porosity areas in Northern Europe's peatlands and Central Europe's forests, providing insights into soil densification processes linked to compaction from traffic or organic matter depletion due to land use changes. This understanding is crucial for assessing compaction and erosion risk.

Key words: Bulk density, soil organic matter, stoichiometric modelling, LUCAS soil survey, soil science

Journal Pre-proof

1 INTRODUCTION

Soil porosity, the proportion of void space among soil particles (Dippenaar, 2014; Nimmo, 2004), is essential for gas fluxes and the infiltration, movement, transport, and retention of soil water. Moreover, it is an important indicator of soil health with declining porosity indicating densification and potential compaction which can lead to enhanced runoff and greater erosion risk (Gupta, Borrelli, Panagos, & Alewell, 2024; Holz, Williard, Edwards, & Schoonover, 2015). Increasing the pore space, for example using plant cover to which porosity is correlated (Thomas et al., 2024), is an important way to mitigate such effects (Zuazo & Pleguezuelo, 2009). Moreover, recent findings indicate that soil macroporosity is dynamic on continental and decadal time scales with unknown consequences for soil hydrological functioning (Hirmas et al., 2018).

Hirmas et al. (2018), found that predictions based on effective porosity, in five different physiographic regions of the USA, based on predicted changes in mean annual precipitation up to 2100 mm, resulted in soil saturated hydraulic conductivity altering between −55 to 34%. Two important advances came from this work, 1) that we understand that soil porosity, especially macroporosity alters on much shorter time scales than previously considered, 5-10 yr time cycles. 2) that the porosity alters due to feedback from climate, presumably through alteration of the flora and physical cracking of soils. A similar recent study for China (Kang, Zhang, Wu, & Zhao, 2024) found similar results with effective porosity higher in drylands compared to humid regions resulting in dryland soils being less conducive to soil water conservation and vegetation development. Several important studies have gone on to examine the implications of better incorporating soil structure into large scale regional or global models. Fatichi et al. (2020) found that the inclusion of better soil structure characterization in Earth System Models affected local hydrologic response. However, they concluded that the implications for global-scale climate

remains elusive in current Earth System Models. More recently, Wankmüller et al. (2024) have shown that global influence of soil texture on ecosystem water limitation, hence factors, such as porosity, affecting the ability of soils to retain moisture will impact the drought resilience of ecosystems. In temperate systems the porosity and water retention are also likely to be influenced strongly by soil organic matter (Robinson et al., 2025; Robinson et al., 2022; Thomas et al., 2024). This increasingly active area of research indicates the importance of improving our ability to measure and predict soil characteristics in space and time related to hydrological function in order to better understand climate soil feedback. Pore sizes can vary significantly, encompassing both large macropores, which promote swift water drainage and air circulation, and smaller micropores, which are responsible for the retention of water and nutrients. The intricate network of pores is vital for soil conservation, ecosystem functionality, effective water management, agricultural productivity, and the sustainability of environmental systems. A comprehensive understanding of soil porosity and bulk density is critical for assessing soil health, particularly in relation to water resilience, ecosystem processes, biomass generation, and carbon storage (Robinson et al., 2022). Moreover, porosity is susceptible to degradation by compaction or consolidation which densify the soil. Porosity and bulk density are essentially emergent properties based on a hierarchy of structures from grains to clay domains (tactoids), micro- and macro-aggregates and peds. Adopting a stereoscopic perspective, encompassing both microscopic and macroscopic dimensions (Scarlett, Van Der Kraan, & Janssen, 1998) can therefore clarify what leads to this emergence.

From a macroscopic perspective, soil porosity is quantified as the volume of pore space relative to the total volume of soil Eq. (1) and exhibits an inverse relationship with bulk density. Fundamentally, the packing of granular particles leads to the macroscopic – bulk relationship

between porosity (φ, $cm^3\ cm^{-3}$), bulk density (ρ_b, $g\ cm^{-3}$), the soil particle density (ρ_{ss}, $g\ cm^{-3}$), and the packing fraction (η, $cm^3\ cm^{-3}$), Eq. (1):

$$\varphi = 1 - \left(\frac{\rho_b}{\rho_{ss}}\right) = (1 - \eta) \quad (1)$$

Numerous elements affect the relationship between bulk density and soil porosity, such as soil texture, structural composition, and the content of clay and organic matter (Robinson et al., 2022). However, a significant challenge arises when attempting to convert bulk density, which is frequently assessed, into porosity due to the necessity of knowing the soil particle density, a parameter that is not typically measured, or difficult to measure routinely. A value of 2.65 $g\ cm^{-3}$ is often used as a proxy for the soil particle density in textbooks as it represents the particle density of quartz, a common constituent of many temperate soils (Brady & Weil, 2008).

Better estimates of soil porosity require knowledge of the particle densities of the soil materials, primarily organic and mineral constituents. However, as noted by (Rühlmann, Körschens, & Graefe, 2006), the calculation of particle density is not merely a summation of the individual densities. This complexity arises because the impact of each component on the overall particle density is contingent upon both its mass fraction and the volume it occupies. Consequently, since the densities do not scale linearly with their mass fractions, it is essential for the equation to incorporate variations in the intrinsic volume contributions of the different components (Ruehlmann, 2020; Ruehlmann & Körschens, 2020). To address this, Ruehlmann (2020) proposed a suitable mixing equation for soils Eq. (2).

$$\rho_{ss} = \frac{1}{\frac{SOM}{\rho_{sOM}} + \frac{(1-SOM)}{\left[\frac{Clay}{\rho_{sC}} + \frac{Silt}{\rho_{sSi}} + \frac{Sand}{\rho_{sSa}}\right]^{-1}}} \quad (2)$$

Where ρ_{sOM} represents the particle density of organic matter and the particle density of the mineral fraction is divided between, clay (ρ_{sC}), silt (ρ_{sSi}) and sand (ρ_{sSa}), respectively. SOM is the fraction of soil organic matter (0-1), hence the mineral matter fraction is 1-SOM. *Clay*, *Silt* and *Sand* are the respective fractions of the texture components which all add together to 1. Ruehlmann and Körschens (2020) utilized a comprehensive global dataset to estimate the particle densities of *Clay*, *Silt*, and *Sand* fractions, which were found to be 2.76, 2.69, and 2.66 g cm^{-3}, respectively. Additionally, using a regression method Ruehlmann and Körschens (2020) suggested that soil organic matter (*SOM*) could be categorized into low-density (1.27 g cm^{-3}) and high-density (1.43 g cm^{-3}) fractions. However, the determination of the most suitable value, or values, of SOM particle density remains an open research question, particularly when assessing large spatial areas. Although not explored, Ruehlmann (2020) suggested that using the stochiometric values of the components of SOM could be one way to determine a value for the particle density of the organic fraction.

In addition, the same modelling approach can also be applied to determine the soil bulk density (Adams, 1973) Eq. (3) and the total porosity Eq. (4), according to:

$$\rho_b = \frac{1}{\frac{SOM}{\rho_{bOM}} + \frac{1-SOM}{\rho_{bM}}} \tag{3}$$

$$\varphi = 1 - \left[\left[\frac{1}{\frac{SOM}{\rho_{bOM}} + \frac{1-SOM}{\rho_{bM}}} \right] \div \left[\frac{1}{\frac{SOM}{\rho_{sOM}} + \frac{1-SOM}{\rho_{sM}}} \right] \right] \tag{4}$$

Where the bulk density of the organic matter (ρ_{bOM}) and mineral material (ρ_{bM}) is for the end members of all OM or all mineral material, SOM is a fraction (0-1). The values of end members will vary quite widely, but Robinson et al. (2022) proposed values of 1.98 g cm^{-3} for ρ_{bM} and ~0.1 g cm^{-3} for ρ_{bOM} that captured the general response of a national data set. A value of 1.98 g cm^{-3} is

105 equivalent to the bulk density of a binary mixture of hard spheres with a ~10:1 size ratio where the
106 small completely infill the voids between the large, assuming a porosity of 0.252 and particle
107 density 2.65 g cm^{-3}; and 0.1 g cm^{-3} was based on data from bogs. As above ρ_{sOM} represents the
108 particle density of organic matter and ρ_{sM} represents the particle density of mineral material.

109 The transformation of bulk density into porosity is advantageous for hydrological studies and
110 offers a more intuitive understanding of the pore space that is comparable, and of the water
111 retention capabilities of soil layers. Moreover, water retention and water holding capacity are both
112 descriptors in the EU monitoring law to which porosity is valuable in determining. The conversion,
113 from bulk density to porosity, necessitates knowledge of the soil particle density, Eq. (1).
114 Accurately measuring this value poses challenges; nonetheless, significant advancements have
115 been achieved by Rühlmann et al. (2006) in predicting the particle densities of both mineral and
116 organic matter (Ruehlmann, 2020; Ruehlmann & Körschens, 2020). Rühlmann et al. (2006)
117 reported that soil organic matter (SOM) particle density exhibited variability within the range of
118 approximately 1.13 to 1.50 g cm^{-3}. Their findings indicated that as the quantity of SOM increased,
119 so did its density, which they attributed to qualitative changes in SOM resulting from
120 decomposition processes. This observed range aligns well with the values suggested by Redding
121 and Devito (2006), which span from 0.9 to 1.55 g cm^{-3}.

122 Ruehlmann (2020), referencing the findings of Tipping, Somerville, and Luster (2016),
123 asserted that lower soil particle densities corresponded with nutrient-rich cropland soils, while
124 higher densities were indicative of nutrient-poor soils. He classified these into two categories: the
125 low-density fraction (SOMld) and the high-density fraction (SOMhd). Earlier, Rühlmann et al.
126 (2006) had provided significant insights by positing that the density of soil organic matter (SOM)
127 is influenced by both the quality of the SOM and the content of soil organic carbon (SOC).

Journal Pre-proof

(Ruehlmann, 2020) indicated that low- density SOM is typically found in nutrient rich soils abundant in nitrogen (N), phosphorus (P), and sulfur (S), although these soils may also exhibit lower SOC content in the SOM, such as 0.42 g g^{-1}, and reduced density. In contrast, high-density SOM is likely derived from nutrient-poor soils characterized by low N:C, P:C, and S:C ratios, along with a higher average carbon concentration in the SOM, exemplified by a value of 0.53 g g^{-1} as noted by Pribyl (2010). Rühlmann et al. (2006) suggested the importance of microbes, while Ruehlmann (2020), based on the observations of Tipping et al. (2016) also suggested that the prevalence of microbes with low organic matter density, approximately 1.15 g cm^{-3}, in nutrient-rich soils could explain these observations. Ruehlmann (2020) concluded that employing a stoichiometric approach could unveil new avenues for exploring SOM density.

Given the aim of the paper, to provide EU wide maps of topsoil porosity, the objective of this paper is to convert total bulk density (Panagos et al., 2024) to fine earth porosity for the EU scale. The added value of this conversion is to have a direct assessment of the amount of void space related to functional capacity, with porosity providing a more standardized metric than bulk density which varies distinctly across soil textures. The novel aspect of this work is the use of stoichiometric data to predict both soil organic matter and organic matter density as a function of land use, or cover, to tighten the estimate of the soil particle density for generic land use or covers appropriate to the EU scale. Moreover, we use an additional novel conversion of SOC to SOM using values appropriate to the carbon density of different habitats. Given the sand, silt and clay fractions, this provides a framework for estimating organic matter particle density based on land cover and hence improving the continental scale prediction of porosity. This advancement facilitates a deeper understanding of how porosity may be influenced by changes in SOM affected by land management or shifts in land use.

Journal Pre-proof

2 MATERIALS AND METHODS

2.1 LUCAS topsoil data

The geographic scope of the study covers the 27 Member States of the European Union (EU) and the United Kingdom. It captures a temporal snapshot of the porosity, as it is based on topsoil data (0-20cm) from the 2018 Land Use and Cover Area Frame Statistical Survey (LUCAS), sampled in this region between June-August. The LUCAS topsoil survey 2018 included just under 20,000 topsoil data points for measured physical, chemical and biological properties (Orgiazzi, Ballabio, Panagos, Jones, & Fernández-Ugalde, 2018). This was the third campaign of LUCAS and for the first time total bulk density was measured for almost 6,000 locations across the EU and UK for 0-10 and 10-20 cm (Orgiazzi et al., 2022). The sampling strategy for bulk density points is like the one used to select the LUCAS 2018 points which includes criteria such as land use/cover, soil properties and topography. The highest number of points were surveyed in Spain, France, Sweden, Poland, Finland and Italy (Panagos et al., 2024).

The bulk density samples were left to air-dry followed by a recording of their weight. A subsample (3-5 g of soil) was then oven-dried at 105°C until it reached a constant weight. The final total bulk density for each location was then calculated following the adapted ISO 11272:2017 (Fernandez-Ugalde et al., 2022). Spurious points were rejected from the analysis data set, any points with a bulk density less than 0.1 g cm^{-3} or greater than 2.0 g cm^{-3}. After quality controls, a bulk density database of 5,659 well distributed points (0-10 and 10-20 cm samples) based on stratification from across the EU was developed, of these 5,659 covered a complete depth of 0-20cm. The total bulk density ($\rho_{b\,T}$) was converted to the bulk density of the fine earth ($\rho_{b\,FE}$)

according to the conversion ($\rho_{b\,FE} = (\rho_{b\,T} \times (1\text{-}M_{CF}) \times 2.6)/(2.6\text{-}\rho_{b\,T} \times M_{CF})$, where the particle density of the coarse fragments is assumed to be 2.6, and M_{CF} is the mass of coarse fragments.

2.2 Stochiometric analysis and data

Soil organic matter particle density is considered to range between 1.1 and 1.5 g cm^{-3} according to (Ruhlmann, 2006) and the references therein. In the development of pedo-transfer functions to predict soil particle densities Ruehlmann (2020) suggested that the approach could be refined using stoichiometry such as used by Tipping et al. (2016), to analyze the quality of organic matter. Kuwata, Zorn, and Martin (2012) presented such an approach using stoichiometry to predict organic compound density of organic matter (ρ_{sOM}). They developed an equation based on hydrogen, carbon and oxygen (H:C and O:C ratios). The basic approach uses the molecular weight (*MW*), molecular volume (V_m) and the intermolecular volume (V_{im}), with A as a unit conversion factor:

$$\rho_{SOM} = \frac{mass}{volume} = \frac{1}{A}\frac{MW}{(V_m + V_{im})} \quad (5)$$

They state that Eq. (5) predicts particle density for pure compounds to within an error of ±5% However, the inputs are not always known and hence, Kuwata et al. (2012) proposed to make predictions using elemental ratios instead which are more commonly measured. They thus formulated Eq. (6) to predict organic material particle density.

$$\rho_{SOM}\ (g\ cm^{-3}) = \frac{12+1(\mathrm{H:C})+16(\mathrm{O:C})}{7.0+5.0(\mathrm{H:C})+4.15(\mathrm{O:C})} \quad (6)$$

Kuwata et al. (2012) tested Eq. (6) on 31 pure compounds and found an error of ±12%, where densities ranged from 0.77 – 1.9 g cm^{-3}. This was about double that of Eq. (5) but represented a

practical more measurable alternative. Certain compounds such as oxalic acid, xylitol, and cholesterol drove the nonconformity error. In addition, Kuwata et al. (2012) were interested in predicting the density of complex mixtures of organic materials such as those found in aerosols. They created a range of mixed secondary organic materials and tested Eq. (6) which predicted the particle density within the 12% error envelope. They concluded that Eq. (6), developed using pure organic compounds, was also accurate for predicting the density of secondary organic matter that constituted a complex mixture of organic compounds. Here we assume the equation holds for soil organic matter. Further validation, specific to soil organic materials to confirm this would be a welcome addition to the literature. Hence, assuming the applicability of Eq. (6) for SOM and given the elemental ratios of SOM, Eq. (6) provides a means to estimate the density of the soil organic matter.

Stoichiometric data sets for soils focus on organic soils to avoid complications with mineral components. Hence, the approach is used more for wetland studies such as the values for natural ecosystems such as bog, fen and swamp (Moore, Large, Talbot, Wang, & Riley, 2018). In addition, Leifeld, Klein, and Wüst-Galley (2020) recently published results for organic soils across Switzerland under different land uses. This data set, comprised of 1165 soil samples from four different land uses, and provides a way to predict ρ_{sOM} based on Eq. (6). The dataset was used to obtain median values for organic matter densities based on land use for, woodland (including shrubland), grassland, cropland, and bare soil. A data set adding bog, fen and swamp was obtained from (Moore et al., 2018), where fens, are fed by streams and rivers; bogs, fed by rainwater; and swamps, distinguished by the presence of trees and shrubs. The data offered a mixture of depths to >5 m and so the data was split into those values for 0-20cm and the entire dataset with all depths for comparison.

Journal Pre-proof

2.3 Mapping

Soil organic carbon (SOC) was measured in LUCAS and was converted to SOM for the purposes of calculations. To convert SOC to SOM a single conversion factor such as a value of 1.82 is often used for the conversion, equivalent to 0.55 for the conversion of SOM to SOC (Lebron et al., 2024). SOC-SOM conversion will also depend on the stoichiometry and could thus be refined. Improvement on this approach was undertaken by obtaining SOC-SOM ratios from the literature for different habitats (Reinsch et al., 2025). This is consistent with the different SOC-SOM ratios for plants in the meta-analysis of (Ma et al., 2018). SOC-SOM conversion factors are provided in Table 1.

The fractions of clay, silt, and sand used in the present work were taken from the set of pan-European maps produced by (Ballabio, Panagos, & Monatanarella, 2016) from 6,140 observations of the LUCAS 2009 database. The bulk density information for the topsoil (20 cm) derived from the map of (Panagos et al., 2024) who used 6,140 points of the LUCAS 2018 database (Orgiazzi et al., 2022), and the soil organic carbon map was derived from observations of all LUCAS campaigns. The CORINE land cover dataset was adopted and reclassified into the broad categories (cropland, grassland, shrubland, forest) to which stoichiometric values were derived. All datasets were resampled to the common spatial resolution of 1 km and cropped to the 27 Member States of the European Union, plus the United Kingdom and Switzerland.

We followed a sequence of five steps to derive the topsoil porosity map: 1) Calculate organic matter particle density based on stoichiometry using Eq. (5) and determine the median for each of the land cover categories; 2) Assign each map land cover to a median value of particle density; 3) Convert soil organic carbon to soil organic matter for LUCAS data; 4) for each map pixel with bulk density, calculate the soil particle density according to Eq. (5); and 5) convert each total bulk

density pixel to $\rho_{b\,FE}$ based on a coarse fragment correction described previously, then determine the porosity of the fine earth using Eq. (1) to produce a porosity map. By using the almost 6,000 points of total bulk density from LUCAS 2018 topsoil survey and advanced machine learning methods (Cubist), Panagos et al. (2024) developed a high-resolution total bulk density map (100 m) for topsoil (0-20 cm) covering the EU, UK and Switzerland. The predicted total bulk density map values were used as the basis for the conversion of total bulk density to fine earth bulk density and then porosity.

2.4 Statistical modelling

We constructed statistical models to explore the extent to which porosity may be predicted directly from SOM, allowing the relationship to vary between land cover types. This type of statistical approach can be used to estimate porosity when data are limited and helps to build understanding of the influence of land cover on trends in the data. Models were fit using k fold cross validation, stratified by landcover type. Due to the bimodal distribution of residuals in models derived from SOM, Gaussian distribution was not appropriate. Hence, a Tweedie distribution was used with variance power *p* assigned during model fitting using the "gam" function in the R package "mgcv" (v1.8-42; (Wood, 2011)). Fitting Tweedie distribution variance power *p* value to the model should capture the distribution of residuals, which was assessed using residual plots for the models.

To test for variation in the relationship of SOM to porosity between land cover types, we constructed two separate models. In Model 1, we allowed the model to vary the gradient of the SOM to porosity relationship between land covers (using "fs" to fit a variable smooth by land cover). We compared this to a Model 2 which instead specified a consistent nonlinear relationship to SOM (using "cs" to specify a cubic spline). In both models, we also included land cover as a

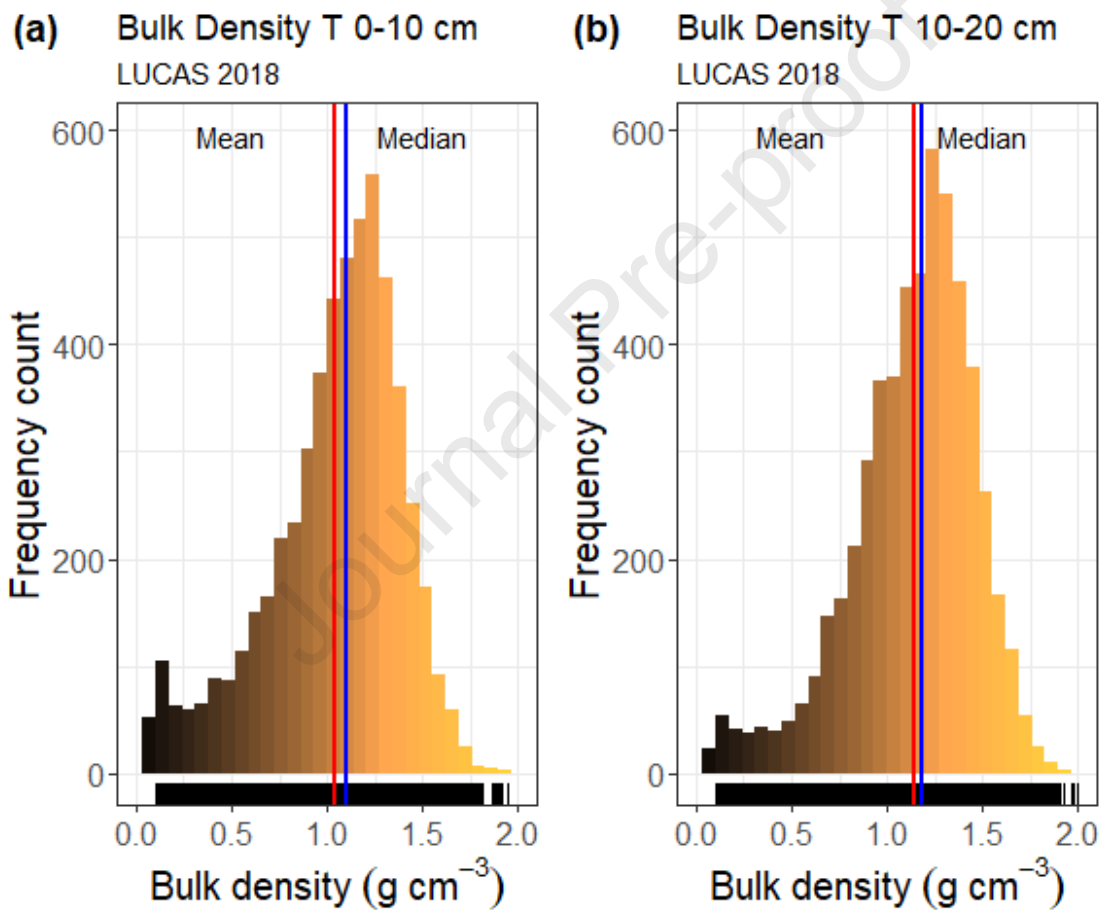
fixed effect, to account only for variation in intercept (i.e. a shift in the trend). The use of penalized smoothers in both models can capture nonlinear relationships between SOM and porosity. The "cs" smoother applies a double penalty (Marra & Wood, 2011), which allows the penalized regression routine to shrink spurious covariates out of the model. The "fs" smoother fits separate smooths by habitat and applies penalization to avoid overfitting, allowing the smooths to be shrunk toward simpler or more similar trends, unless the data strongly justifies greater complexity. Using these methods, the influence of a variable in the model may be interpreted as indicative of improving the fit (since influence of the variable would otherwise be shrunk out). The penalization approach should avoid overfitting if implemented correctly, which was assessed from comparison of estimated degress of freedom (edf) with reference degrees of freedom (ref.df) and inspection of smooth plots to look for implausible patterns. The separate inclusion of land cover as a fixed effect in both models allows us to also capture average variation in SOM porosity relationships between land cover types, rather than assuming that differences should be only related to a trend with SOM. We used a Wilcoxon signed-rank test to evaluate whether the additional flexibility in allowing the relationship to SOM to vary between landcover types significantly improves model fit. To better explore the relationship by land cover, we filtered out wetland data points due to low n (<3).

3 RESULTS

Histograms illustrating the soil total bulk density data for the LUCAS 2018 dataset are presented in Fig. 1, encompassing two depth ranges: 0-10 cm and 10-20 cm. The bulk density values span from 0.1 g cm^{-3} to 2.0 g cm^{-3}, with the 0-10 cm histogram (Fig. 1a) indicating a higher prevalence of organic material, as evidenced by the noticeable increase in data points at lower bulk densities. In contrast, this trend diminishes in the 10-20 cm dataset (Fig. 1b). Furthermore, this

286 decline in organic matter correlates with a rise in the mean bulk density across the two layers,
287 shifting from 1.04 g cm^{-3} for the 0-10 cm range to 1.14 g cm^{-3} for the 10-20 cm range (1.09 g cm^{-3}
288 for the 0-20 cm range). The conversion of bulk density from Total ($\rho_{b\,T}$) to fine earth ($\rho_{b\,FE}$) for
289 0-20cm is found in (Supplementary Fig. S1). The mean bulk density of the $\rho_{b\,FE}$ transitions from
290 1.09 cm^{-3} to 1.01 cm^{-3} after coarse fraction removal.

291



292

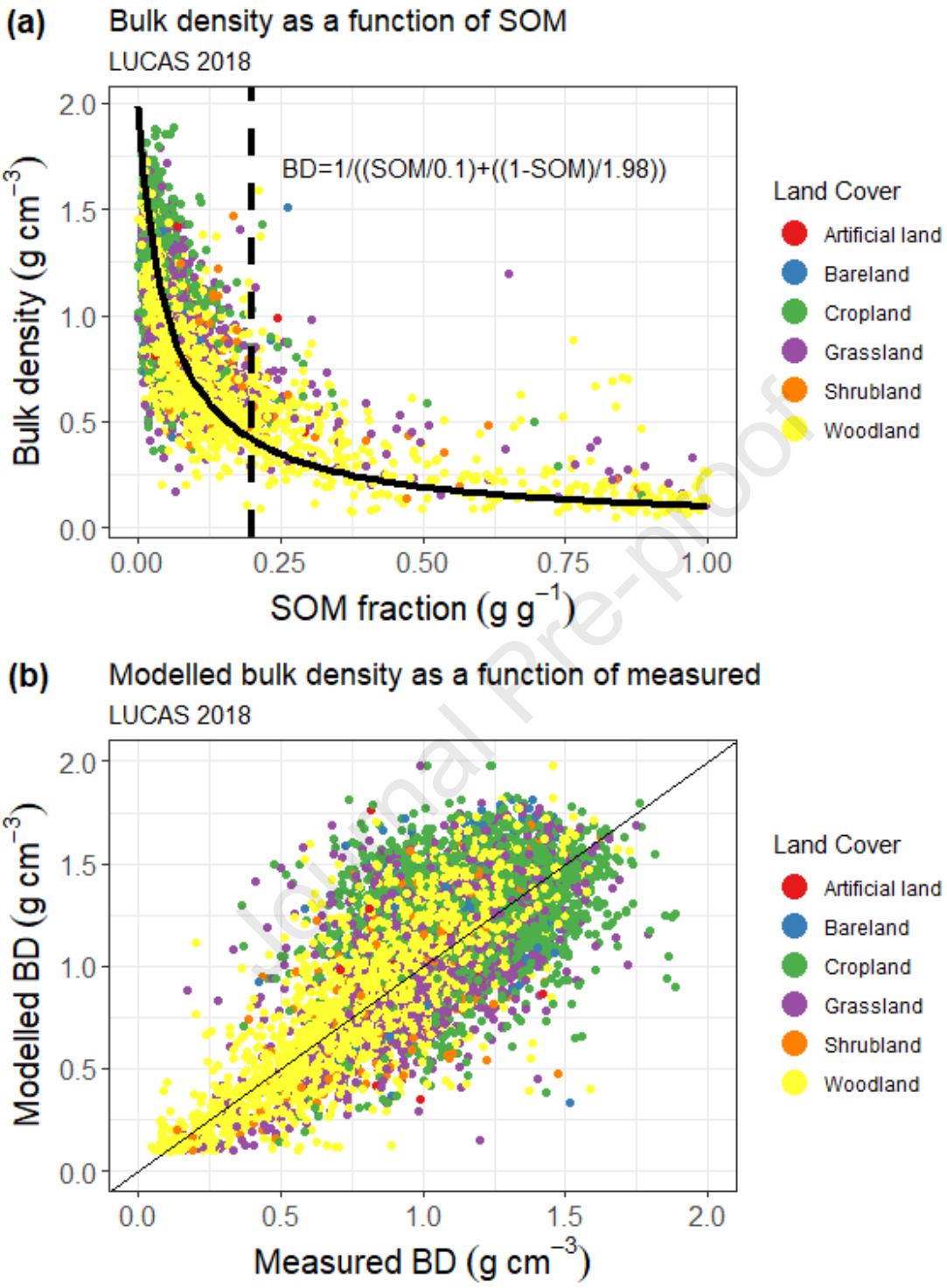
293 Figure 1. Soil total bulk density histograms for (a) 0-10 cm (n=5,518) and (b) 0-20 cm
294 (n=5,518) filtered from the LUCAS 2018 topsoil survey (n=5,659). The red line indicates the mean
295 values for the distribution (0-10cm = 1.04 g cm^{-3}; 10-20cm = 1.14; 0-20cm 1.09), which is slightly
296 lower than the median value (blue line) (0-10cm = 1.10 g cm^{-3}; 10-20cm = 1.18; 0-20cm 1.15).

Journal Pre-proof

297 The data are colored from dark to pale signifying the greater organic matter content in the low bulk
298 density soils and the higher mineral content in the high bulk density soils.

299 The bulk density data are illustrated in Fig. 2a, which depicts the relationship between soil total
300 bulk density and soil organic matter fraction. This fraction is derived from the soil organic carbon
301 concentration within the LUCAS dataset, calculated by applying the conversion factors in Table
302 1. The data points are color-coded according to their respective land covers, revealing a robust
303 relationship consistent with findings by Panagos et al. (2024) and Thomas et al. (2024).
304 Furthermore, an interpretive model is introduced, Eq. (6), with bulk density values at the extreme
305 ends set to be 0.1 g $cm^{-3}$and 1.98 g cm^{-3}. This model, grounded in physical principles (Robinson
306 et al., 2022), effectively captures the observed trends and curvature of the data. In Fig. 2b, the
307 modeled data is compared to the measured data, with a 1:1 line included for reference. The figure
308 demonstrates a relatively uniform distribution of values around the model, emphasizing the
309 prevalence of woodland in areas with low bulk density soils, while cropland and grassland are
310 associated with higher bulk density soils. Additional figures, differentiated by clay content, are
311 provided in Supplementary Fig. S2, indicating that grasslands and croplands exhibit greater
312 consistency with mineral soils.
313

Journal Pre-proof



314

315 Figure 2 Fine earth bulk density relationships for major EU land covers. (a) Bulk density (0-20 cm

316 FE) as a function of soil organic matter for the LUCAS 2018 topsoil dataset. (b) Model predicted

bulk density FE, Eq. (6) vs. the measured bulk density FE. Artificial land represents urban brown field sites for example while bare land is that without vegetation more generally.

The bulk density values categorized by land cover type are detailed in Table 1, which also includes the soil clay fraction. Table 1 also contains predictions discussed later in the context of Fig. 5. These estimated values are derived from modeled bulk density as per Eq. (4), with the soil organic matter bulk density end members set at 0.1 g cm^{-3} and the mineral soil bulk density at 1.98 g cm^{-3} (Robinson et al., 2022). The calculation of soil particle density involved the integration of mineral and organic particle densities, following the methodology outlined in Eq. (4). The mineral particle density was obtained from the particle densities of the clay, silt, and sand fractions, as reported by (Ruehlmann & Körschens, 2020), utilizing a comprehensive global dataset (2.76, 2.69, and 2.66 g cm^{-3}), while the soil organic matter particle density was ascertained from the current study.

Table 1. Soil metrics, measured and estimated either from the model predictions, Eq. (2), (3) & (6) for six land cover types. Values represent the means with the standard deviation following in brackets. Estimated values, the bulk density is calculated using Eq. (3). The soil particle density with Eq. (2) and the porosity Eq. (1) using the modelled particle density to convert the bulk density of the fine earth. Numbers in parentheses are the standard deviations.

Metric	Bareland	Cropland	Grassland	Shrubland	Woodland
Bulk Density (ρ_b) (T) measured g cm^{-3}	1.23 (0.199)	1.25 (0.212)	1.10 (0.270)	1.05 (0.316)	0.83 (0.353)
Bulk Density (ρ_b) (FE) measured g cm^{-3}	1.11 (0.210)	1.17 (0.232)	1.00 (0.283)	0.90 (0.290)	0.74 (0.338)
Clay %	24.1 (11.1)	23.7 (13.9)	21.1 (13.6)	18.1 (12.9)	11.5 (10.7)
Estimated values					

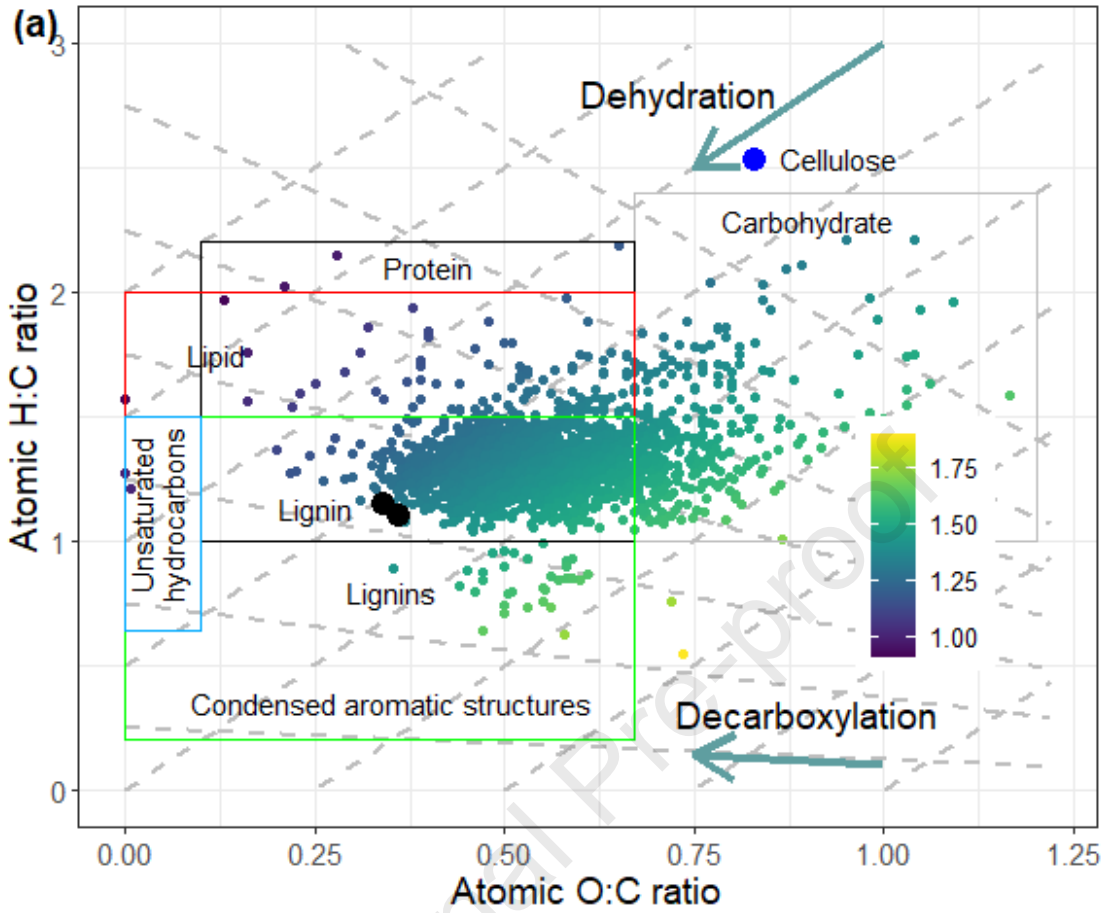
Soil Particle Density (ρ_{ss}) modelled, Eq. (6) & (2) (g cm^{-3})	2.62 (0.05)	2.61 (0.07)	2.53 (0.16)	2.50 (0.20)	2.40 (0.29)
Porosity (FE) – modelled, Eq. (6) & (4) (cm^3 cm^{-3})	0.576 (0.080)	0.553 (0.087)	0.609 (0.102)	0.646 (0.104)	0.702 (0.122)
Porosity (FE) assuming ρ_M 2.65 (cm^3 cm^{-3})	0.581 (0.079)	0.560 (0.088)	0.624 (0.107)	0.662 (0.110)	0.721 (0.127)
Number obs.	200	2328	1167	181	1617

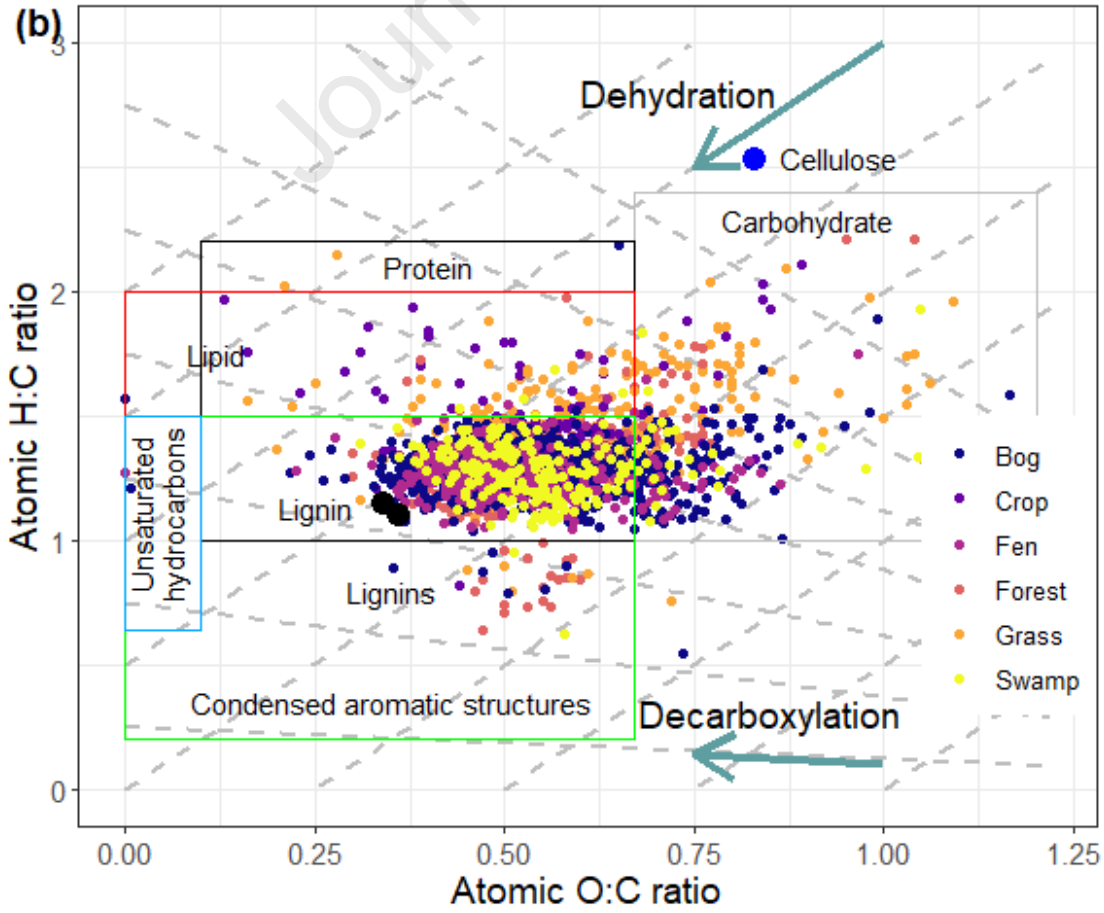
335

336 Particle density was calculated based on Eq. (6) using the data of Moore et al. (2018) and
337 Leifeld et al. (2020) with the results presented in the form of a Van Krevelen diagram (Fig. 3) that
338 plots the H:C versus O:C ratios (van Krevelen, 1950). This diagram illustrates the anticipated
339 positioning of organic materials based on their stoichiometric ratios, thereby capturing the
340 potential relationships and transitions that SOM may experience. In Fig. 3a, the data points are
341 color-coded according to their predicted density. The predicted density is consistent with a gradient
342 ranging from lightweight, lipid-dominated compounds (such as stearic acid at 0.94) to denser
343 substances like lignin (1.3) and carbohydrates (cellulose at 1.5). The densest SOM is linked to the
344 condensed aromatic structures depicted (Fig. 3a). The dashed lines, marked with arrows, represent
345 the pathways of dehydration and decarboxylation. Therefore, Eq. (6) predicts particle density
346 values consistent with where we would expect them to fall on the van Krevelen plot. Fig. 3b
347 presents the same diagram, but the data points are categorized by habitat. The distribution of these
348 data points suggests potential clustering among habitats, indicating that compounds of specific
349 densities are more prevalent in certain environments. Lighter organic matter is predominantly
350 associated with nutrient-rich habitats such as cropland and grassland, while denser organic matter
351 is more commonly found in woodlands and peatlands.

352

Journal Pre-proof





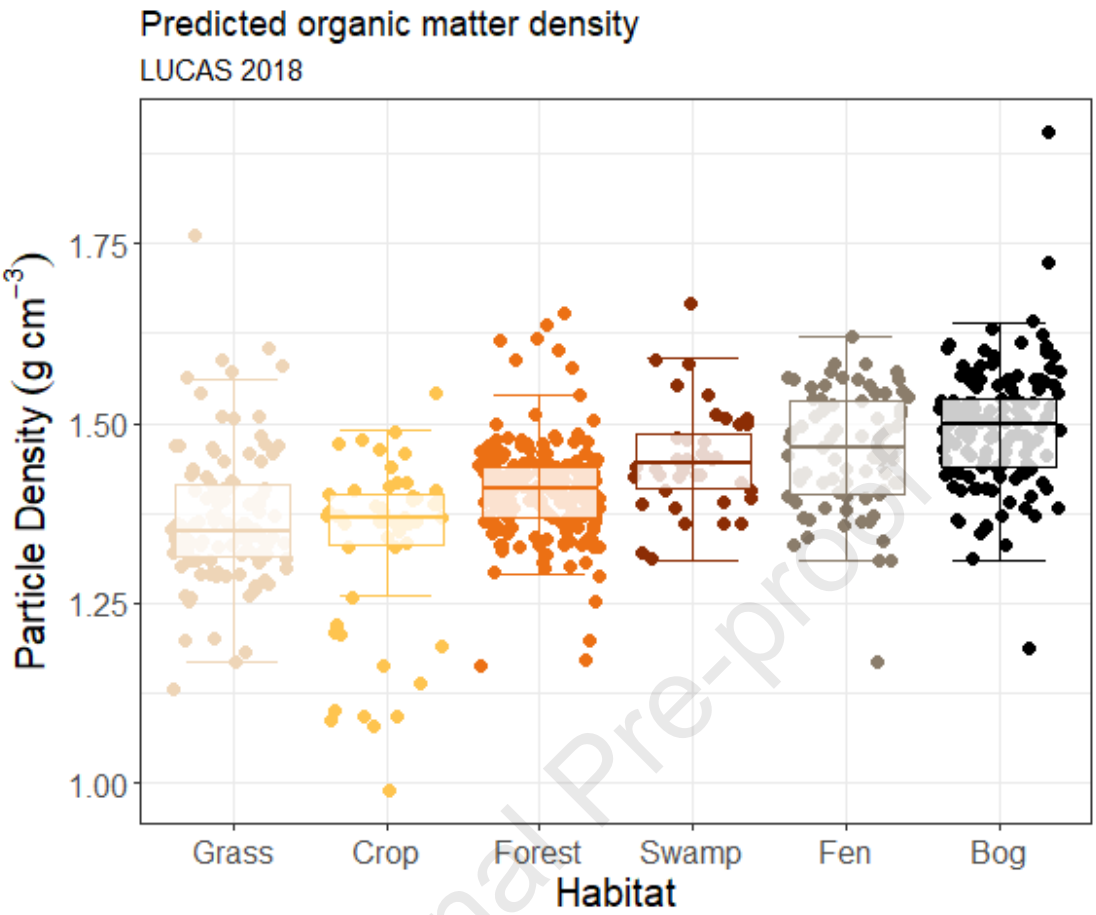
Journal Pre-proof

353

Figure 3. (a) Organic matter particle density (g cm^{-3}) calculated by using the atomic hydrogen to carbon (H:C) and oxygen to carbon (O:C) ratios from (Leifeld et al., 2020) and (Moore et al., 2018) and plotted on the Van Krevelen diagram. The legend is particle density (g cm^{-3}) (b) The same diagram but colored by habitat

The predicted particle density values (0-20cm) for different habitats are presented in Fig. 4 and in Supplementary Table S1; these are for organic soils, where fens, are fed by streams and rivers; bogs, fed by rainwater; and swamps, distinguished by the presence of trees and shrubs. The mean particle density ranged from 1.35 g cm^{-3} in grassland soils to 1.50 g cm^{-3} in nutrient-poor peatland for 0-20cm; this narrowed to 1.35 to 1.41 if data from the whole profiles was included (Table S2). This change perhaps reflecting a decrease in O_2 with depth. The data, represented by median values in Fig. 4, suggest a potential gradient in density from nutrient-rich to nutrient-poor habitats.

Journal Pre-proof



366

367 Figure 4. Organic matter density for differing habitats predicted according to Eq. (6) for the
368 stoichiometric data sets in Leifeld et al. (2020) and Moore et al. (2018) (0-20cm).

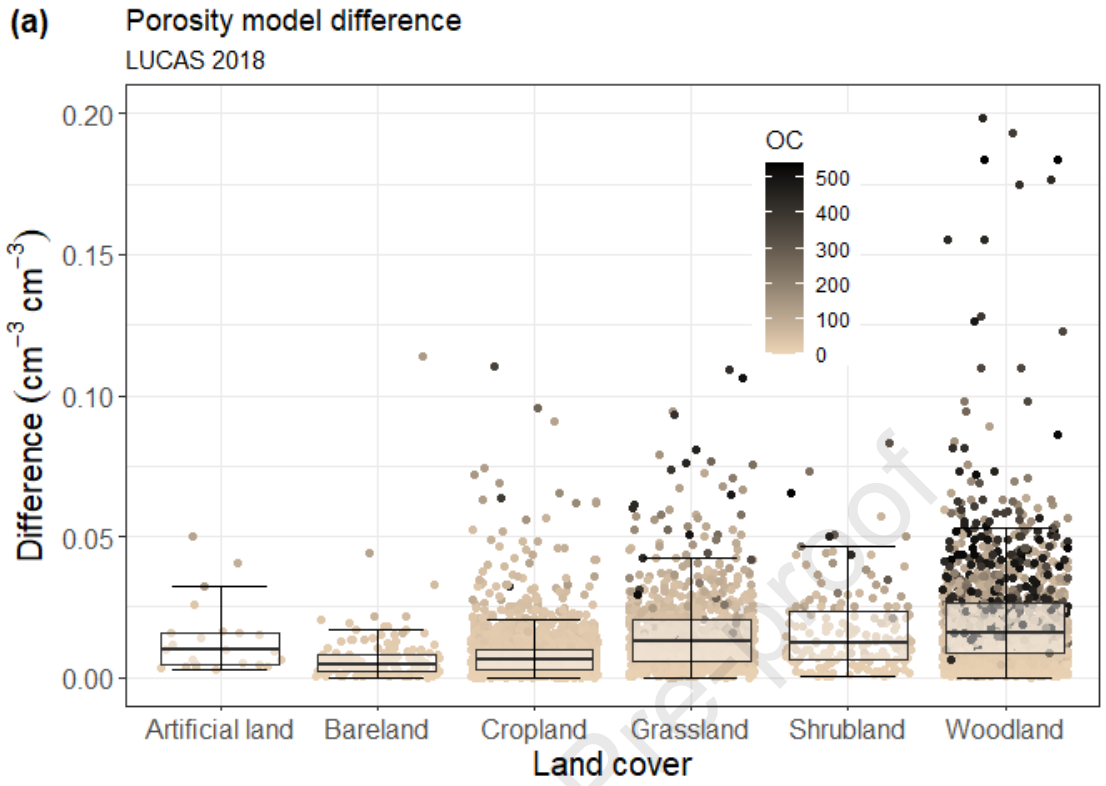
369

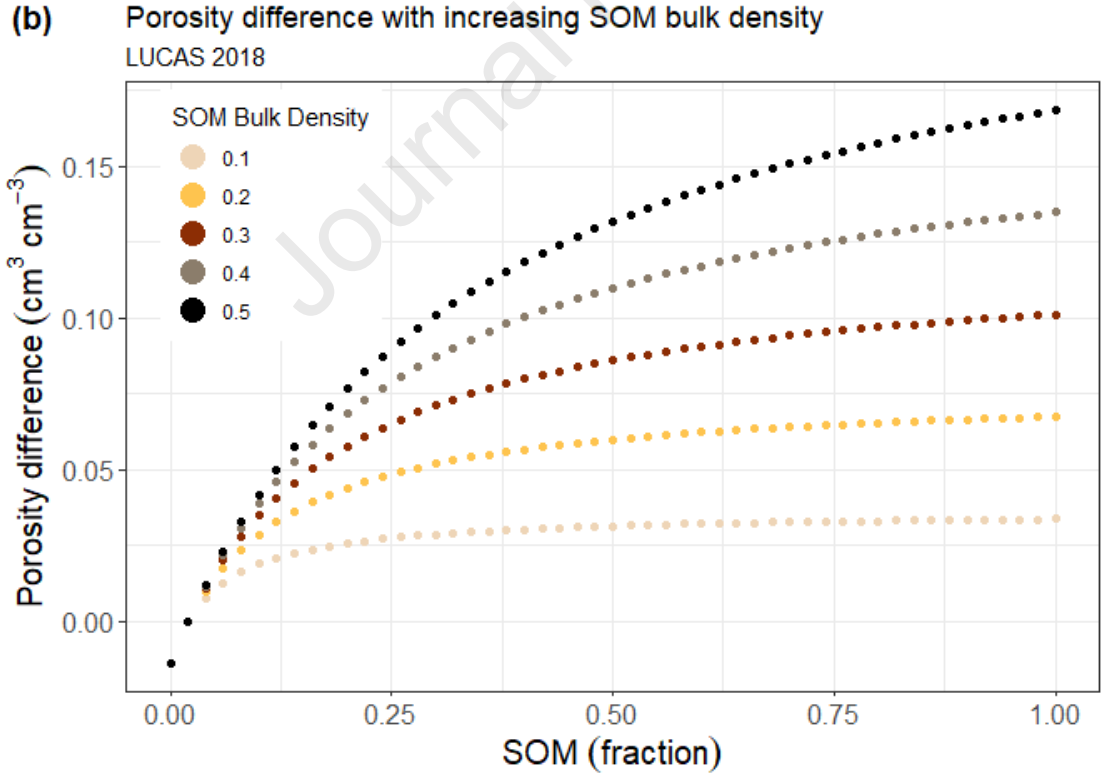
370 Particle densities presented in (Supplementary Table S1) were utilized to ascertain the soil
371 particle density (Table 1), which subsequently facilitated the calculation of soil porosity in
372 accordance with Eq. (1). The mean porosity values derived from the novel methodology introduced
373 in this study are displayed in Table 1, alongside those obtained using a conventional particle
374 density value of 2.65 g cm^{-3}, commonly applied in mineral soil conversions, for comparative
375 purposes (see Fig. 5). Fig. 5a shows the difference in porosity between an assumption of 2.65 and
376 the calculated particle density based on the mineral and organic fraction weighting. The findings

Journal Pre-proof

from the new model suggest that when compared to the assumption of a particle density of 2.65, the porosities calculated using a combination of organic and mineral particle densities are lower as expected, within 0.05 m^3m^{-3}, due to the lower density of organic matter. The coloring shows that soils where the difference is greater than 0.05 tend to have more organic matter. This is explored in Fig. 5b using Eq. (4) to calculate that happens when the bulk density of the organic fraction is increased. The figure clearly shows that the absolute error in terms of porosity is small (~0.03) when the bulk density of the organic material is low, as there is little of the organic material resulting in a small error. However, as expected, as the amount of organic material increases so the absolute porosity difference increases substantially such that the use of 2.65 substantially overestimates porosity. This difference in organic soils with bulk densities of 0.5 g cm^{-3}, represents an over estimation of more than 0.15 cm^3 cm^{-3} when 2.65 is assumed, which is beyond the standard deviation of the combined LUCAS data set porosity (SD = 0.119).

Journal Pre-proof





Journal Pre-proof

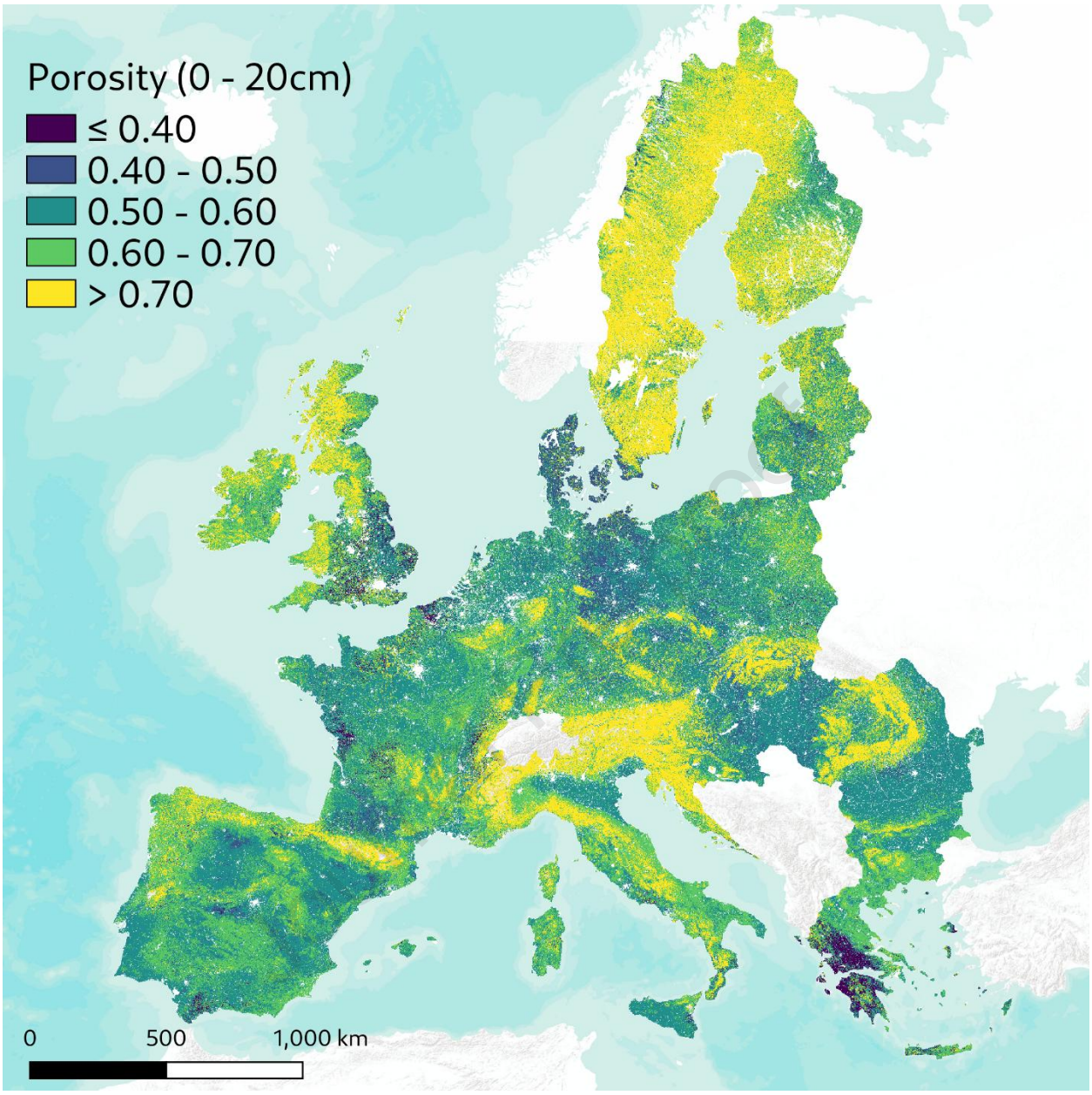
391

392

Figure 5. (a) Boxplots of the difference in porosity (ρ_{ss}=2.65 – ρ_{ss}=calculated) for different LUCAS land covers. Particle densities calculated with, Eq. (6), (a generic value of 2.65 g cm^{-3}, often used for mineral soils (Brady & Weil, 2008)). (b) The porosity (FE) difference calculated with ρ_{ss}=2.65-ρ_{ss}=1.4, simulated for the SOM range with Eq. (4) ρ_b mineral = 1.98 ρ_b SOM adjusted.

The spatial extension of these findings, utilizing the maps produced for bulk density (Panagos et al., 2024) and subsequently converting to porosity through the application of mineral and organic matter particle densities, culminates in the European map depicted in Fig. 6 (0-20 cm). The regions marked in yellow in Northern Europe correspond to soils rich in organic matter, while those in Central Europe align with extensive areas of forested soils. Conversely, the soils represented in blue, characterized by low porosity, are typically associated with agricultural croplands.

Journal Pre-proof



406

407 Figure 6. Soil porosity (FE) (cm^3 cm^{-3}) for the topsoil (0-20 cm) for Europe and UK based on the

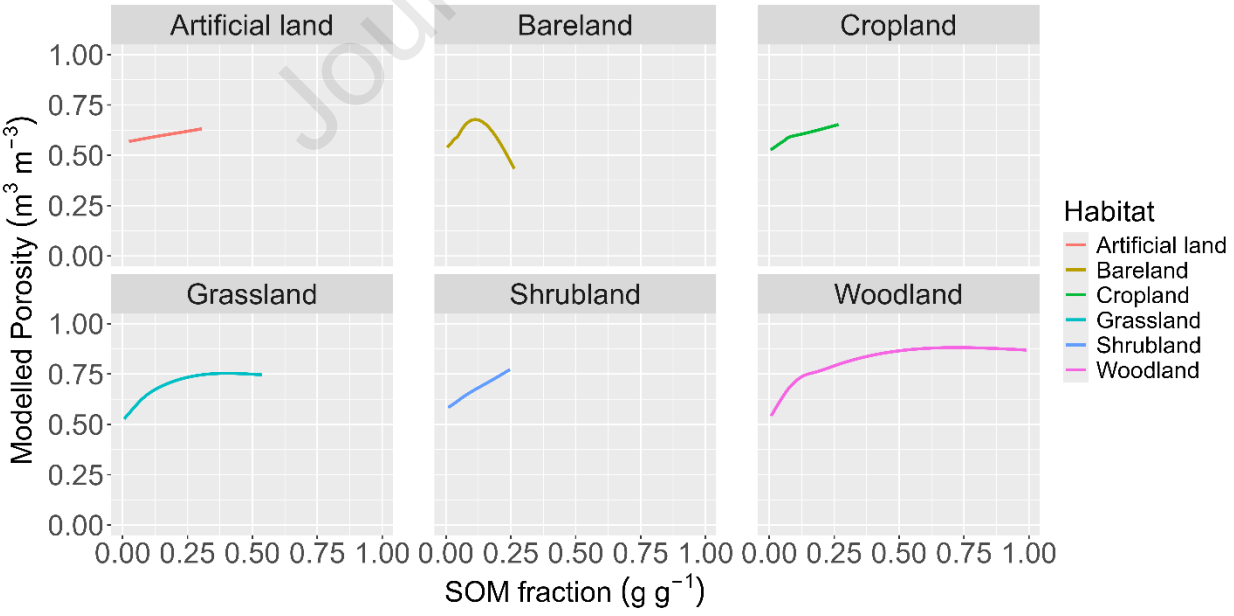
408 conversion of bulk density to porosity using the particle density determination approach introduced

409 in this work.

410

411 Fig. 7 examines the relationship between soil organic matter and porosity across various
412 vegetation covers. A Wilcoxon signed-rank test showed that this model which allowed the SOM
413 trend to vary by landcover provided significantly better fit ($p<0.001$, see Table S4) than a model
414 2 without landcover specific relationship to SOM: Porosity ~ s(SOM Fraction, bs="cs") +
415 landcover. The observed gradients indicate that vegetation cover significantly influences the SOM-
416 porosity relationship. Notably, the relationships observed in artificial and bare land differ markedly
417 from those in vegetated habitats, highlighting the distinct biotic drivers and processes at play. In
418 vegetated habitats, the overall trend exhibited less variability. Grassland, shrubland, and woodland
419 display the characteristic curvature of the empirical model, Eq. (4); (Fig. S4), while cropland does
420 not have the initial steep rise, suggesting that porosity increases less with lower levels of soil
421 organic matter compared to semi-natural habitats. This phenomenon may be indicative of a higher
422 degree of settlement and potential compaction associated with this land cover.

423



424

Journal Pre-proof

Figure 7. Plots of predicted porosity using a Generalized Additive Model (GAM) for LUCAS 0-20cm datapoints using model 1: Porosity ~ s(SOM Fraction, by = landcover, bs="fs") + landcover. The relationship to soil organic matter (SOM) was allowed to vary by land cover, and model statistics indicate variation in porosity between land cover types, as well as nonlinear trends with SOM for all landcover types except artificial (see Table S5; Fig. S5). Model constructed using k-fold cross validation; each point in the data used for plotting was predicted by a model which excluded it.

4 DISCUSSION

The bulk density data obtained from the LUCAS topsoil survey constitutes an internally consistent dataset that is essential for comprehending the condition and transformation of soils throughout Europe. Bulk density serves as a critical indicator of soil health and has been demonstrated by Seaton et al. (2021) to be closely associated with concentrations of carbon and nitrogen at a national scale. Recent advancements in understanding the physical relationship between bulk density and soil organic matter have been articulated by Robinson et al. (2022) and Thomas et al. (2024). This enhanced comprehension suggests that the significance lies not only in the mere presence of soil organic matter and its differing particle density compared to mineral materials, but also in the morphology of soil organic matter particles, as noted by (Robinson et al., 2022). This principle similarly applies to porosity. The data distinctly illustrate a robust correlation with land cover or habitat type, as highlighted by (Panagos et al., 2024) and Thomas et al. (2024) for porosity. Furthermore, Thomas et al. (2024) has indicated that for porosity this relationship

447 extends beyond soil organic matter and is further refined by the inclusion of habitat variables, as
448 evidenced in Models 14 and 15 of their study. We found even greater influence of land cover, to
449 the extent that the gradient of the relationship between soil organic matter and porosity for the
450 LUCAS data varies between land cover, whilst Thomas et al. (2024) found only a shift in the trend
451 for UK data, probably due to the smaller area studied. The additional influence of land cover is
452 likely attributable to plant roots and pores created by other organisms altering the geometry of the
453 soil beyond that of soil organic matter alone.

454 The mean particle density of organic matter was observed to range from 1.35 in grassland to
455 1.50 in bog, according to this methodology (Supplementary Table S1). We concur with Ruehlmann
456 (2020); (Rühlmann et al., 2006) that a distinct gradient exists in soil organic matter density from
457 low to high. Additionally, we hypothesize that this gradient is likely attributable to the quality of
458 SOM. This corresponds to Fig 3 that indicates that as the H:C ratio alters, to more carbon and less
459 hydrogen, the particle density increases as expected. Hence, as carbon concentrates in Bog's, the
460 SOM becomes denser. Recent investigations by Ma et al. (2018) have examined plant components
461 globally for their organic carbon content, revealing notable patterns; for instance, organic carbon
462 content varies from 0.382 in the roots of cropland plants to 0.474 in woody plant roots, which we
463 believe undergo transformation and are subsequently represented in the soil organic matter
464 stoichiometry. A comprehensive analysis conducted by (Reinsch et al., 2025) across European
465 soils corroborated these findings regarding SOM. Therefore, we align with Rühlmann et al. (2006)
466 in asserting that the observed gradient in soil organic matter density likely reflects variations in
467 organic matter quality and the SOC content.

468 An examination of Fig. 4b indicates that cropland soils are characterized by a higher presence
469 of lighter compounds, such as lipids and proteins, and a lower presence of lignin, in contrast to

Journal Pre-proof

woody habitats that exhibit increased lignin content. (Ma et al., 2018) research indicated a positive correlation between carbon content and lignin, while no such correlation was found with cellulose. This suggests that the concentrations of lignin and lipid proteins within soil organic matter are significant determinants of overall organic matter particle density. Furthermore, we endorse the notion that this gradient aligns with the "soil continuum model" conceptual framework proposed by (Lehmann & Kleber, 2015); who suggest a gradient from particulate organic matter into smaller and smaller molecules as the SOM breaks down.

The integration of soil texture for assessing mineral density alongside average land cover values for organic matter particle density, Eq. (6) facilitates a more refined prediction of porosity for large scales based on Eq. (1). This approach is predicated on a detailed estimation of soil particle density, Eq. (6), utilizing metrics that are readily accessible through soil monitoring and remote sensing technologies. As illustrated in Fig. 5, the predictions of porosity derived from this method differ from those obtained by employing a standard factor, such as 2.65 g cm^{-3}, for particle density. It is anticipated that a mineral soil particle density conversion factor would yield suboptimal results in woodland and peatland environments where organic soils are prevalent. Nevertheless, it is noteworthy that the predictions also fall short in cropland and bare soil contexts, where one might reasonably assume a particle density reflective of quartz. Analysis of the LUCAS dataset indicates that the average value for particle density for all EU soils used to convert bulk density to porosity, Eq. (1), in the absence of texture and organic matter data, is 2.53 g cm^{-3}. However, it’s better to use the land use specific values in Table 1.

The porosity map for Europe (Fig. 6 (0-20 cm) illustrates a significant anthropogenic influence on soil porosity throughout the EU and UK. Those datasets at 100m resolution will be available in the European Soil Data Centre (ESDAC) (Panagos et al., 2022). The generated porosity map aligns

493 closely with the soil organic carbon distribution across Europe, highlighting the critical role of soil
494 organic carbon in this relationship. In Southern Europe, soils exhibit low levels of SOM, primarily
495 attributed to rising pH levels and a transition from organic to inorganic carbon stability.
496 Consequently, the inherent low SOM content in these soils is reflected in their reduced porosity.
497 While land cover plays a crucial role in determining porosity, latitude also influences organic
498 carbon distribution across the continent, further affecting porosity levels. This can be observed
499 when upscaling to global scales, such as when looking at products from soilgrids (Hengl et al.,
500 2017) or openland map (Hengl et al., 2025) for bulk density. When observing such global data, it
501 is apparent that a gradient exists with lower bulk densities in the peatlands of the North and higher
502 bulk densities in hot dry places like deserts. SOM and parent material will each contribute to the
503 expression of porosity, but this generalization is supported by analysis of global data based on
504 mean annual precipitation and temperature, where soils in hot dry climates tend to have higher
505 bulk densities than those that follow a gradient from cold and dry to hot and humid (Zhao, Yang,
506 Shen, Geng, & Fang, 2019). Global soil mapping is yet to incorporate temporal change in the way
507 Hirmas et al. (2018) observed on years to decade time scales across the USA. Moreover, changes
508 in porosity are also likely to be observed in some latitudes due to seasonality with both land
509 management and moisture and temperature cycles altering the porosity (Alletto & Coquet, 2009;
510 Hu, Shao, & Si, 2012; Nottingham, Thompson, Turk, Li, & Connolly, 2015).

511 Europe spans the latitudes where soils are very porous in the Northern peatlands to where soils
512 are increasingly dense close to the equator, or in deserts with no organic matter; this represents a
513 span in porosity from ~0.95 to 0.4 m^3m^{-3}. In such locations without SOM and only granular
514 material bulk densities revert to values around 1.6 g cm^{-3} corresponding to porosities of ~0.4 m^3m^{-3}.
515 The results imply that the greatest changes to porosity will likely be observed initially in the

temperate latitudes where changes in soil organic matter due to land use or climate change are likely to be more pronounced.

While this work has focused on mechanistic modelling, increasingly data driven modelling using AI is being applied in soil science (Minasny & McBratney, 2025; Wadoux, 2025). As porosity is not measured directly AI and machine learning approaches have focused on bulk density (Chen et al., 2024; Hengl et al., 2025; Hengl et al., 2017; Panagos et al., 2024). These black box methods continue to improve (Minasny & McBratney, 2025). One of the challenges for soil mapping is presented by biased data sets (Liu, Ikonnikova, Scott Hamlin, Sivila, & Pyrcz, 2021) or bimodal data (Nussbaum, Zimmermann, Walthert, & Baltensweiler, 2023). Traditional soil sampling, for example based on a grid, will often produce machine learning predictions that have the tails of the distribution poorly represented. The tails are often important in soil science as they may represent less spatially abundant but functionally important soils like peats. One way of dealing with this is using stratified sampling which results in more balance data sets. Alternatively, methods continue to develop to deal with such data e.g. selecting the best approach for rebalancing data (Dal Pozzolo, Caelen, Waterschoot, & Bontempi, 2013). These methods continue to evolve and as data becomes more abundant will significantly improve the prediction of soil change in space and time across scales, but caution and care should be applied in the selection of covariates to improve soil knowledge and understanding (Wadoux, Samuel-Rosa, Poggio, & Mulder, 2020). These issues are not unique to AI modelling, for example, statistical models like the GAM presented here must also be applied with consideration for unbalanced data-sets e.g. through careful selection of response distribution (Wood, 2017). Similarly, data-driven statistical modelling implemented with poorly selected co-variates may identify spurious relationships which

do not reflect underlying processes or drivers and may not hold outside of the sample dataset (Tredennick, Hooker, Ellner, & Adler, 2021).

Total porosity represents an initial aspect of soil characterization; however, comprehending variations in size distribution is a crucial subsequent consideration. Macropores play a significant role in facilitating swift infiltration (Beven & Germann, 1982; Watson & Luxmoore, 1986), thereby minimizing surface water accumulation and mitigating the risks associated with runoff, localized flooding, and soil erosion. Recent investigations conducted throughout the continental United States have revealed that soil macroporosity is undergoing changes over decadal periods (Hirmas et al., 2018), over and above rapid field scale changes due to agriculture. Their findings suggest that "drier and warmer climates foster the formation of macropores in the surface layer, while more humid and cooler climates inhibit the manifestation of macroporosity." While macropores facilitate rapid drainage, micropores serve to retain water against gravitational forces, thus supplying moisture to plants and soil microorganisms. This dual porosity characteristic of soils is vital for sustaining biological functions. Any reduction in total porosity will have significant implications for hydrological processes, the extent of which will be influenced by the distribution of pore sizes.

Many sustainable land management practices that promote soil health play a significant role in the restoration, maintenance, and enhancement of total soil porosity, as well as the distribution and connectivity of soil pores. These practices encompass the application of organic amendments, and the use of cover crops and reduced or no tillage (Bai et al., 2019). It is essential to acknowledge that soil pore space is inherently dynamic and requires careful management to ensure the health of agricultural soils. This dynamic is influenced by various processes, including wetting and drying cycles, shrink-swell phenomena, and soil aggregation (Ghezzehei, 2012). Additionally, biotic

561 factors, such as plant root systems and ecosystem engineers like earthworms, ants, and termites,
562 contribute to the dynamic characteristics of soil porosity (Kodešová, Kodeš, Žigová, & Šimůnek,
563 2006). This evolving nature of soil porosity and its implications for structural development is
564 gaining attention, particularly concerning its potential interactions with the hydrological cycle and
565 climate change (Fatichi et al., 2020; Hirmas et al., 2018).

566 The anthropogenic impact, particularly in central and northern Europe, suggests a potential
567 increase in soil compaction, which is likely associated with a decrease in SOM resulting from
568 agricultural practices that has been observed both in measurement (Reynolds et al., 2013) and
569 modelling (Janes-Bassett et al., 2021). Furthermore, the relationship between bulk density or
570 porosity and SOM implies that significant feedback mechanisms may arise concerning local soil
571 hydrology in response to changes in land use and soil organic carbon management. Alterations in
572 porosity are expected to occur over various temporal scales. Seasonal variations in porosity may
573 arise due to cycles of wetting and drying, while longer-term changes may be observed over decades
574 because of shifts in land use (De Rosa et al., 2024; Or, Keller, & Schlesinger, 2021). In the United
575 States, changes in porosity have been documented over similar time frames, although the
576 underlying causes remain unclear (Hirmas et al., 2018). Datasets like the LUCAS for European
577 topsoils, which is expanding its monitoring of bulk density, offer a robust foundation for
578 monitoring the magnitude and extent of soil conditions and changes as the survey progresses over
579 time. This will enhance our comprehension of the dynamic characteristics of soil porosity at large
580 scales and its responses to various factors, including seasonal variations, climate influences, and
581 land use modifications.

582

583

Journal Pre-proof

Journal Pre-proof

584

585

586

587

Journal Pre-proof

Journal Pre-proof

ACKNOWLEDGMENTS

Disclaimer: Work funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Funding: The research was funded by, The Research Council of Norway, Climasoil, Project number: 325253. Natural Environment Research Council award to UKCEH for the National Capability for UK Challenges Programme NE/Y006208/1. This project acknowledges funding from the European Union's Horizon Europe research and innovation program under grant agreement No.-101086179; and funding from the UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee [grant number 10053484] as part of AI4SoilHealth. The work was a collaborative effort within the Collaboration Agreement No 36667 between JRC and the AI4SoilHealth Soil Mission project.

CONFLICT OF INTEREST

"The authors declare no conflict of interest".

ORCID

DAR: 0000-0001-7290-4867

Author Contributions Statement: DAR handled the conceptualization and writing – original draft. Formal analysis was conducted by DAR the main analysis, AF produced the maps and AT conducted the GAM modelling. JL, TM and JS were responsible for providing datasets with organic soils. All were responsible for review & editing. DAR handled the funding acquisition.

Declaration of generative AI and AI-assisted technologies in the writing process: During the preparation of this work the author used ahrefs sentence rewriting tool in order to improve the

consistency of the language across multiple international authors. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

REFERENCES

Adams, W. (1973). The effect of organic matter on the bulk and true densities of some uncultivated podzolic soils. *Journal of Soil Science, 24*(1), 10-17.

Alletto, L., & Coquet, Y. (2009). Temporal and spatial variability of soil bulk density and near-saturated hydraulic conductivity under two contrasted tillage management systems. *Geoderma, 152*(1-2), 85-94.

Bai, X., Huang, Y., Ren, W., Coyne, M., Jacinthe, P. A., Tao, B., et al. (2019). Responses of soil carbon sequestration to climate-smart agriculture practices: A meta-analysis. *Global change biology, 25*(8), 2591-2606.

Ballabio, C., Panagos, P., & Monatanarella, L. (2016). Mapping topsoil physical properties at European scale using the LUCAS database. *Geoderma, 261*, 110-123.

Beven, K., & Germann, P. (1982). Macropores and water flow in soils. *Water resources research, 18*(5), 1311-1325.

Brady, N. C., & Weil, R. R. (2008). *The nature and properties of soils* (Vol. 13): Prentice Hall Upper Saddle River, NJ.

Chen, S., Chen, Z., Zhang, X., Luo, Z., Schillaci, C., Arrouays, D., et al. (2024). European topsoil bulk density and organic carbon stock database (0–20 cm) using machine-learning-based pedotransfer functions. *Earth System Science Data, 16*(5), 2367-2383.

Dal Pozzolo, A., Caelen, O., Waterschoot, S., & Bontempi, G. (2013). *Racing for unbalanced methods selection.* Paper presented at the International conference on intelligent data engineering and automated learning.

De Rosa, D., Ballabio, C., Lugato, E., Fasiolo, M., Jones, A., & Panagos, P. (2024). Soil organic carbon stocks in European croplands and grasslands: How much have we lost in the past decade? *Global Change Biology, 30*(1), e16992.

Dippenaar, M. A. (2014). Porosity reviewed: quantitative multi-disciplinary understanding, recent advances and applications in vadose zone hydrology. *Geotechnical and Geological Engineering, 32*(1), 1-19.

Fatichi, S., Or, D., Walko, R., Vereecken, H., Young, M. H., Ghezzehei, T. A., et al. (2020). Soil structure is an important omission in Earth System Models. *Nature communications, 11*(1), 1-11.

Ghezzehei, T. A. (2012). Soil structure. *Handbook of soil science, 2*, 1-17.

Gupta, S., Borrelli, P., Panagos, P., & Alewell, C. (2024). An advanced global soil erodibility (K) assessment including the effects of saturated hydraulic conductivity. *Science of the Total Environment, 908*, 168249.

Hengl, T., Consoli, D., Tian, X., Nauman, T. W., Nussbaum, M., Isik, M. S., et al. (2025). OpenLandMap-soildb: global soil information at 30 m spatial resolution for 2000–2022+

based on spatiotemporal Machine Learning and harmonized legacy soil samples and observations. *Earth System Science Data Discussions, 2025*, 1-66.

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., et al. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLoS one, 12*(2), e0169748.

Hirmas, D. R., Giménez, D., Nemes, A., Kerry, R., Brunsell, N. A., & Wilson, C. J. (2018). Climate-induced changes in continental-scale soil macroporosity may intensify water cycle. *Nature, 561*(7721), 100.

Holz, D. J., Williard, K. W., Edwards, P. J., & Schoonover, J. E. (2015). Soil erosion in humid regions: A review. *Journal of Contemporary Water Research & Education, 154*(1), 48-59.

Hu, W., Shao, M., & Si, B. (2012). Seasonal changes in surface bulk density and saturated hydraulic conductivity of natural landscapes. *European Journal of Soil Science, 63*(6), 820-830.

Janes-Bassett, V., Bassett, R., Rowe, E. C., Tipping, E., Yumashev, D., & Davies, J. (2021). Changes in carbon storage since the pre-industrial era: A national scale analysis. *Anthropocene, 34*, 100289.

Kang, W., Zhang, Y., Wu, S., & Zhao, W. (2024). Spatial distribution of theoretical soil macropores on a continental scale and its eco-hydrological significance in China. *Journal of Soils and Sediments, 24*(2), 563-574.

Kodešová, R., Kodeš, V., Žigová, A., & Šimůnek, J. (2006). Impact of plant roots and soil organisms on soil micromorphology and hydraulic properties. *Biologia, 61*, S339-S343.

Kuwata, M., Zorn, S. R., & Martin, S. T. (2012). Using elemental ratios to predict the density of organic material composed of carbon, hydrogen, and oxygen. *Environmental science & technology, 46*(2), 787-794.

Lebron, I., Cooper, D. M., Brentegani, M. A., Bentley, L. A., Dos Santos Pereira, G., Keenan, P., et al. (2024). Soil carbon determination for long-term monitoring revisited using thermo-gravimetric analysis. *Vadose Zone Journal, 23*(1), e20300.

Lehmann, J., & Kleber, M. (2015). The contentious nature of soil organic matter. *Nature, 528*(7580), 60-68.

Leifeld, J., Klein, K., & Wüst-Galley, C. (2020). Soil organic matter stoichiometry as indicator for peatland degradation. *Scientific reports, 10*(1), 7634.

Liu, W., Ikonnikova, S., Scott Hamlin, H., Sivila, L., & Pyrcz, M. J. (2021). Demonstration and mitigation of spatial sampling bias for machine-learning predictions. *SPE Reservoir Evaluation & Engineering, 24*(01), 262-274.

Ma, S., He, F., Tian, D., Zou, D., Yan, Z., Yang, Y., et al. (2018). Variations and determinants of carbon content in plants: a global synthesis. *Biogeosciences, 15*(3), 693-702.

Marra, G., & Wood, S. N. (2011). Practical variable selection for generalized additive models. *Computational Statistics & Data Analysis, 55*(7), 2372-2387.

Minasny, B., & McBratney, A. B. (2025). Machine Learning and Artificial Intelligence Applications in Soil Science. *European Journal of Soil Science, 76*(2), e70093.

Moore, T. R., Large, D., Talbot, J., Wang, M., & Riley, J. L. (2018). The stoichiometry of carbon, hydrogen, and oxygen in peat. *Journal of Geophysical Research: Biogeosciences, 123*(10), 3101-3110.

Nimmo, J. R. (2004). Porosity and pore size distribution. *Encyclopedia of Soils in the Environment, 3*(1), 295-303.

Nottingham, A. C., Thompson, J. A., Turk, P. J., Li, Q., & Connolly, S. J. (2015). Seasonal dynamics of surface soil bulk density in a forested catchment. *Soil Science Society of America Journal, 79*(4), 1163-1168.

Nussbaum, M., Zimmermann, S., Walthert, L., & Baltensweiler, A. (2023). Benefits of hierarchical predictions for digital soil mapping—An approach to map bimodal soil pH. *Geoderma, 437*, 116579.

Or, D., Keller, T., & Schlesinger, W. H. (2021). Natural and managed soil structure: On the fragile scaffolding for soil functioning. *Soil and Tillage Research, 208*, 104912.

Orgiazzi, A., Ballabio, C., Panagos, P., Jones, A., & Fernández-Ugalde, O. (2018). LUCAS Soil, the largest expandable soil dataset for Europe: a review. *European Journal of Soil Science, 69*(1), 140-153.

Orgiazzi, A., Panagos, P., Fernández-Ugalde, O., Wojda, P., Labouyrie, M., Ballabio, C., et al. (2022). LUCAS soil biodiversity and LUCAS soil pesticides, new tools for research and policy development. *European Journal of Soil Science, 73*(5), e13299.

Panagos, P., De Rosa, D., Liakos, L., Labouyrie, M., Borrelli, P., & Ballabio, C. (2024). Soil bulk density assessment in Europe. *Agriculture, Ecosystems & Environment, 364*, 108907.

Panagos, P., Van Liedekerke, M., Borrelli, P., Köninger, J., Ballabio, C., Orgiazzi, A., et al. (2022). European Soil Data Centre 2.0: Soil data and knowledge in support of the EU policies. *European Journal of Soil Science, 73*(6), e13315.

Redding, T., & Devito, K. (2006). Particle densities of wetland soils in northern Alberta, Canada. *Canadian Journal of Soil Science, 86*(1), 57-60.

Reinsch, S., Lebron, I., de Jonge, L. W., Weber, P. L., Norgaard, T., Arthur, E., et al. (2025). The fraction of carbon in soil organic matter as a national-scale soil process indicator. *Global Change Biology (accepted)*.

Reynolds, B., Chamberlain, P., Poskitt, J., Woods, C., Scott, W., Rowe, E., et al. (2013). Countryside Survey: National "Soil Change" 1978–2007 for Topsoils in Great Britain—acidity, carbon, and total nitrogen status. *Vadose Zone Journal, 12*(2), 1-15.

Robinson, D., Friedman, S., Thomas, A., Hirmas, D., Sullivan, P., & Nemes, A. (2025). Soil bulk density and porosity connecting macro-and micro-scales through geometry. *Earth-Science Reviews*, 105173.

Robinson, D., Thomas, A., Reinsch, S., Lebron, I., Feeney, C., Maskell, L., et al. (2022). Analytical modelling of soil porosity and bulk density across the soil organic matter and land-use continuum. *Scientific reports, 12*(1), 1-13.

Ruehlmann, J. (2020). Soil particle density as affected by soil texture and soil organic matter: 1. Partitioning of SOM in conceptional fractions and derivation of a variable SOC to SOM conversion factor. *Geoderma, 375*, 114542.

Ruehlmann, J., & Körschens, M. (2020). Soil particle density as affected by soil texture and soil organic matter: 2. Predicting the effect of the mineral composition of particle-size fractions. *Geoderma, 375*, 114543.

Rühlmann, J., Körschens, M., & Graefe, J. (2006). A new approach to calculate the particle density of soils considering properties of the soil organic matter and the mineral matrix. *Geoderma, 130*(3-4), 272-283.

Scarlett, B., Van Der Kraan, M., & Janssen, R. (1998). Porosity: a parameter with no direction. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 356*(1747), 2623-2648.

Seaton, F. M., Barrett, G., Burden, A., Creer, S., Fitos, E., Garbutt, A., et al. (2021). Soil health cluster analysis based on national monitoring of soil indicators. *European Journal of Soil Science, 72*(6), 2414-2429.

Thomas, A., Seaton, F., Dhiedt, E., Cosby, B., Feeney, C., Lebron, I., et al. (2024). Topsoil porosity prediction across habitats at large scales using environmental variables. *Science of the Total Environment, 922*, 171158.

Tipping, E., Somerville, C. J., & Luster, J. (2016). The C: N: P: S stoichiometry of soil organic matter. *Biogeochemistry, 130*, 117-131.

Tredennick, A. T., Hooker, G., Ellner, S. P., & Adler, P. B. (2021). A practical guide to selecting models for exploration, inference, and prediction in ecology. *Ecology, 102*(6), e03336.

van Krevelen, D. (1950). Graphical-statistical method for the study of structure and reaction processes of coal. *Fuel, 29*, 269-284.

Wadoux, A. M. C. (2025). Artificial intelligence in soil science. *European Journal of Soil Science, 76*(2), e70080.

Wadoux, A. M. C., Samuel-Rosa, A., Poggio, L., & Mulder, V. L. (2020). A note on knowledge discovery and machine learning in digital soil mapping. *European Journal of Soil Science, 71*(2), 133-136.

Wankmüller, F. J., Delval, L., Lehmann, P., Baur, M. J., Cecere, A., Wolf, S., et al. (2024). Global influence of soil texture on ecosystem water limitation. *Nature, 635*(8039), 631-638.

Watson, K., & Luxmoore, R. (1986). Estimating macroporosity in a forest watershed by use of a tension infiltrometer. *Soil Science Society of America Journal, 50*(3), 578-582.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology, 73*(1), 3-36.

Wood, S. N. (2017). *Generalized additive models: an introduction with R*: chapman and hall/CRC.

Zhao, X., Yang, Y., Shen, H., Geng, X., & Fang, J. (2019). Global soil–climate–biome diagram: linking surface soil properties to climate and biota. *Biogeosciences, 16*(14), 2857-2871.

Zuazo, V. c. H. D., & Pleguezuelo, C. R. o. R. (2009). Soil-erosion and runoff prevention by plant covers: a review. *Sustainable agriculture*, 785-811.

Journal Pre-proof

Conflicts of interest statement

Soil porosity prediction across Europe with a focus on soil particle density determination

The authors whose names are listed immediately below certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Author names:

David A. Robinson[1], Arthur Fendrich[2], Amy Thomas[1], Sabine Reinsch[1], Jens Leifeld[3], Tim Moore[4], James Seward[4], Pasquale Borrelli[5], Cristiano Ballabio[2], Panos Panagos[2]

Affiliations:

1 UK Centre for Ecology & Hydrology, Bangor, UK

2 European Commission, Joint Research Centre (JRC), Ispra, Italy

3 Agroscope, Climate and Agriculture Group, Zurich, Switzerland.

4 Department of Geography, McGill University, Montreal, QC, Canada

5 Department of Science, Roma Tre University, Rome, Italy

Journal Pre-proof