



First insights into soil fauna mapping across Europe using data from multiple data sources for three different taxa[☆]

Flore Marchal^{a,b}, Maria Tsiafouli^{c,*}, Alexandre M.J.-C. Wadoux^d, Maria J.I. Briones^e, Cristina Fiera^f, Paul Henning Krogh^g, Ladislav Miko^h, Christian Mulderⁱ, Maria Papadopoulou^c, Anton Potapov^{j,k,l}, David Russell^j, Lucia Santorufo^m, Andrey S. Zaitsev^{j,n}, Jérôme Cortet^b

^a LISAH, Univ Montpellier, INRAE, IRD, Institut Agro, Montpellier, France

^b CEF, Université Paul Valéry Montpellier 3, Université Montpellier, CNRS, EPHE, IRD, Montpellier, France

^c Department of Ecology, School of Biology, Aristotle University of Thessaloniki, 54124, Thessaloniki, Greece

^d College of Science and Engineering, James Cook University, Cairns, Queensland, Australia

^e Departamento de Ecología y Biología Animal, Universidad de Vigo, Spain

^f Institute of Biology Bucharest, Romanian Academy, Bucharest, Romania

^g Department of Ecoscience, Aarhus University, C.F. Møllers All'e 4, 8000, Aarhus C, Denmark

^h Institute for Environmental Sciences, Faculty of Science, Charles University, Prague, Czech Republic

ⁱ Department of Biological, Geological and Environmental Sciences, University of Catania, Catania, Italy

^j Senckenberg Museum of Natural History Göttingen, Am Museum 1, 02826, Göttingen, Germany

^k German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Puschstrasse 4, 04103, Leipzig, Germany

^l International Institute Zittau, TUD Dresden University of Technology, Markt 23, 02763, Zittau, Germany

^m Department of Biology, University of Naples "Federico II", via Cinthia, 80126, Naples, Italy

ⁿ A.N. Severtsov Institute of Ecology and Evolution, Russian Academy of Sciences, Leninsky pr., 33, Moscow, 119071, Russia

ARTICLE INFO

Keywords:

Soil biodiversity
Community ecology
Collembola
Diplopoda
Lumbricidae
Soil health
Soil conservation

ABSTRACT

Soil fauna is one of the most important components of the soil ecosystem. It is highly diverse and abundant, supporting essential ecosystem services. However, its use in management decisions has been limited by now. This is partly a result of missing data and knowledge for large-scale mapping and application. In this study, our objective was to shed light on these aspects by mapping the spatial distribution of representative groups of macro- and mesofauna. Specifically, we used publicly available data for Earthworms, Millipedes, and Springtails and mapped their abundance, richness, and diversity across Europe. We applied machine learning techniques to understand the spatial configuration of these soil fauna taxa as influenced by topography, climate, vegetation, land use, and soil properties.

Springtail abundance ranged from 1000 to 15,000 individuals per m², while rarefied species richness ranged from 2 to 7 species per sample, with higher values in mountainous regions compared to plains. Earthworm abundance ranged from 20 to 300 individuals per m², while richness ranged from 2 to 6 species, with higher values in Scandinavia compared to central and eastern Europe. Millipede abundance ranged from 20 to 400 individuals per m², while species richness ranged from 1 to 6 species, with the highest values in central Europe and the lowest in northern Europe. Although the three taxonomic groups displayed marked differences in spatial patterns, in general, the diversity values were higher in northern and central Europe compared to South Europe. In the case of the latter, however, the data available were less, which might have led to underestimations of diversity in this area. Our results showed further that for Springtails, Earthworms and Millipedes, altitude, climate, vegetation, and land use were among the strongest predictors of species richness and diversity patterns, while abundance was mainly influenced by vegetation and soil properties. Our results indicate that climate changes, particularly rising temperatures, might impact on the evolution of soil fauna communities in the coming decades, especially in the mountain areas. Although mapping provides a visual presentation of soil ecosystems and conservation trends, as well as soil biodiversity hotspots, we also found that soil mapping has a high degree

[☆] This article is part of a Special issue entitled: 'Soil Biodiversity in Europe' published in Applied Soil Ecology.

* Corresponding author at: Department of Ecology, School of Biology, Aristotle University of Thessaloniki, Thessaloniki, Greece.

E-mail address: tsiafoul@bio.auth.gr (M. Tsiafouli).

of uncertainty (e.g. in areas of low data availability, such in the case of Southern Europe), largely underestimated due to the highly endemic and patchy nature of soil fauna, with significant disparities in current data, including historical differences in methodologies, sampling efforts and scope of studies. Our mapping methodology can be applied at regional, national or European scale, providing actionable management and policy implementation pathways. Our study highlights the need to continue soil biodiversity monitoring efforts, especially in areas with less data, promote data sharing and conduct further research on soil biodiversity conservation and management.

1. Introduction

Soil biodiversity is an essential component of soil ecosystems and is particularly diverse, accounting for 59 % of species living on Earth (Anthony et al., 2023). It plays multiple roles in soil functioning, such as nutrient cycling, soil aeration, decomposition of organic matter, regulation of water, building and maintenance of soil structure, and soil pedogenesis, therefore supporting essential ecosystem services. Soil fauna is a major component of soil biodiversity. This includes macrofauna, such as earthworms or millipedes; mesofauna, such as springtails and mites; and microfauna, such as nematodes or protists, all being actors in the processes of aggregation and organic matter decomposition (de Vries and Shade, 2013; Lavelle et al., 2016; Nicolay et al., 2021). Among them, earthworms and springtails have been largely considered as good indicators of soil health and representative groups of soil macro- and mesofauna (Bispo et al., 2009). Millipedes, although less frequently used as soil bio-indicators, are representatives of epigeic detritivores and are commonly found in various ecosystems. These different fauna groups and size classes are fundamental to soil health and productivity (Bardgett et al., 1999; Wall et al., 2004; Gilarte et al., 2024).

Soil biodiversity is currently under significant pressure due to threats such as urbanization, pollution, intensive agriculture, land use change and climate change (e.g. Tsiafouli et al., 2015; Orgiazzi et al., 2016; Joimel et al., 2021). At the local scale, usually represented by the alpha diversity, the factors influencing structural and functional characteristics of soil fauna include vegetation, climatic factors, soil properties, as well as organism interactions (Remelli et al., 2024). Studies describing soil fauna patterns at landscape scales (i.e. with beta diversity) have recently emerged and showed the contribution of specific environmental parameters, for instance, the urban contexts for earthworms (Eydoux et al., 2024) or Collembola (Milano et al., 2018). Despite these few examples, soil fauna and soil biodiversity remain often understudied, (Guerra et al., 2022) and large-scale spatial patterns are poorly known (Cameron et al., 2019; Fiera et al., 2017). Even though some progress has been made to include soil fauna in large-scale assessment and mapping studies, most of the maps produced in recent years at the global scale have mainly revealed variations linked to biome distribution (Phillips et al., 2019; Potapov et al., 2023).

Although there are initiatives for soil biodiversity conservation (Farfan et al., 2025), currently, there is still a lack of policy and relevant legislative regulation (Römbke et al., 2016; Zeiss et al., 2022; Köninger et al., 2022). However, progressive change and a new status have recently been reached in Europe with the Soil Monitoring and Resilience Directive (European Commission, 2023), in line with the EU Soil Strategy for 2030 (European Commission, 2021) and the European Green Deal (Fetting, 2020). The Directive points out the need for comprehensive data and maps on soil biodiversity for policymakers on a local scale by introducing the concept of soil districts and soil units (Wadoux et al., 2024). Indeed, the production of soil biodiversity-related maps is essential for research and informed policy-making decisions in the area of environmental and agricultural soil policy. There is a requirement for maps and a user-friendly representation of the data that will advance soil biodiversity from a theoretical concept to practical aspects of the environmental and agricultural policy agenda. Maps can support decision-makers in formulating evidence-based policies and strategies that prioritize the preservation of soil biodiversity while ensuring the sustainable use of land resources (Tsiafouli et al., 2023). Further, these maps

can provide information on the stressors affecting soil biodiversity and identify potential threats, as illustrated in resources such as the Global Soil Biodiversity Atlas (Orgiazzi et al., 2016) and the European Atlas of Soil Biodiversity (Artz et al., 2010) or identify soil biodiversity hotspots that need to be protected (Montanarella and Panagos, 2021).

In the last years, maps depicting soil biodiversity-related information have been produced for some taxa on a global or European scale, showing the feasibility of combining empirical data from different sources to create large-scale mapping and provide some first insights into this scientific area (Römbke et al., 2016; Rutgers et al., 2016; Phillips et al., 2019; Potapov et al., 2023; Xue et al., 2024). Local actions require precise maps with high spatial resolution, covering multiple taxonomic groups. Issuing such maps and updating them as more data become available is a challenge. During the last few years, several research and monitoring projects on soils have produced datasets containing some soil biodiversity information at the national or European level (e.g. EcoFINDERS, RMQS, BiSQ, LUCAS Soil Survey). There are, however, large disparities in spatial coverage of soil fauna data, with strong differences in representation for the different groups between countries and regions across Europe (Guerra et al., 2020; Topole et al., 2025).

There have been considerable efforts these last decades to obtain and compile data on soil fauna, but these data need now to be transformed into comprehensive ways for policymakers, which implies mapping. There is an increasing demand for comprehensive biodiversity assessments, a need emphasized more than a decade ago by Wall and Nielsen (2012). However, this effort is subject to various constraints: 1) collating data from different sources is only possible when sufficient metadata are available, such as coordinates and the year of sampling (Guerra et al., 2020); 2) the accuracy of soil biodiversity estimates depends strongly on the sampling effort, which carries a significant risk of underestimating biodiversity (Cameron et al., 2018); 3) the sampling methodology can be different within the collection, making homogenization of data a challenge, but also different methods in taxonomical identification (i.e. morphologically based taxonomic identification vs identification based on molecular methods) may lead to different results (Köninger et al., 2025). This work builds on previous efforts made during the EU COST Action Edaphobase (<https://www.eudaphobase.eu/>) to collect and collate relevant data on soil fauna and store them in a public database warehouse (Russell et al., 2024) and to further produce the European Atlas on Soil Fauna (Tsiafouli et al., 2023). To explore the pattern and diversity of soil fauna, we retrieved different open data. We relied primarily on data from the Edaphobase data-query portal (<https://portal.edaphobase.org/>), focusing on three major taxa: Collembola (Springtails), Lumbricidae (Earthworms), and Diplopoda (Millipedes). These taxa were selected as they belong to the most represented ones in the Edaphobase portal to ensure both a sufficient amount of data and good spatial coverage. We employed mapping techniques to predict soil biodiversity and highlight the gaps in our knowledge of soil fauna on a European scale. This study focuses on three main objectives: 1) to build a consistent dataset from existing databases and point out data gaps, 2) to build a predictive model and analyse the role of environmental variables, and 3) to draw preliminary maps of soil fauna diversity distribution across Europe and discuss the relevance of the prediction with existing literature. We hypothesized that collating data from different sources and mapping biodiversity for various soil fauna groups would incorporate difficulties but would be feasible and would indicate the

strengths but also the shortcomings in our state of knowledge on soil fauna diversity distribution across Europe. While previous studies have focused on single soil fauna taxa at global or regional scales (e.g., Potapov et al., 2023 for collembola; Rutgers et al., 2016 for earthworms), to our knowledge, our study provides the first European wide assessment encompassing multiple taxa within a unified modelling framework. By applying shared model parameters across various taxonomic groups, we ensured methodological consistency and enabled direct comparison of responses to environmental drivers.

2. Materials and methods

2.1. Study area

The study area includes 44 European countries spanning the European Union Member States and the transcontinental regions of Iceland, Turkey, Georgia, Azerbaijan, Armenia, and the Western part of Russia (See Supplementary Material, Fig. 1). This large area has a broad variety of climatic conditions, from Arctic to Mediterranean (Corcho Alvarado et al., 2009). It features a diverse topography: mountains such as the Alps, Pyrenees, and Carpathians, as well as extensive plains such as the Great European Plain from the Atlantic coast to mountains in the East. Vegetation varies significantly across the region, with five main biomes: tundra, cold evergreen needle-leaf forest (taiga), cool mixed forest, temperate deciduous forests, and Mediterranean scrublands. A bioclimatic stratification of this area was made by Metzger et al. (2005), which yielded 84 strata distributed in 13 environmental zones in Europe.

2.2. Collecting and harmonizing soil fauna data

2.2.1. Data compilation

Soil biodiversity data (see Table 1) was extracted from the Edaphobase portal (<https://portal.edaphobase.org/>). We selected the records included in the Edaphobase portal with the following keywords: “Collembola”, “Diplopoda” and “Lumbricidae”. The dataset included species occurrences, relative abundances and metadata for each sampling site, such as information on sampling methods, collection dates, geographic coordinates, vegetation cover, and number of sample replicates.

We extracted the selected records and performed the following pre-processing: we removed sites located outside the study area (i.e. outside the European continent) and excluded records with insufficient

Table 1

Dataset used in the analysis alongside with number of records retained. The pipeline depicting all the steps to reach this pre-processed number is shown in Fig. 2 of the Supplementary Material.

Taxa	Database	Number of records	Reference
Springtails	Edaphobase	42,369	Russell et al. (2024)
	GBIF	48,877	GBIF.org (2023)
	#GlobalCollembola	91,198	Potapov et al. (2023)
	Additional articles	274	e.g. Sevgili and Özata (2014)
Earthworm	Edaphobase	28,720	Russell et al. (2024)
	GBIF	10,465	GBIF.org (2023)
	Phillips et al., 2021 dataset	54,268	Phillips et al. (2021)
	Bouché, 1972 dataset	19,888	Bouché (1972)
Millipedes	Edaphobase	47,710	Russell et al. (2024)
	GBIF	20,498	GBIF.org (2023)

information (i.e. absence of geographic coordinates, inadequate traceability of the data such as missing data sources or authors) and samples with a lack of information on the number of individuals (i.e. abundance) or samples that were not collected from soil. We also excluded samples collected before 1980, artificial habitats (e.g. urban areas) and glaciers. It has to be noted that data included in our study derive from various studies throughout the years not following the same sampling protocols (e.g. sampling size or sampling depth). This might be a source of variation in our data set that we could not specifically account for. As this source of variation concerns the entire dataset, it is unlikely that it influenced the results of our models on the European scale.

To allow for comparisons across sites, we used the rarefied species richness, which standardizes richness estimates by accounting for differences in sampling effort. As species richness is highly sensitive to sampling effort, with more intensively sampled sites appearing to be richer in species (Engemann et al., 2015), we excluded those samples with fewer than 10 individuals as they are sensitive to under-sampling, prone to reflect insufficient sampling effort than the actual community composition. This threshold was defined with statistical evaluation based on a histogram of the distribution of the number of individuals in the samples. The histogram showed a right-skewed distribution, with many samples containing very few individuals, which would decrease the robustness of the rarefied species diversity estimation and had to be removed.

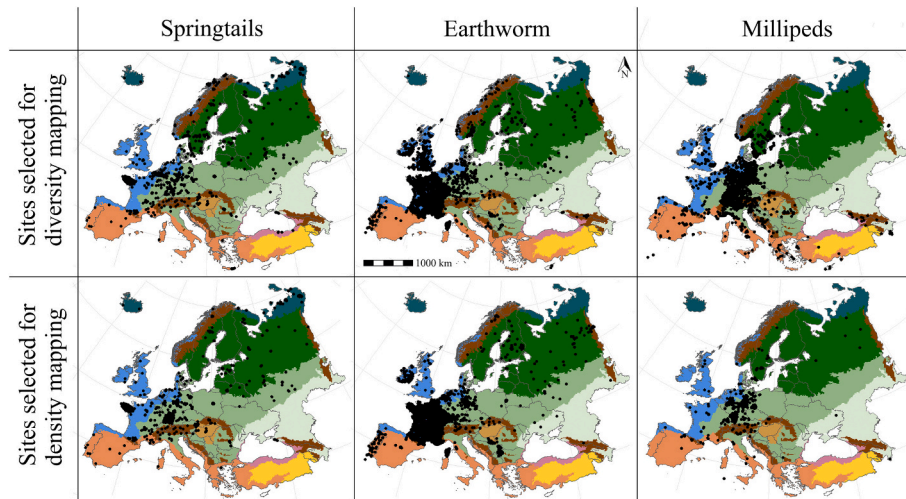


Fig. 1. Geographical locations (black dots) of the sampling sites retained for analysis after pre-processing and based on datasets used to calculate diversity indices and those containing abundance information for Springtails, Earthworms and Millipedes. The background colors show the five main biomes: tundra, cold evergreen needle-leaf forest (taiga), cool mixed forest, temperate deciduous forests, and Mediterranean scrublands. A map in the Supplementary Material (Fig. 3) presents the sites included from model fitting and prediction. All maps are projected using the Lambert Azimuthal Equal-Area projection.

This exclusion could result in the removal of sites with potentially high biodiversity, in case they would have very few individuals belonging to different species. Our heatmap however shows (Fig. 4 Supplementary Material) that samples with few individuals generally have low richness, and that richness becomes more evenly distributed when more individuals are sampled (in contrast, we did not remove samples with an exceptionally high number of individuals as they provide valuable ecological information and are not keen to under-sampling). In addition, we tested our model with different thresholds for minimum sample size (5, 10, and 15 individuals) and produced prediction maps the comparison of which showed that our results are robust to the choice of threshold and the decision to remove the samples with fewer than 10 individuals was necessary to ensure that our analysis was based on a sufficient number of individuals per sample (see Fig. 5 and Table 1 in the Supplementary Material). It also has to be mentioned that models linking low-count samples (i.e. samples with <10 individuals) increased model uncertainty, while not providing additional information for biodiversity mapping at the European scale nor influencing the identification of biodiversity hotspots at this scale. Hence, excluding extremely under-sampled sites was essential to ensure the reliability of our rarefaction process and to avoid distortions in biodiversity patterns.

To further expand our dataset, we incorporated additional records from the literature that were not stored in the Edaphobase portal. We used data from the Global Biodiversity Information Facility-GBIF (<https://www.gbif.org/>), and the Marcel Bouché Earthworm dataset (Bouché, 1972), as well as data from the studies of Potapov et al. (2023) and Phillips et al. (2019). Each of these datasets was subject to the same pre-processing steps described previously. Specific filters were applied to remove opportunistic records (e.g. data from GBIF focusing on one species observation). For Bouché's dataset, we only used those records that corresponded to the current described land use to ensure habitat consistency over time.

We excluded data concerning soil fauna collected by non soil-specific sampling methods, such as vegetation beating or canopy sweeping, to maintain a clear focus on edaphic (i.e. soil-dwelling and soil surface) fauna. We also harmonized the taxonomical levels by retaining only species-level or genus-level data using the GBIF Species Lookup tool. Finally, we eliminated duplicate records by comparing geographic coordinates and sampling references across all data sources. Our compiled dataset is based on species occurrences, meaning that we only consider the presence of species in each sample and do not account for the absence of species in non-sampled areas. The entire cleaning process described in this section resulted in the removal of a significant amount of data due to missing or inconsistent data, leading to a notable reduction in the initial size of the dataset. A summary of the data counts and removals at each processing step is provided in Fig. 2 of the Supplementary Material.

We excluded data concerning soil fauna collected by non soil-specific sampling methods, such as vegetation beating or canopy sweeping, to maintain a clear focus on edaphic (i.e. soil-dwelling and soil surface) fauna. We also harmonized the taxonomical levels by retaining only species-level or genus-level data using the GBIF Species Lookup tool. Finally, we eliminated duplicate records by comparing geographic coordinates and sampling references across all data sources. Our compiled dataset is based on species occurrences, meaning that we only consider the presence of species in each sample and do not account for the absence of species in non-sampled areas. The entire cleaning process described in this section resulted in the removal of a significant amount of data due to missing or inconsistent data, leading to a notable reduction in the initial size of the dataset. A summary of the data counts and removals at each processing step is provided in Fig. 2 of the Supplementary Material.

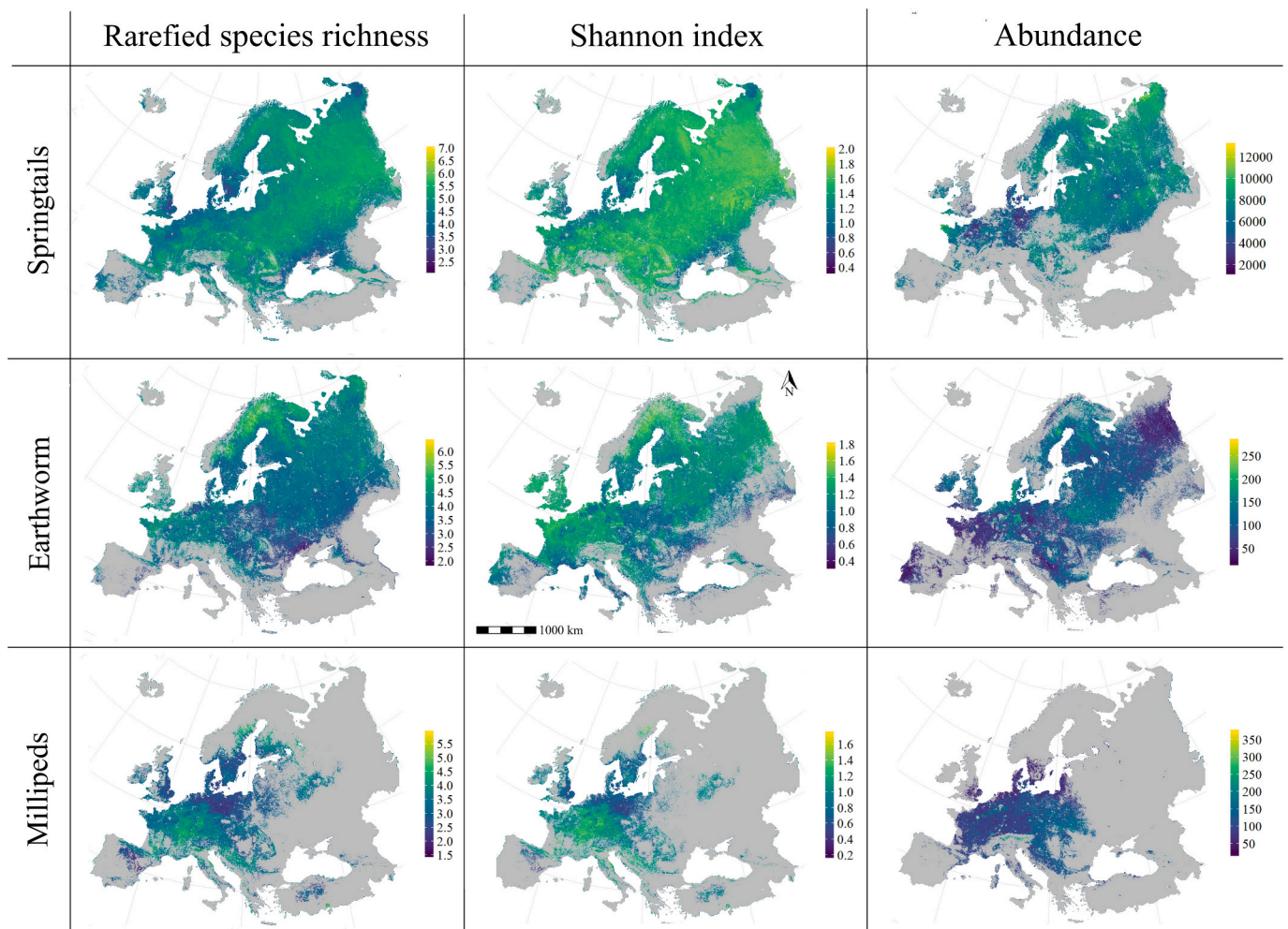


Fig. 2. Maps of Rarefied species richness, Shannon index, and abundance across Europe for Springtails, Earthworms and Millipedes. Grey areas indicate regions where the model prediction fall outside the AOA. The dissimilarity index (DI) thresholds for each model are given in Table 2 in the supplementary material. All maps are projected using the Lambert Azimuthal Equal-Area projection.

2.2.2. Calculating community descriptors

We focused on the local scale by computing community indicators at the sample level (α -diversity). After filtering and sorting data, we obtained two distinct datasets (Table 2). The first dataset, referred to as the diversity dataset, included species composition and the number of individuals per species for each sample. It was used to compute rarefied species richness and the Shannon diversity index (to have a complementary understanding on the distribution of abundance among the different species in the community) as described below. The second dataset, called the density dataset, refers to those samples for which data on individuals per surface area were available. For the diversity dataset, we characterize soil fauna communities with three metrics: rarefied species richness, the Shannon diversity index, and the species density (i.e. individuals per m²). These metrics provide complementary insights into community structure by capturing different aspects of diversity and abundance (Bardgett and Putten, 2014). They are described hereafter.

Rarefied species richness measures biodiversity while accounting for sample size differences. Since larger samples tend to capture more species, we applied rarefaction to standardize richness to the smallest sample size. Rarefied species richness is calculated as (Hurlbert, 1971):

$$E_{\text{rarefied}} = \sum_{i=1}^S \left[1 - \frac{\binom{N - N_i}{m}}{\binom{N}{m}} \right], \quad (1)$$

where S represents the number of observed species, N is the total number of individuals in the sample, N_i is the number of individuals for species i , and m is the standardized sample size. The Shannon diversity index quantifies diversity based on species' relative abundance, incorporating both richness and distribution of abundance among species. The Shannon index was calculated using the vegan package in R (Oksanen et al., 2015). For the density dataset, the density (i.e. individuals per surface area) provides an estimate of community size. This metric was either directly obtained from the datasets or calculated based on sampling area or sampling device specifications. Our analysis focused on the alpha diversity, with all descriptors computed at the sample level rather than the site level. Since sites may include multiple temporally or spatially distinct samples, treating each sample independently captures within-site variability without averaging local fluctuations.

2.3. Environmental covariates

We collected a set of spatially exhaustive environmental covariates known to influence soil fauna community structure and spatial distribution. The set contains 21 climate, topography, soil, vegetation and land use variables: climate data from WorldClim (mean annual precipitation, mean annual air temperature, air temperature seasonality, air temperature annual range, precipitation seasonality, precipitation of the driest quarter, and aridity index), topographical features (i.e. elevation and slope), soil physicochemical properties within the 0–30 cm layer from SoilGrids (i.e. pH of water, texture, soil organic carbon concentration, bulk density, etc.), vegetation indices (e.g. Enhanced Vegetation Index (EVI) - EVI shannon, NDVI), net ecosystem productivity (NEP), land cover and Population density as an index of human pressure.

Table 2

Summary of the main characteristics of the two datasets used in this study that focus on three major soil taxa: Springtails, Earthworms, and Millipedes.

Taxa	Diversity Data		Density Data	
	Number of sites	Number of samples	Number of sites	Number of samples
Springtails	1005	5440	688	3064
Earthworms	1764	2233	1583	3549
Millipedes	1615	2949	347	477

A description of the variables along with their unit and source is provided in the Table 3. Since these covariates originate from different sources and have different spatial resolutions, we standardized them to a common grid resolution of 250 m × 250 m using the *project* function from *terra* package, which performs a resampling with bilinear interpolation. Finally, the covariates were projected with the same spatial extent to the Europe Lambert Azimuthal Equal-Area (LAEA, EPSG:3035) coordinate system.

2.4. Mapping

2.4.1. Quantile regression forest

We employed a digital soil mapping (DSM) approach (Mcbratney et al., 2003) to relate measured values of soil fauna diversity and abundance to environmental covariates using Quantile Regression Forests (QRF, Meinshausen, 2006). QRF is an ensemble learning algorithm that has the same basis as random forests. It works by aggregating decision trees through bootstrap aggregation (i.e. bagging). Unlike

Table 3

Set of environmental covariates along with their acronym, unit and source.

Type	Covariate	Abbreviation	Units	Reference
Climate	Global aridity index	Aridity	no unit	Zomer et al. (2022) A
	Precipitation	Mean Prec	mm	Fick and Hijmans (2017)
	Seasonality	Seas Prec	no unit	Fick and Hijmans (2017)
	Temperature seasonality	Seas Temp	°C	Fick and Hijmans (2017)
	Prec. of the driest month	DM Prec	mm	Fick and Hijmans (2017)
	Temperature mean	Mean Temp	°C	Fick and Hijmans (2017)
	Temperature of the driest quarter	DQ Temp	°C	Fick and Hijmans (2017)
Soil	Clay content (0–30 cm)	Clay	percent	Poggio et al. (2021)
	Sand content (0–30 cm)	Sand	percent	Poggio et al. (2021)
	Bulk density (0–30 cm)	B dens	cg/cm ³	Poggio et al. (2021)
	pH Water (0–30 cm)	pH H ₂	no unit	Hengl et al. (2017)
	Soil organic Carbon (0–30 cm)	SOC	percent	Hengl et al. (2017)
	Water content at field capacity (at 15 cm)	WFC	cm ³ /cm ³	Tóth et al. (2017)
	Wilting point (at 15 cm)	WP	cm ³ /cm ³	Tóth et al. (2017)
Topography	Elevation	Eleva	meter	HO et al. (2023)
	Slope	Slope	percent	Yamazaki et al. (2017)
Vegetation	Diversity of EVI (EVI Shannon)	EVI Sha	no unit	Tuanmu and Jetz (2015)
	NDVI	NDVI	no unit	U.S. Geological Survey (2023)
Anthropogenic pressure	Net ecosystem productivity	NEP	gC/m ² /an	Han and Liu (2024)
	Land cover for 2017	Land C	no unit	Malinowski et al. (2020)
	Population Density	Pop km2	persons/km ²	Lloyd et al. (2017)

standard random forests, which provide only the conditional mean of the response variable, QRF estimates the entire conditional distribution by retaining all response values within each leaf node. This enables the derivation of conditional quantiles, thereby enabling the estimation of prediction intervals. The QRF has three user-defined parameters: *ntree*, which defines the number of trees, *Mtry*, which corresponds to the number of covariates randomly selected at the splitting of the tree, and *min.node.size*, which is a stopping criterion of the minimum leaf at the end of the node stops to split. A grid search was performed to identify the optimal number of trees, the *Mtry* parameter, and the minimum node size. We investigated values between 50 and 1000 trees in increments of 50, *Mtry* values between 5 and 50 in increments of 5, and minimum node sizes between 1 and 10 in increments of 1.

The modelling with QRF followed four steps (Wadoux et al., 2020): 1) model fitting where the model is calibrated using the measured values of soil fauna with the matching values of environmental covariates, 2) model validation where the prediction quality is assessed using a spatially-weighted cross-validation strategy, 3) prediction where a final model is fitted and used for prediction purposes across the study area, and 4) uncertainty and applicability assessment where prediction uncertainty is quantified using the 90 % prediction interval and the area of applicability (AOA) is calculated. More information on these steps is provided in the next subsections.

2.4.2. Model fitting and validation

Individual QRF models were fitted for each taxon and diversity index as the response variable, using the set of environmental covariates as predictors. The three main QRF parameters (i.e. *Mtry*, *ntree*, and *min.node.size*) were optimised by using a cross-validation procedure described in Probst et al. (2018). The implementation was carried out using the *train* function of the *caret* R package (Kuhn, 2008), which allowed testing various parameter combinations for using a grid search and 100 iterations. The best set of parameters evaluated on the root mean squared error (RMSE) was then used to fit all subsequent models. All QRF models were trained using the *ranger* package (Wright and Ziegler, 2015).

To evaluate the quality of predictions, we used a 10-fold cross-validation approach, which involves splitting the dataset into 10 subsets. The model is trained on nine folds, and a prediction is made on the remaining fold. This ensures that every fold is used for testing once. Spatial projection of sites showed that they were moderately clustered. To avoid over-optimistic map accuracy assessment due to the spatial correlation of the errors, we adopted a cross-validation strategy recommended in de Bruin et al. (2022) for datasets with moderate to strong spatial clustering. This method incorporates spatial density by assigning weights to the residuals (i.e. measured minus predicted values) obtained with cross-validation. This ensures that areas with higher sampling intensity are given smaller importance in the calculation of the validation statistics. The weights were derived from the inverse of the sampling intensity based on a kernel density estimate. We calculated using this approach two statistics: the root mean square error (RMSE), which measures the average magnitude of prediction errors with lower values indicating lower error, and the model efficiency coefficient (MEC), which has an optimal value of one but can be negative (Wadoux et al., 2022). All statistics were calculated using weights applied to the residuals, as follows:

The root mean square error:

$$\text{RMSE} = \sqrt{\frac{\sum (\text{resid}^2 \cdot \text{weight})}{\sum (\text{weight})}}, \quad (2)$$

The modelling efficiency coefficient:

$$\text{MEC} = 1 - \frac{\sum (\text{weight} \cdot \text{resid}^2)}{\sum (\text{weight} \cdot (\text{obs} - \mu_{\text{ref}})^2)}, \quad (3)$$

with

$$\mu_{\text{ref}} = \frac{\sum (\text{weight} \cdot \text{obs})}{\sum (\text{weight})}, \quad (4)$$

In Eq. 2 and 3, *resid* are the residuals (i.e. the differences between measured and predicted values). The term *weight* reflects spatial proximity, assigning higher influence to points that are far from the rest of the dataset. The kernel width criterion used in this study was obtained from *bw.CvL* and intensity estimates from the *density.ppp* functions from the package *spatstat* (Turner and Baddeley, 2005). The term μ_{ref} is the weighted mean of observed values and *obs* denotes the measured values.

2.4.3. Prediction and uncertainty quantification

Predictions for the study area were generated using the final QRF model onto a fine grid with a resolution of 250 m × 250 m. This was implemented with the *predict* function from the *terra* package. Prediction uncertainty was obtained by simulations of the QRF model predictions using the *what = function(x) sample(x)* option in the *predict* function, from which we obtained the location-specific standard deviation value using the *stdev* function from the *terra* package.

2.4.4. Area of applicability

To assess the spatial reliability of our models, we calculated the AOA as proposed in Meyer and Pebesma (2021). The AOA identifies regions where predictions may be unreliable due to insufficient representation of environmental covariates in the calibration dataset. This enables highlighting areas where predictions should be interpreted carefully, and additional data may be necessary to confirm the predictions. Obtaining the AOA involves comparing the predictor space of the study area to that of the calibration dataset to calculate a dissimilarity index (DI). The DI is obtained by a distance metric in the covariate and applying a threshold above which a new location is considered as falling outside the area of applicability. This suggests that a low DI indicates that the prediction point is similar to the calibration dataset, whereas a high DI indicates an important dissimilarity. The AOA threshold and cross-validation procedure is made following Meyer and Pebesma (2021) while the implementation was done using the *aoa* function from the *CAST* package (Meyer et al., 2024).

We map the AOA hereafter by cropping the predictions that fall outside of the AOA. This avoids showing predictions in areas where the model cannot be trusted and where further data collection is needed. To enhance result clarity, we present the prediction maps with overlaid areas of applicability.

2.5. Variable importance

Variable importance was obtained with Shapley values. Shapley values are based on game theory. They measure the contribution of each variable to the model prediction relative to the mean of the prediction by considering possible combinations of variables. More explanations on Shapley values and model interpretation for soil mapping are provided in Wadoux and Molnar (2022). The values were calculated using the *explain* function from the *fastshap* package on the final model used in prediction. Shapley values were determined with 50 Monte-Carlo simulations, which has been proved as sufficient from previous studies (Wadoux et al., 2022; Wadoux, 2023; Wadoux et al., 2023). We tested whether 50 simulations were sufficient by plotting the Shapley values obtained with 50 simulations against the Shapley values obtained with another run of 50 simulations. The results suggested a good agreement of the two runs in a scatterplot, which means 50 simulations were enough to obtain stable results. The Shapley values were aggregated by adding all absolute values of the contributions and then sorted from the largest absolute contributor to the smallest to identify the 10 most important covariates in prediction.

3. Results

3.1. Soil fauna data

The sites retained for the spatial models are shown on the map in Fig. 1, where major European and national monitoring projects are clearly identifiable. For instance, the RMQS database in Brittany is prominently represented by springtails, and the BiSQ initiative sites are visible for all three taxa. The Edaphobase database provides, in general, good coverage, but data are clustered in the central part of Europe. Databases like GBIF, Potapov et al. (2023) and Phillips et al. (2019) contribute to broader spatial coverage, particularly across Eastern and Northern Europe. The map in Fig. 1 reveals the high degree of geographical clustering and site density across Europe. The map also show important regional variability in sampling efforts and many areas with limited data. For instance, the Steppic, Anatolian, and southern Mediterranean regions are notably under-sampled and often have no data at all in the dataset. The number of sampled sites decreases drastically towards the east of Europe, whereas Western and Central Europe have more sites. The Alpine regions of the Alps are the most represented in the data.

3.2. Mapping

3.2.1. Model validation

The validation statistics obtained by the density-weighted cross-validation are presented in Table 4. Model prediction varied across taxa and community metrics, with diversity indices having higher predictive accuracy than abundance for the three taxa. The Shannon index consistently yielded the best results, particularly for Earthworms and Millipedes, where the model explained approximately 67 % of the variance (MEC = 0.67) with low error (RMSE = 0.23). Rarefied species richness also performed well for these taxa, with MEC values around 0.64. In contrast, abundance was predicted with lower accuracy across all groups, especially for Springtails, where only 31 % of the variance was explained and the prediction error was relatively high (i.e. the RMSE is 2061). Overall, all models were sufficient to explain some of the spatial variation for the community indices of the three taxa and can be used as the basis for the prediction and interpretation steps.

3.2.2. Distribution of biodiversity levels and environmental predictors

The maps showing the spatial distribution of alpha diversity and abundance for Springtails, Earthworms and Millipedes are presented in Fig. 2 (see Table 2 in the Supplementary material for Dissimilarity Index thresholds of each model). As previously noted, areas outside the model were masked. The maps reveal marked differences in spatial patterns among the taxonomic groups. For all three taxa, values were generally higher in central and northern Europe, while southern and eastern regions, particularly the Mediterranean (i.e. areas with lower data points), show consistently lower values across all metrics. For springtails, rarefied species richness ranged from 2 to 7 species per sample, with higher

values in mountainous regions compared to plains. The Shannon index varied from 0.4 to 2.0, with lower values in the plains of western and northern Europe (e.g. Paris Basin, England, southern Scandinavia, southern Ukraine) and higher values in central (Alps) and eastern Europe (Russia, Finland). Abundance ranged from 1000 to 15,000 individuals per m², with the highest values in eastern Europe (Russia) and lower values in western and central Europe (northern France, Germany, Denmark). Earthworm species richness ranged from 2 to 6 species, with higher values in Scandinavia (northern Sweden and Norway) compared to central and eastern Europe (Poland, southern Ukraine). The Shannon index (0.2 to 2.0) was highest in western (northern France, Great Britain) and northern Europe (Scandinavia), and lowest in eastern (Poland, Czech Republic) and southern (southern France, Italy, Spain) Europe. Abundance ranged from 20 to 300 individuals per m², with the highest values in Scandinavia (Denmark, Sweden, Finland), central (Germany) and eastern Europe (Belarus). Millipede species richness ranged from 1 to 6 species, with the highest values in central Europe (eastern France, southern Germany) and the lowest in northern Europe (Great Britain, northern Germany, Poland, southern Scandinavia). The Shannon index (0.2 to 1.8) followed a similar spatial pattern. Millipede abundance ranged from 20 to 400 individuals per m², with higher values in eastern Europe (Poland, Belarus) and the Alps, and lower values in western Europe. Fig. 2 also shows that the model is consistently predicting outside the AOA in southern and eastern Europe, as well as in north-eastern Europe and Scotland, where grey areas are dominant. The maps of the sampling locations (Fig. 1) confirm this pattern since grey areas represent zones with no data or a relatively low density of sampled sites.

Fig. 3 shows the main contributors to model predictions. For species richness and Shannon index, elevation is one of the most influential variables across all three taxa. Climate parameters such as temperature seasonality and temperature of the driest quarter for Springtails and Millipedes, and mean precipitation for Earthworms, are the next most important factors. Abundance is more strongly driven by vegetation-related variables such as NDVI and net ecosystem productivity for Springtails and earthworms, while soil texture, particularly sand content, is most influential for Millipede abundance. Springtail abundance is also affected by soil properties, especially soil organic carbon, while sand content influences Earthworm abundance as well. Secondary variables, such as soil pH and bulk density, are relevant for Springtail species richness and abundance; bulk density also contributes to earthworm abundance; and pH influences the species richness and the Shannon index of Earthworms and Millipedes. Land cover is a secondary but influential factor for most taxa and metrics, except for Springtail abundance.

4. Discussion

The compilation of data for several soil fauna taxa, progressively developed by the scientific community over time, for various purposes across Europe, and stored in several sources (multiple repositories, such as Edaphobase, GBIF, and other sources, such as data papers) was especially challenging. In this study, our mapping effort demonstrated the potential of these collated data, while it also highlighted limitations in the use and interpretation that should be addressed in future research.

Our results demonstrate that it is possible to predict alpha diversity at a fine spatial resolution (i.e. 250 m²) across Europe for key taxa used in soil health assessments, such as Springtails and Earthworms, as well as less frequently considered groups like Millipedes. The predictive map values are consistent with existing literature and the known ecology of the target taxa. Specifically, climatic parameters, altitude, land use and vegetation strongly influenced all three model groups, while soil organic carbon was relevant only for endogeic taxa, such as springtails and earthworms. These findings highlight the possibility of using public databases to support soil biodiversity monitoring and, more broadly, soil health assessments by stakeholders and policy-makers. Nevertheless, the results also point out major challenges that must be addressed to

Table 4
Validation statistics for the maps for Springtails, Earthworms and Millipedes. The statistics are obtained by 10-fold cross-validation [MEC: model efficiency coefficient, RMSE: root mean square error].

Taxa	Community index	MEC	RMSE
Springtails	Rarefied species richness	0.51	0.94
	Shannon index	0.53	0.34
	Abundance	0.31	2061
Earthworm	Rarefied species richness	0.64	0.54
	Shannon index	0.67	0.23
	Abundance	0.49	31.9
Millipedes	Rarefied species richness	0.64	0.54
	Shannon index	0.67	0.23
	Abundance	0.52	25.7

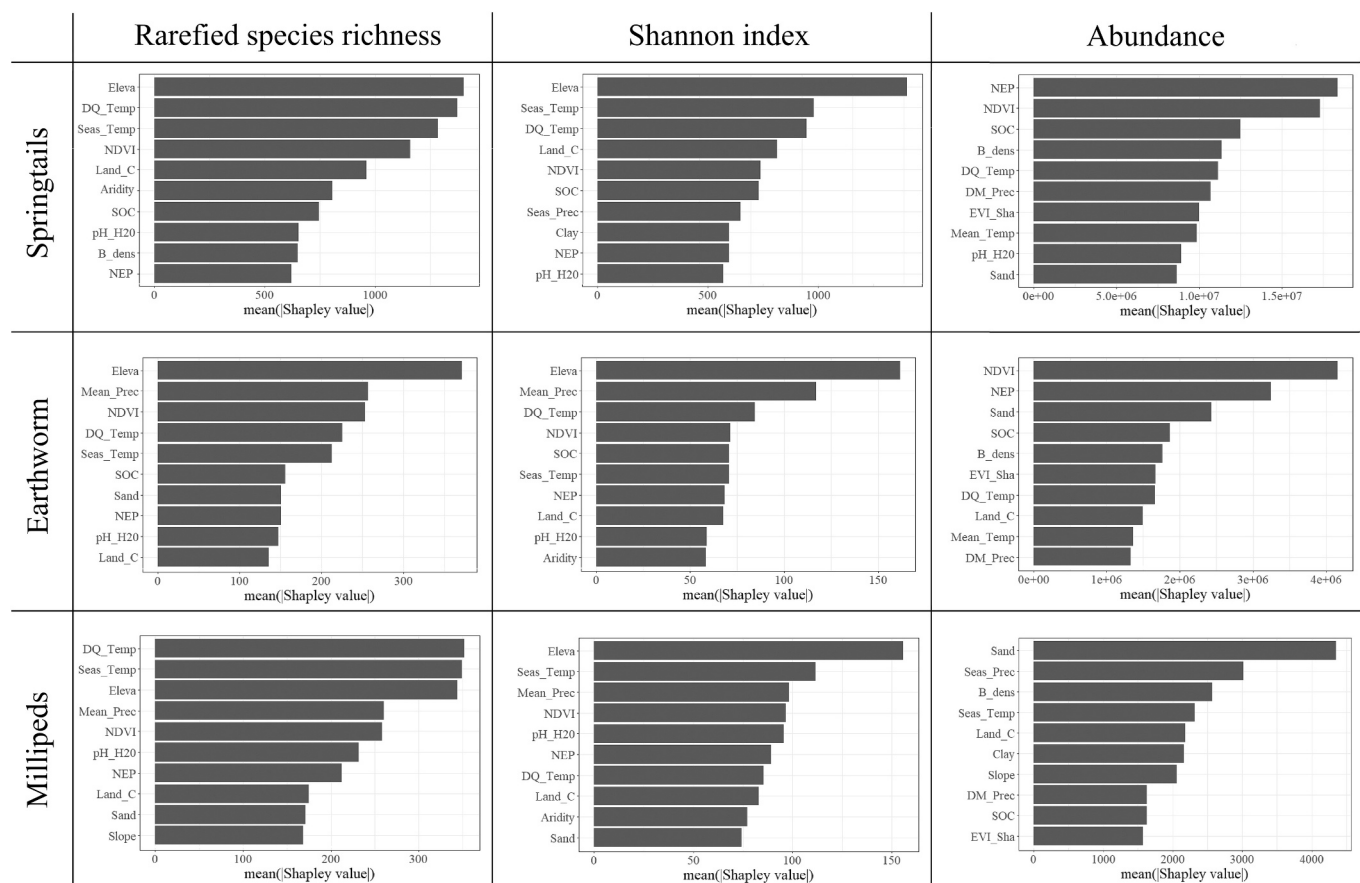


Fig. 3. Importance of the covariates in model prediction as calculated with the average of absolute Shapley values. Each row corresponds to a different taxon: Springtails (top row), Earthworms (middle row), and Millipedes (bottom row). The columns represent an index, either local rarefied species richness, Shannon diversity index, or the abundance (i.e. the number of individuals per m^2). Covariates are ranked from most to least important and only the 10 most important variables are reported [for abbreviations see Table 3].

advance these efforts.

4.1. Biodiversity levels and factors of variation

For Springtails, predicted levels of rarefied species richness appear to be underestimated compared to those reported by Joimel et al. (2017), who found values ranging from 3.2 in vineyards to 21.7 in forests, and Potapov et al. (2023), who reported approximately 20 species in temperate agricultural systems. This under-estimation could be attributed to the smoothing effect of regression models, which is a common problem in soil mapping studies. However, both studies also observed a wide range of values across European latitudes, from 0 to 100 species. Predicted springtail abundance, ranging between 2000 and 14,000 individuals per m^2 , is consistent with the findings of Joimel et al. (2017) and Potapov et al. (2023). The Shannon index is more difficult to compare, as it has not previously been modelled, but our values, ranging from 0.4 to 2.0, fall within the range reported in the literature for springtails.

For earthworms, the species richness predicted by our models is consistent with values reported in previous studies. Cluzeau et al. (2012) observed a mean species richness of eight per site in Brittany (France), while Phillips et al. (2021) reported an average of around six species across Europe. Rutgers et al. (2016) found values ranging from one to seven species in Western Europe. Similarly, our predicted Shannon index values align with those of Rutgers et al. (2016), who reported values between 0.8 and 1.6. Reported earthworm abundance in the literature ranges from 5 to 500 individuals per m^2 according to Rutgers et al. (2016), and from 50 to 350 individuals per m^2 on average in

Brittany, as noted by Cluzeau et al. (2012). Phillips et al. (2021) recorded values up to 150 individuals per m^2 across Europe. Here too, our predicted values therefore fall within the range reported in the literature.

Data on Millipedes are relatively scarce compared to those for springtails and earthworms. However, a literature review by Saulnier and Athias-Binche (1986) reported abundances ranging from 40 to 190 individuals per m^2 in forest systems across several countries (i.e. Denmark, the Netherlands, Belgium, France, and Russia). Aubert et al. (2005) recorded an abundance of 39 individuals per m^2 in French beech forests, while David (1989) reported 97 individuals per m^2 in the Orleans forest (France), with mean species richness ranging from five to ten species per site. These values are consistent with our model predictions.

The integrative approach focusing on multiple taxa in this study demonstrates that different parameters are influenced by distinct climatic, edaphic, and land-use factors, thereby offering novel insights into soil biodiversity patterns and advancing methodological developments in multi-taxa ecological modelling. Specifically, our results indicate that species richness and diversity are primarily driven by elevation and climatic parameters, whereas abundance is mainly influenced by ecosystem productivity variables related to vegetation and soil properties, particularly texture, and soil organic carbon. These findings are consistent with ecological theory, as climate and elevation are well-established determinants of species distribution and community composition (Wilson et al., 2005), while soil fauna abundance is closely associated with carbon availability and energy flow, typically higher in forest ecosystems. Indeed, for example, the reason why millipede

abundance is mainly influenced by sand content, and therefore texture, appears difficult to explain in light of the research on the subject. Thus, most of the studies mentioned above did not test these parameters on millipede communities because these animals live mainly in hologranic horizons and are therefore assumed to be mainly influenced by factors related to vegetation or litter. One could hypothesise that highly sandy textures are favourable to the development of hologranic horizons of the mor or moder type, which are themselves very favourable to the development of certain species, leading to high abundances. Sandy textures might also be more favourable for burrowing activities of larger animals such as millipedes, which lack the burrowing functional capabilities of worms; whereas smaller animals, like springtails, can take advantage of the spaces associated with microporosity.

In addition, land cover, reflecting anthropogenic pressure, emerges as one of the most influential factors, underscoring the substantial impact of land management on soil fauna communities. However, this would require further investigation.

4.2. Environmental predictors

Our maps provide preliminary insights into the relationship between biodiversity and potential environmental stressors, supporting adaptive management approaches (Phillips et al., 2019). Covariates were selected a priori based on their known relevance to soil biodiversity, without applying statistical variable selection methods. Their choice was also constrained by the spatial extent of the study area. Model performance was evaluated using MEC and RMSE. Although various evaluation metrics are used in other studies, our approach incorporates observation density through weighting, allowing adjustment for sampling bias. The resulting MEC values ranged from 0.31 to 0.67, indicating overall acceptable performance and falling within the range reported in other studies on earthworm biodiversity using random forest models (i.e., the MEC range is usually 0.4–0.7) (Johnston et al., 2014; Salako et al., 2023). While predictions for species richness and Shannon index were relatively accurate, abundance, particularly for springtails, was more variable. This might be due to the stochastic nature of abundance data, which are strongly influenced by micro-habitat conditions and temporal variation (Eriksen-Hamel and Whalen, 2007; Mathur and Mathur, 2023). Springtails, in particular, are highly sensitive to such fluctuations (Potapov et al., 2023). But these micro-environmental factors are not totally captured at the coarse resolution of our European-scale environmental rasters, leading to reduced predictive accuracy at this geographical extent. Also, the underlying datasets combine records from multiple sources with varying sampling and extraction methods, adding methodological noise to abundance estimates. Together, these ecological and methodological factors explain the lower model performance observed for Collembola.

4.3. Spatial patterns and their implications

As with aboveground biodiversity, high-resolution mapping has historically been a key tool in successful conservation policies (Montanarella and Panagos, 2021). Crucially, such maps provide policymakers with evidence-based insights for identifying biodiversity hotspots and developing mitigation strategies against land-use intensification and pollution, while promoting sustainable agricultural practices (Tsiafouli et al., 2023). Our analysis reveals a significant spatial gap in soil fauna biodiversity data across Europe, with Southern and Eastern regions notably underrepresented. This pattern aligns with findings by Topole et al. (2025). Regions such as the Mediterranean, which harbor high biodiversity and endemic species, are particularly affected (Marchal et al., 2022). Cameron et al. (2019) highlighted global disparities in soil biodiversity data, often linked to differences in economic investment and research capacity. Under-representation in these regions has important implications for biodiversity modelling, increasing uncertainty and risking misidentification of biodiversity hotspots (Guerra

et al., 2020).

In our study, areas with low sampling density were de facto excluded from model predictions (this is shown in grey in Fig. 2). These regions could be considered “blind spots” for conservation and policy development. Rather than include potentially misleading predictions, we chose to highlight these areas as zones of unreliable model predictions, which could be used as a guide for future sampling efforts. Acknowledging such a problem allowed us to illustrate both the spatial structuring of ecological processes and the limitations of current models and data coverage. Identifying these knowledge gaps helps prevent biased policy decisions that could lead to ineffective or misdirected management strategies (Gardi et al., 2009).

Our findings also show the urgent need to expand sampling and enhance data sharing, while emphasizing the lack of harmonized sampling designs and protocols across large regions of Europe, which would allow both a good coverage of existing data and accurate model fitting and prediction (Hermans et al., 2019; Wadoux et al., 2020). We therefore recommend that relevant authorities should invest in soil biodiversity monitoring programs based on standardized and harmonized protocols validated by the scientific community. Data gap analyses can help prioritize under-sampled regions such as the Balkans and parts of Eastern Europe. More comprehensive datasets are essential to reduce model prediction uncertainty and support subsequent soil biodiversity and health monitoring. These efforts align with initiatives such as the Global Soil Biodiversity Atlas (Orgiazzi et al., 2016), the harmonization goals of Rutgers et al. (2016), and emerging policies, including the European Soil Monitoring and Resilience Directive (European Commission, 2023).

4.4. Data harmonization: Selection and curation of data

The main challenge in implementing this synthesis was data harmonization. The data were sourced from multiple studies with different sampling protocols, methods, and sampling dates, and in some cases, critical information was missing. Harmonization required balancing the need for reliability and having a dataset with sufficient spatial coverage.

Another issue arises from the risk of conflating alpha- and beta diversity. To address this, we implemented a sorting process to manage the considerable disparity in the number of individuals collected per sample. We worked at the sample level, setting thresholds for under-sampling to allow rarefied diversity index calculations. This ensures that comparisons were made based on an equal number of samples. This approach reduces sampling error but means that our maps do not reflect all biodiversity observations. Specifically, during the curation process, we had to exclude a significant amount of data (see Fig. 2 in supplementary materials for details of the steps). While this is a significant loss, we preferred to use smaller but more complete/robust datasets for analysis. Even at the site level, methodological choices had to be made. For example, when considering datasets including different treatments it is of question whether the treated samples or only the control samples should be considered. Additionally, how should occurrence data be considered when no abundance was measured? These decisions must be clearly stated and justified. It is also essential to consider the distribution of samples to understand the limitations and uncertainties of the mapping. Such transparency helps avoiding false assumptions about soil biodiversity and ensures that policy decisions are based on robust evidence.

4.5. Future needs for mapping soil fauna biodiversity

An important challenge to address is the focus on species composition, which was not explored here but would provide insights into β diversity. Exploring β diversity could fundamentally alter our understanding of soil biodiversity maps by highlighting biomes known for their heterogeneity at the landscape level. Indeed, β diversity, which

measures dissimilarity between sites, can reveal highly variable patterns that may differ from the results obtained through other metrics. For instance, the Mediterranean climate, with its seasonal variability in humidity and temperature, promotes greater habitat heterogeneity (Gasith and Resh, 1999). This can lead to higher β diversity, as species adapted to specific microhabitats thrive in this environment. Such dynamics may also foster endemism, particularly in springtails, where certain groups show pronounced endemism in Mediterranean regions like the Pyrenean mountains (Deharveng, 1996). In contrast, in more uniform temperate regions, local diversity is often higher, but the same species tend to occupy larger areas (Medina-Sauza et al., 2019). Moreover, focusing on species composition would allow the incorporation of biodiversity conservation concerns, especially regarding rare species. While databases may contain this information, two challenges remain: 1) sampling effort, which is not always clearly documented in datasets, and 2) the need for expertise in data curation.

This study focused on taxonomic diversity, specifically rarefied species richness. However, other dimensions of biodiversity, such as phylogenetic and functional diversity, also merit attention. Extending this approach to these other types of diversity would require additional data. Modelling phylogenetic diversity necessitates information on the evolutionary relationships and phylogenetic distances between all identified species. Similarly, functional diversity requires data on the ecological and functional traits of species, which are not yet available for all taxonomic groups.

On the other hand, our methodology applied in this study, including steps of data collection, collation, cleaning modelling and mapping supports soil assessment and monitoring initiatives, such as those required by the European Soil Monitoring and Resilience Directive for soil biodiversity. Our methodology can be applied at regional, national or European scale for the same or other groups of soil organisms. Depending on data availability, produced maps could be used for locating biodiversity hotspots or biodiversity changes (decline, increase etc.) or indicating data gaps, providing actionable management and policy implementation pathways.

Although our model predictions are produced at a fine spatial resolution (250 m $\times$ 250 m), their practical relevance lies primarily at the regional scale targeted by the European Soil Monitoring and Soil Resilience Directive. The high-resolution outputs provide spatially consistent information on the variability of soil functions across Europe, which can support policy design and implementation by identifying priority areas for intervention, optimizing sampling strategies, and improving resource allocation. Beyond direct map use, the predictions can be aggregated to broader geographic units, such as soil districts, which are the operational scale defined in the Directive for assessing soil health. Aggregated outputs at this scale contribute to the estimation of soil descriptors and functions required for soil health assessment, periodic reporting, and trend analysis. In addition, the associated uncertainty estimates can inform the evaluation of sampling density and estimation error constraints (e.g., the 5 % precision target for the area of healthy soils), thereby supporting the methodological and statistical requirements set out in the Directive.

5. Conclusions

We provided preliminary maps of soil fauna biodiversity across Europe from the compilation of several existing databases. We discussed the spatial pattern, the important environmental predictors that drive the model prediction, and further discussed the implications of our work and soil fauna mapping for large-scale conservation policy. The main findings are as follows:

- We integrated data from multiple repositories, such as Edaphobase and GBIF. While this is a challenging task for data harmonization, it is possible to account for variations in sampling protocols, methods, and dates if careful curation is made. An additional need for future

soil fauna studies is the collection of accompanying data/metadata at the sampling sites to be used as predictors at the same scale.

- There were significant gaps in soil fauna biodiversity data spatial coverage, particularly in Southern and Eastern Europe. These gaps lead to reduced model extrapolation and areas where model prediction cannot be trusted, especially in under-sampled regions. There is an urgent need for data coverage of these areas (e.g. through systematic fieldwork and sampling financed by the relevant regional, national and/or European sources).
- Prediction of alpha diversity for key soil fauna taxa, including Springtails, Earthworms, and Millipedes, at a fine spatial resolution (250 m²) aligns with existing literature and underscores the role of environmental variables, such as elevation and climate, but also soil organic carbon and land use (reflecting anthropogenic pressure), in shaping soil fauna communities' spatial variation. These results suggest that the climate changes currently occurring in Europe, particularly rising temperatures, combined with changes in land use, will have a significant impact on the evolution of soil fauna communities in the coming decades, especially in the mountain areas, as altitude and temperature appear as the main drivers of species richness.
- We found that predicted springtail species richness may be underestimated due to the smoothing effects of regression models, a common limitation in digital soil mapping. However, predicted springtail abundance and Shannon index values fall within the ranges reported in the literature.
- Species richness and diversity were primarily driven by climatic factors and elevation, while abundance was more strongly influenced by ecosystem productivity and soil conditions, particularly organic carbon content.

We acknowledge that significant areas lacked data in this European-scale mapping. However, this issue can be mitigated by excluding regions from the predictions where the predictive model is unreliable. In the future, harmonized sampling protocols and comprehensive data sharing will be essential to support soil biodiversity monitoring and policy initiatives, such as the European Soil Monitoring and Resilience Directive.

CRedit authorship contribution statement

Flore Marchal: Writing – original draft, Formal analysis, Data curation. **Maria Tsiafouli:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Conceptualization. **Alexandre M.J.-C. Wadoux:** Writing – review & editing, Methodology, Formal analysis, Conceptualization. **Maria J.I. Briones:** Writing – review & editing, Conceptualization. **Cristina Fiera:** Writing – review & editing, Conceptualization. **Paul Henning Krogh:** Writing – review & editing, Conceptualization. **Ladislav Miko:** Writing – review & editing, Conceptualization. **Christian Mulder:** Writing – review & editing, Conceptualization. **Maria Papadopolou:** Writing – review & editing, Conceptualization. **Anton Potapov:** Writing – review & editing, Methodology, Conceptualization. **David Russell:** Writing – review & editing, Methodology, Conceptualization. **Lucia Santorufo:** Writing – review & editing, Conceptualization. **Andrey S. Zaitsev:** Writing – review & editing, Methodology, Conceptualization. **Jérôme Cortet:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: A. Z. serves as executive editor of the journal Applied Soil Ecology M.T. and J.C. serve as guest editors for the special issue entitled “Soil biodiversity at the European scale” If there are other authors, they declare that they

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

Acknowledgments

Edaphobase was developed with the help of multiple partners throughout its lifetime (see <https://www.edaphobase.org> for details). Initial development of Edaphobase was funded by the German Ministry for Education and Research (BMBF) under the support codes 01LI0901A (2009-2013) and 01LI1301A (2014-2018). From October 2019 to March 2024, further development of Edaphobase 2.0 took place within the framework of the EU COST Action “EUDaphobase - European Soil-Biology Data Warehouse for Soil Protection” (CA18237), supported by the COST Association (European Cooperation in Science and Technology) and funded by the European Union. We are immensely grateful to all data providers and especially the members of the COST Action EUDaphobase for developmental input to, and testing of, all software components, nomenclatural standardization, the Edaphobase Data Policy, data re-use procedures, and much more. Alexandre Wadoux has received funding from the European Unions Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 101059012. Flore Marchal was financed by the project RECROP financially supported by the PRIMA European program for Research and Innovation solutions in the Mediterranean region (Project PCI2021-121981 funded 517by MCIN/AEI/10.13039/501100011033 and Next-GenerationEU/PRTR).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.apsoil.2025.106588>.

Data availability

Data will be made available on request.

References

- Anthony, M.A., Bender, S.F., van der Heijden, M.G.A., 2023. Enumerating soil biodiversity. *Proc. Natl. Acad. Sci.* 120, e2304663120.
- Artz, R., Anastasiou, D., Arrouays, D., Bastos, A., Bendetti, A., Bispo, A., Brandmayr, P., Broll, G., Bunning, S., Castracani, C., Campbell, C., Cluzeau, D., Coates, D., Creamer, R., Dias, L., Durrant, T., Fritz, G., Gardi, C., Gemmill-Herren, B., Zanella, A., 2010. European Atlas of Soil Biodiversity. European Commission, Publications Office of the European Union, Luxembourg.
- Aubert, M., Hedde, M., Decaens, T., Margerie, P., Alard, D., Bureau, F., 2005. Facteurs contrôlant la variabilité spatiale de la macrofaune du sol dans une hêtraie pure et une hêtraie charmaie. *Comptes Rendus. Biologies* 328, 57–74.
- Bardgett, R., Putten, W., 2014. Belowground biodiversity and ecosystem functioning. *Nature* 515, 505–511.
- Bardgett, R., Cook, R., Yeates, G., Denton, C., 1999. The influence of nematodes on below-ground processes in grassland ecosystems. *Plant Soil* 212, 23–33.
- Bispo, A., Cluzeau, D., Creamer, R., Graefe, U., Krogh, P.H., Sousa, J.P., P'er'es, G., Rutgers, M., Winding, A., Römke, J., 2009. Indicators for monitoring soil biodiversity. *Integr. Environ. Assess. Manag.* 5, 717–719.
- Bouché, M.B., 1972. Lombriciens de France. *Ecologie et Systématique*. Institut National de la Recherche Agronomique (INRA), Paris, France.
- Cameron, E., Martins, I., Lavelle, P., Mathieu, J., Tedersoo, L., Bahram, M., Gottschall, F., Guerra, C., Hines, J., Patoine, G., Siebert, J., Winter, M., Cesarz, S., Ferlian, O., Kreft, H., Lovejoy, T., Montanarella, L., Orgiazzi, A., Eisenhauer, N., 2019. Global mismatches in aboveground and belowground biodiversity. *Conserv. Biol.* 33, 1187–1192.
- Cameron, E.K., Martins, I.S., Lavelle, P., Mathieu, J., Tedersoo, L., Gottschall, F., Guerra, C.A., Hines, J., Patoine, G., Siebert, J., et al., 2018. Global gaps in soil biodiversity data. *Nature Ecology & Evolution* 2, 1042–1043.
- Cluzeau, D., Guernion, M., Chaussod, R., Martin-Laurent, F., Villenave, C., Cortet, J., Ruiz-Camacho, N., Pernin, C., Mateille, T., Philippot, L., et al., 2012. Integration of biodiversity in soil quality monitoring: baselines for microbial and soil fauna parameters for different land-use types. *Eur. J. Soil Biol.* 49, 63–72.
- Corcho Alvarado, J., Barbecot, F., Purtschert, R., Gillon, M., Aeschbach, W., Kipfer, R., 2009. European climate variations over the past half-millennium reconstructed from groundwater. *Geophys. Res. Lett.* 36.
- David, J.-F., 1989. Les peuplements de diplopodes d'une forêt temp'érée: variations spatiales et stabilité dans le temps. *Revue d'Ecologie et de Biologie du Sol* 26, 75–90.
- de Bruin, S., Brus, D.J., Heuvelink, G.B.M., Tengbergen, T., Wadoux, A.M.J.-C., 2022. Dealing with clustered samples for assessing map accuracy by cross-validation. *Eco. Inform.* 69, 101665.
- de Vries, F.T., Shade, A., 2013. Controls on soil microbial community stability under climate change. *Front. Microbiol.* 4, 265.
- Deharveng, L., 1996. Soil collembola diversity, endemism, and reforestation: a case study in the Pyrenees (France). *Conserv. Biol.* 10, 74–84.
- Engemann, K., Enquist, B., Sandel, B., Boyle, B., Jørgensen, P., Morueta-Holme, N., Peet, R., Violle, C., Svenning, J.-C., 2015. Limited sampling hampers “big data” estimation of species richness in a tropical biodiversity hotspot. *Ecology and evolution* 5.
- Eriksen-Hamel, N.S., Whalen, J.K., 2007. Impacts of earthworms on soil nutrients and plant growth in soybean and maize agroecosystems. *Agric. Ecosyst. Environ.* 120, 442–448.
- European Commission, 2021. EU Soil Strategy for 2030. In: Reaping the Benefits of Healthy Soils for People, Food, Nature and Climate. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the regions. European Commission Brussels, Belgium.
- European Commission, 2023. Proposal for a Directive of the European Parliament and of the Council on Soil Monitoring and Resilience (Soil Monitoring Law). European Commission Brussels, Belgium. [Link. https://environment.ec.europa.eu/publication/s/proposal-directive-soil-monitoring-and-resilience_en](https://environment.ec.europa.eu/publication/s/proposal-directive-soil-monitoring-and-resilience_en).
- Eydoux, L., Cortet, J., Barantal, S., Decaens, T., 2024. Driving factors of earthworm communities in Mediterranean urban parks. *Appl. Soil Ecol.* 202, 105602.
- Farfan, M.A., Guerra, C.A., Ingimarsdottir, M., Hedlund, K., Barrios, E., Cox, N., Dahlberg, A., de Assis Sanchez Crespo, F., de Vries, F., Delgado-Baquerizo, M., Eisenhauer, N., Lundersjö, M., Martin, A., Orgiazzi, A., Parnell, J.J., Potapov, A., Ramirez, K.S., Raschmanová, N., Russell, D., Tchedy, V., Tsiafouli, M., Uhde, L., Zaytsev, A., 2025. Outlook on the knowledge gaps to improve nature conservation of soil biodiversity. *Soils for Europe 1*, e149005. <https://doi.org/10.3897/soils4europe.e149005>.
- Fetting, C., 2020. The European green deal. In: ESDN. Report December. ESDN Office, Vienna.
- Fick, S., Hijmans, R., 2017. Worldclim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* 37.
- Fiera, C., Habel, J.C., Kunz, M., Ulrich, W., 2017. Large-scale phylogenetic structure of European springtails (Collembola) depends on species range size and postglacial colonization history. *Biol. J. Linn. Soc.* 120, 664–674.
- Gardi, C., Montanarella, L., Arrouays, D., Bispo, A., Lemanceau, P., Jolivet, C., Mulder, C., Ranjard, L., Römke, J., Rutgers, M., et al., 2009. Soil biodiversity monitoring in Europe: ongoing activities and challenges. *Eur. J. Soil Sci.* 60, 807–819.
- Gasith, A., Resh, V.H., 1999. Streams in Mediterranean climate regions: abiotic influences and biotic responses to predictable seasonal events. *Annu. Rev. Ecol. Syst.* 30, 51–81.
- GBIF.org, 2023. GBIF Occurrence Download. <https://doi.org/10.15468/dl.abcdcf>. Accessed: 2024-06-05.
- Gilarte, P., Plett, J.M., Pendall, E., Carrillo, Y., Nielsen, U.N., 2024. Soil nematodes modify interactions between nitrogen-fixing and non-fixing tree seedlings from late, but not early, successional stages. *Plant Soil* 1–13.
- Guerra, C., Heintz-Buschart, A., Sikorski, J., Chatzinotas, A., Guerrero-Ramírez, N., Cesarz, S., Beaumelle, L., Rillig, M., Maestre, F., Delgado-Baquerizo, M., Buscot, F., Overmann, J., Patoine, G., Phillips, H., Winter, M., Wubet, T., Küsel, K., Bardgett, R., Cameron, E., Eisenhauer, N., 2020. Blind spots in global soil biodiversity and ecosystem function research. *Nat. Commun.* 11, 3870.
- Guerra, C., Berdugo, M., Eldridge, D., Eisenhauer, N., Singh, B., Cui, H., Abades, S., Alfaro, F., Bamigboye, A., Bastida, F., Blanco Pastor, J., De los R'ios, A., Dur'an, J., Grebenc, T., G. Illan, J., Liu, Y.-R., Makhallanyane, T., Mamet, S., Molina-Montenegro, M., Delgado-Baquerizo, M., 2022. Global hotspots for soil nature conservation. *Nature* 610, 693–698.
- Han, Q., Liu, L., 2024. Upscaling tower-based net ecosystem productivity to global 250m using the data augmentation method by considering their spatial distribution.
- Hengl, T., Mendes de Jesus, J., Heuvelink, G., Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M., Geng, X., Bauer-Marschallinger, B., Guevara, M., Vargas, R., Macmillan, R., Batjes, N., Leenaars, J. G., Ribeiro, E., Wheeler, I., Mantel, S., & Kempen, B., 2017. SoilGrids250m: global gridded soil information based on machine learning. *PLoS One* 12, e0169748.
- Hermans, S.M., Buckley, H.L., Lear, G., 2019. Perspectives on the impact of sampling design and intensity on soil microbial diversity estimates. *Front. Microbiol.* 10, 1820.
- HO, Y.F., Hengl, T., Parente, L., 2023. Ensemble Digital Terrain Model (EDTM) of the World.
- Hurlbert, S.H., 1971. The nonconcept of species diversity: A critique and alternative parameters. *Ecology* 52, 577–586.
- Johnston, A.S.A., Holmstrup, M., Hodson, M.E., Thorbek, P., Alvarez, T., Sibly, R.M., 2014. Earthworm distribution and abundance predicted by a process-based model. *Appl. Soil Ecol.* 84, 112–123.
- Joimel, S., Schwartz, C., Hedde, M., Kiyota, S., Krogh, P.H., Nahmani, J., P'er'es, G., Cortet, J., 2017. Urban and industrial land uses have a higher soil biological quality than expected from physicochemical quality. *Sci. Total Environ.* 584.
- Joimel, S., Schwartz, C., Bonfanti, J., Hedde, M., Krogh, P.H., P'er'es, G., Pernin, C., Rakoto, A., Salmon, S., Santorufu, L., Cortet, J., 2021. Functional and taxonomic diversity of Collembola as complementary tools to assess land use effects on soils biodiversity. *Front. Ecol. Evol.* 9, 630919.

- Köninger, J., Panagos, P., Jones, A., Briones, M., Orgiazzi, A., 2022. In defence of soil biodiversity: towards an inclusive protection in the European Union. *Biol. Conserv.* 268, 109475.
- Köninger, J., Beule, L., Tsiafouli, M., Seeber, J., Blasbichler, H., Sousa, J.P., da Silva, P. M., Frouz, J., Hedlund, K., Orgiazzi, A., Briones, M.J.I., Potapov, A., 2025. Molecular and morphological identification methods show contrasting trends in soil fauna diversity along land-use intensity gradients. *Appl. Soil Ecol.* 216, 106499. <https://doi.org/10.1016/j.apsoil.2025.106499>.
- Kuhn, M., 2008. Building predictive models in r using the caret package. *J. Stat. Softw.* 28, 1–26.
- Lavelle, P., Spain, A., Blouin, M., Brown, G., Decaëns, T., Grimaldi, M., Jimenez, J., McKey, D., Mathi, J., Velasquez, E., Zangerli, A., 2016. Ecosystem engineers in a self-organized soil: a review of concepts and future research questions. *Soil Sci.* 181.
- Lloyd, C.T., Soricchetta, A., Tatem, A.J., 2017. High resolution global gridded data for use in population studies. *Sci Data* 4, 1–17.
- Malinowski, R., Lewin'ski, S., Rybicki, M., Gromny, E., Jenerowicz, M., Krupin'ski, M., Nowakowski, A., Wojtkowski, C., Krupin'ski, M., Krätzschmar, E., et al., 2020. Automated production of a land cover/use map of Europe based on Sentinel-2 imagery. *Remote Sens.* 12, 3523.
- Marchal, D.F., G'érard, S., Hedde, M., Rougerie, R., Decaëns, T., 2022. An updated checklist and a DNA barcode library for the earthworms (Crassicitellata, Oligochaeta) of Corsica, France. *Zoosystema* 44, 439–461.
- Mathur, M., Mathur, P., 2023. Global distribution modelling of Macrophomina phaseolina (Tassi) Goid: A comparative assessment using ensemble machine learning tools. *Australas. Plant Pathol.* 52, 353–371.
- Mcbratney, A.B., Mendonça, Santos, M., Minasny, B., 2003. On digital soil mapping. *Geoderma* 117, 3–52.
- Medina-Sauza, R.M., Álvarez-Jiménez, M., Delhal, A., Reverchon, F., Blouin, M., Guerrero-Analco, J.A., Cerdán, C.R., Guevara, R., Villain, L., Barois, I., 2019. Earthworms building up soil microbiota, a review. *Front. Environ. Sci.* 7, 81.
- Meinshausen, N., 2006. Quantile regression forests. *J. Mach. Learn. Res.* 7, 983–999.
- Metzger, M., Bunce, R., Mucher, S., Watkins, J., 2005. A climatic stratification of the environment of Europe. *Glob. Ecol. Biogeogr.* 14, 549–563.
- Meyer, H., Pebesma, E., 2021. Predicting into unknown space? Estimating the area of applicability of spatial prediction models. *Methods Ecol. Evol.* 12.
- Meyer, H., Ludwig, M., Mil'á, C., Linnenbrink, J., Schumacher, F., 2024. The CAST Package for Training and Assessment of Spatial Prediction Models in R. *arXiv preprint. arXiv:2404.06978*.
- Milano, V., Maisto, G., Baldantoni, D., Bellino, A., Bernard, c., Croce, A., Dubs, F., Strumia, S., Cortet, J., 2018. The effect of urban park landscapes on soil *Collembola* diversity: a Mediterranean case study. *Landsc. Urban Plan.* 180, 135–147.
- Montanarella, L., Panagos, P., 2021. Soil security for the European Union. *Soil Security* 4, 100009.
- Nicolay, R., Jeffery, S., Randall, N., 2021. Comparison of three soil health indicators between different vegetative strip compositions. *Afr. J. Range Forage Sci.* 38, S104–S109.
- Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P., O'Hara, B., Simpson, G., Solymos, P., Stevens, H., Wagner, H., 2015. *Vegan: community ecology package. R Package Version 2.2-1*, 2, 1–2.
- Orgiazzi, A., Bardgett, R., Barrios, E., Behan-Pelletier, V., Briones, M., Chotte, J.-L., Deyn, G., Eggleton, P., Fierer, N., Fraser, T.D., Hedlund, K., Jeffery, S., Johnson, N., Jones, A., Kandel, E., Kaneko, N., Lavelle, P., Lemanceau, P., Miko, L., Wall, D., 2016. *Global Soil Biodiversity Atlas*. Publications Office of the European Union.
- Phillips, H., Guerra, C., Bartz, M., Briones, M., Brown, G., Crowther, T., Ferlian, O., Gongalsky, K., Hoogen, J., Krebs, J.J., Orgiazzi, A., Routh, D., Schwarz, B., Bach, E., Bennett, J., Brose, U., Decaëns, T., König-Ries, B., Loreau, M., Schmidt, O., 2019. Global distribution of earthworm diversity. *Science* 366, 480–485.
- Phillips, H., Bach, E., Bartz, M., Bennett, J., Beugnon, R., Briones, M., Brown, G., Ferlian, O., Gongalsky, K., Guerra, C., König-Ries, B., Krebs, J.J., Orgiazzi, A., Ramirez, K., Russell, D., Schwarz, B., Wall, D., Brose, U., Decaëns, T., Eisenhauer, N., 2021. Global data on earthworm abundance, biomass, diversity and corresponding environmental properties. *Sci Data* 8.
- Poggio, L., De Sousa, L., Batjes, N., Heuvelink, G.B.M., Kempen, B., Ribeiro, E., Rossiter, D., 2021. SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty. *SOIL* 7, 217–240.
- Potapov, A., Guerra, C., Hoogen, J., Babenko, A., Bellini, B., Berg, M., Chown, S., Deharveng, L., Kovac, L., Kuznetsova, N., Ponge, J.-F., Potapov, M., Russell, D., Alexandre, D., Alatalo, J., Arbea, J., Bandyopadhyaya, I., Bernava, V., Bokhorst, S., Scheu, S., 2023. Globally invariant metabolism but density-diversity mismatch in springtails. *Nat. Commun.* 14.
- Probst, P., Boulesteix, A.-L., Wright, M., 2018. Hyperparameters and tuning strategies for random forest.
- Remelli, S., Ghobari, H., Oliveira Filho, L.C.I., 2024. Editorial: the role of soil mesofauna as indicators of sustainable ecosystem management plans. *Front. Ecol. Evol.* 12.
- Römbke, J., Gardi, C., Creamer, R., Miko, L., 2016. Soil biodiversity data: actual and potential use in European and national legislation. *Appl. Soil Ecol.* 97, 125–133.
- Russell, D., Naudts, E., Soudzilovskaia, N., Briones, M., Çakir, M., Conti, E., Cortet, J., Fiera, C., Hackenberger, D., Hedde, M., Hobbeg, K., Indjic, D., Krogh, P.H., Lehmitz, R., Lesch, S., Marjanovic, Z., Mulder, C., Mumlade, L., Murvanidze, M., S. Rick, Roß-Nickoll, M., Schlaghamerský, J., Schmidt, O., Shelef, O., Suhadolc, M., Tsiafouli, M., Winding, A., Zaytsev, A., Potapov, A., 2024. Edaphobase 2.0: advanced international data warehouse for collating and using soil biodiversity datasets. *Appl. Soil Ecol.* 204, 105710.
- Rutgers, M., Orgiazzi, A., Gardi, C., Jansch, S., Keith, A., Neilson, R., Boag, B., Schmidt, O., Murchie, A., Blackshaw, R., P'er'es, G., Cluzeau, D., Guernion, M., Briones, M., Rodeiro, J., Piñero, R., Diaz Cosin, D.J., Sousa, J.P., Zwart, D., 2016. Mapping earthworm communities in Europe. *Appl. Soil Ecol.* 97, 98–111.
- Salako, G., Russell, D.J., Stucke, A., Eberhardt, E., 2023. Assessment of multiple model algorithms to predict earthworm geographic distribution range and biodiversity in Germany: implications for soil-monitoring and species-conservation needs. *Biodivers. Conserv.* 32, 2365–2394.
- Saulnier, L., Athias-Binche, F., 1986. MODALITE'S DE LA CICATRISATION DES E'COSYSTE'MES ME'DITERRANE'ENS APRE'S INCENDIE: CAS DE CERTAINS ARTHRO-PODES DU SOL 2. Les Myriapodes édaphiques. *Vie et Milieu/Life & Environment* 191–204.
- Sevgili, H., Özata, M.A., 2014. Checklist of the springtails (Hexapoda: Collembola) of Turkey. *Zool. Middle East* 60, 162–168.
- Tóth, B., Weynants, M., Pásztor, L., Hengl, T., 2017. 3D soil hydraulic database of Europe at 250 m resolution. *Hydrol. Process.* 31, 2662–2666.
- Topole, M., Marinko, J., Podpečan, V., Mylona, P., Debeljak, M., 2025. Barriers and prospects for soil biodiversity research in the Balkans. *Appl. Soil Ecol.* 206, 105804.
- Tsiafouli, M., Thébault, E., Sgardelis, S., Ruiter, P., Putten, W., Birkhofer, K., Hemerik, L., Vries, F., Bardgett, R., Brady, M., Björnlund, L., Jørgensen, H., Christensen, S., D'Hertefeldt, T., Hotes, S., Hol, G., Frouz, J., Liiri, M., Mortimer, S., Hedlund, K., 2015. Intensive agriculture reduces soil biodiversity across Europe. *Glob. Chang. Biol.* 21, 973–985.
- Tsiafouli, M., Cortet, J., Russell, D., 2023. Call for collaboration a call for collaboration to create the European atlas of soil fauna. *Soil Organisms* 175–181.
- Tuanmu, M.-N., Jetz, W., 2015. A global, remote sensing-based characterization of terrestrial habitat heterogeneity for biodiversity and ecosystem modeling. *Glob. Ecol. Biogeogr.* 24.
- Turner, R., Baddeley, A., 2005. SPATSTAT: an R package for analyzing spatial point patterns. *J. Stat. Softw.* 12.
- U.S. Geological Survey, 2023. USGS EROS Archive Vegetation Monitoring VIIRS Global NDVI. <https://www.usgs.gov/centers/eros/science/usgs-eros-archive-vegetation-monitoring-eviirs-global-ndvi>.
- Wadoux, A., Minasny, B., Mcbratney, A., 2020. Machine learning for digital soil mapping: applications, challenges and suggested solutions. *Earth Sci. Rev.* 210C, 103359.
- Wadoux, A.M.J.-C., Molnar, C., 2022. Beyond prediction: methods for interpreting complex models of soil variation. *Geoderma* 422, 115953.
- Wadoux, A.M.J.-C., Walvoort, D.J.J., Brus, D.J., 2022. An integrated approach for the evaluation of quantitative soil maps through Taylor and solar diagrams. *Geoderma* 405, 115332.
- Wadoux, A.M.J.-C., Courteille, L., Arrouays, D., Gomes, L., Cortet, J., Creamer, R., Eberhardt, E., Greve, M., Harhoff, R., Heuvelink, G., Krah, L., Lagacherie, P., Miko, L., Mulder, V., Pásztor, L., Pieper, S., Richer-de Forges, A., S'anchez-Rodríguez, A. R., & Wetterlind, J., 2024. On soil districts. *Geoderma* 452C, 117065.
- Wadoux, Alexandre M.J.-C., 2023. Interpretable spectroscopic modelling of soil with machine learning. *Eur. J. Soil Sci.* 74 (3), e13370.
- Wadoux, A.M.J.-C., Saby, N.P.A., Martin, M.P., 2023. Shapley values reveal the drivers of soil organic carbon stock prediction. *Soil* 9, 21–38.
- Wall, D., Bardgett, R., Covich, A., Snelgrove, P., 2004. The need for understanding how biodiversity and ecosystem functioning affect ecosystem services in soils and sediments. In: *Sustaining biodiversity and ecosystem Services in Soils and Sediments*, pp. 1–12.
- Wall, D.H., Nielsen, U.N., 2012. Biodiversity and ecosystem services: is it the same below ground. *Nature Education Knowledge* 3, 8.
- Wilson, R.J., Gutiérrez, D., Gutiérrez, J., Martínez, D., Agudo, R., Monserrat, V.J., 2005. Changes to the elevational limits and extent of species ranges associated with climate change. *Ecol. Lett.* 8, 1138–1146.
- Wright, M., Ziegler, A., 2015. Ranger: a fast implementation of random forests for high dimensional data in C++ and R. *J. Stat. Softw.* 77.
- Xue, P., Minasny, B., Wadoux, A.M.J.-C., Dobarco, M.R., Mcbratney, A.B., Bissett, A., de Caritat, P., 2024. Drivers and human impacts on topsoil bacterial and fungal community biogeography across Australia. *Glob. Chang. Biol.* 30, e17216.
- Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J., Sampson, C., Kanae, S., Bates, P., 2017. A high-accuracy map of global terrain elevations: accurate global terrain elevation map. *Geophys. Res. Lett.* 44.
- Zeiss, R., Eisenhauer, N., Orgiazzi, A., Rillig, M., Buscot, F., Jones, A., Lehmann, A., Reitz, T., Smith, L., Guerra, C.A., 2022. Challenges of and opportunities for protecting European soil biodiversity. *Conserv. Biol.* 36, e13930.
- Zomer, R., Xu, J., Trabucco, A., 2022. Version 3 of the global aridity index and potential evapotranspiration database. *Sci Data* 9, 409.