Geoderma 467 (2026) 117735

Contents lists available at ScienceDirect

Geoderma

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

Accounting for soil class variability in Mediterranean Europe using legacy soil maps and the topsoil LUCAS survey

G. Belvisi [a,*], C. Schillaci [b], L. Gristina [a], A. Delgado [c], D. Triantakonstantis [d,e], N. Lolos [e], M. Batsalia [e], A. Jones [b], L. Ribeiro Roder [f], C. Zucca [f], R. Scalenghe [a]

[a] Università degli Studi di Palermo, Palermo, Italy
[b] JRC European Commission, Ispra, Italy
[c] Department of Agronomy (ETSIA), University of Seville, Ctra. Utrera Km 1, Seville 41013, Spain
[d] Department of Sustainable Agriculture, University of Patras, 2 Seferi, 30100 Agrinio, Greece
[e] Institute of Soil and Water Resources, Hellenic Agricultural Organization—DIMITRA, 1 Sofokli Venizelou, 14123 Athens, Greece
[f] Department of Agricultural Sciences and Centre for Sustainable Management of Soil and Landscape (SMSL), University of Sassari, Viale Italia 39, Sassari 07100, Italy

ARTICLE INFO

Handling Editor: Jingyi Huang

Keywords:
Soil surveys
Mediterranean soils
Spatial representativeness
Legacy data
Sampling design
Soil variability

ABSTRACT

Mediterranean soils exhibit high pedological diversity, shaped by complex climatic, geomorphological, and anthropogenic drivers. Assessing this variability at the continental scale requires harmonized sampling frameworks. The LUCAS topsoil module provides standardized EU-wide topsoil physico-chemical and biological data, but its capacity to represent soil type diversity, particularly in heterogeneous Mediterranean landscapes, has not yet been evaluated. To perform this evaluation, this study compared LUCAS topsoil sampling points with legacy soil maps (LSMs) across multiple spatial scales (continental, national, sub-national) in Mediterranean Europe. We employed two complementary metrics: (i) the Average Absolute Coverage Difference (AACD), (ii) and a proportional mean index. Results show that continental-scale maps (e.g., WRB-FULL, WRB-LEV1) achieved AACD values below 1%, showing that sampling point distribution reflects well the soil type variability at this spatial scale and taxonomic levels. The performance was good also when sub-national maps with higher level of pedodiversity (Soil Map of Sicily) were used, with AACD between 1.2% and 1.7%. Overall, higher capacity to capture pedological diversity was observed in maps featuring higher soil diversity, larger spatial extent, and a greater number of soil map units (SMUs). However, some specific RSG, Histosols and Leptosols were underrepresented, whereas Calcisols were systematically oversampled. Overall, the LUCAS sampling design captures soil type variability at multiple scales, and incorporating local pedodiversity could reduce biases and improve compliance with the European Soil Monitoring Law. Legacy soil maps remain essential for optimising sampling design and extending harmonised soil monitoring to data-scarce regions such as the Near East and North Africa (NENA).

1. Introduction

Soils in the Mediterranean region are the result of a complex long-term interplay of unique climatic, geomorphological, and anthropogenic factors (Zdruli et al., 2010). In the central Mediterranean in particular, soil development has been shaped by unique natural processes, such as the accumulation of evaporitic deposits during the Messinian Salinity Crisis, with important outcrops in Sicily, as well as continuous inputs of desert dust – especially from the Sahara – which actively contribute to pedogenisis. Yaalon (1997) identified four key attributes characterizing Mediterranean soils: a specific seasonal climate, the presence of mountains, desert dust inputs, and a long history of human intervention.

Human influence on Mediterranean landscapes is especially pronounced. Land cover has been modified and even shaped for over 30,000 years, with intensive transformations documented during the Greek and Roman periods, including large-scale deforestation (Ellis et al., 2021), which led to significant soil erosion. In the last century, urbanization and infrastructure development have accelerated soil degradation, with soil sealing being among the most critical land cover change processes (Gardi et al., 2021; Carrubba and Scalenghe, 2012).

Recognizing the growing pressure on soil resources, the Global Soil

* Corresponding author.
E-mail address: giacomo.belvisi@unipa.it (G. Belvisi).

https://doi.org/10.1016/j.geoderma.2026.117735
Received 21 August 2025; Received in revised form 29 January 2026; Accepted 9 February 2026
Available online 21 February 2026

Partnership (GSP), promoted by FAO has identified soil data and information availability of as one of its five key pillars of action for mitigating the processes that threaten the soil resources. Despite this environmental and anthropogenic complexity, soil data across Mediterranean Europe remains patchy. Most available datasets are national or sub-national, developed for disparate objectives, and based on non-standardized sampling designs, spatial resolutions, and laboratory protocols. Furthermore, data accessibility is often limited or fragmented (Cornu et al., 2023; Jones et al., 2005; Morvan et al., 2008; van Leeuwen et al., 2017). This heterogeneity poses a major challenge for continental-scale monitoring, modelling, and evidence-based policy-making.

1.1. The LUCAS soil survey

In response to the need for harmonized land use and soil data, the European Commission launched in 2006 the Land Use and Coverage Area frame Survey (LUCAS). Coordinated by the European Statistical Office (Eurostat) with the Joint Research Centre (JRC). LUCAS was designed to provide standardized, field-based observations of land cover and land use across the European Union (EU). Since 2009, LUCAS has introduced a soil module, collecting topsoil samples according to harmonized protocols (Tóth et al., 2013). The LUCAS soil surveys (2009/12, 2015, 2018, 2022) represent a milestone in continental-scale soil monitoring, offering consistent, comparable, and freely accessible datasets on physical and chemical soil properties. More recently, additional parameters such as bulk density, biodiversity indicators, and erosion assessments have been integrated into the sampling protocol (Pacini et al., 2023, Chen et al., 2024, Panagos et al., 2024).

The LUCAS 2015 soil module aimed to resample approximately 90% of the sites surveyed in 2009 while extending coverage to all 28 EU Member States and to areas above 1000 m altitude. The remaining 10% of sampling locations were relocated within each Member State. Overall, around 22,000 samples underwent analysis for pH, soil organic carbon (SOC), and soil nutrient key elements (N, P, K), with electrical conductivity (EC) also measured in 2018 (over 23,000 samples if including non-EU countries) (Schillaci et al., 2025).

While LUCAS offers a unique framework for soil data harmonization, its sampling design was originally optimized for land cover monitoring, without considering at all the soil type distribution. As a result, the spatial pattern and density of soil observations, approximately one sample every 400 km^2, may be insufficient to fully capture the intrinsic spatial variability of soils, particularly in heterogenous and geomorphologically complex regions such as the Mediterranean.

1.2. Soil legacy data (SLD)

Soil legacy data (SLD) refers to historical soil information collected using varying sampling strategies, analytical methods, standards, or technologies. SLD include observations on soil physical, chemical, and biological properties, as well as soil map products derived from past surveys and monitoring programs. Despite potential limitations in consistency or resolution compared to modern datasets, SLDs play crucial role in soil science. They provide historical baselines for assessing long-term changes, enable retrospective analyses, support trend detection related to environmental or land-use dynamics, and can complement contemporary datasets when integrated appropriately. However, SLDs are frequently stored in non-digital formats, physically contained in old libraries and not accessible to the public (Rossiter, 2008; Arrouays et al., 2017; Cambule et al., 2015; Rasaei et al., 2020; Dent and Ahmed, 1995; Sarmento et al., 2014). A time-consuming digitization work is often necessary to recover data and preserve valuable information (Arrouays et al., 2017). At the European level, the European Soil Data Centre (ESDAC) hosts a large collection of georeferenced SLDs, including several soil maps made publicly available through the EUSDAM repository (https://esdac.jrc.ec.europa.eu/resource-type/national-soil-maps-eudasm).

1.3. LUCAS sampling design

The LUCAS topsoil survey builds upon the general LUCAS sampling framework, which is based on a regular 2 x 2 km grid covering the EU. Approximately 10% of the general LUCAS points are selected for soil sampling, ensuring proportional coverage of the main land cover types, including cropland annual and permanent crops, woodland, shrubland and grassland.

To ensure the effective representativeness of soil monitoring in the first survey, a multi-stage stratified random sampling method, as proposed by McKenzie et al. (2008), was adopted (Tóth et al., 2013). Stratification was primarily driven by land cover and topographic information, allowing the definition of landscape strata from which candidate sampling locations were selected. Operational flexibility during fieldwork was ensured through the use of alternative sampling locations. The core principles of the first LUCAS topsoil sampling design are summarized in Fig. 1, while detailed operational aspects are described in Tóth et al. (2013). Interestingly, recognizing the availability of a good quantity of soil data for forest areas (Hiederer and Durrant, 2010), LUCAS strategy redistributed a portion of forest points to arable land and grassland areas, so the first survey was not purely design-based.

1.4. Land cover vs soil type in sampling design applications

The LUCAS topsoil sampling design is strongly anchored in land cover classes. Within the SCORPAN framework (McBratney et al., 2003), progression of the Jenny (1941) seminal model, soil properties can be expressed as a function of environmental and spatial factors:

$S_c=S_a = f$ (s,c,o,r,p,a,n) (1)

where Sc = soil classes, Sa = soil attributes, s = soil, other soil properties, c = climate, o = organisms, r = topography, p = parent material, a = time, n = space. Land cover is undoubtedly a visible attribute, observable even via satellite, unlike soil types. Designing

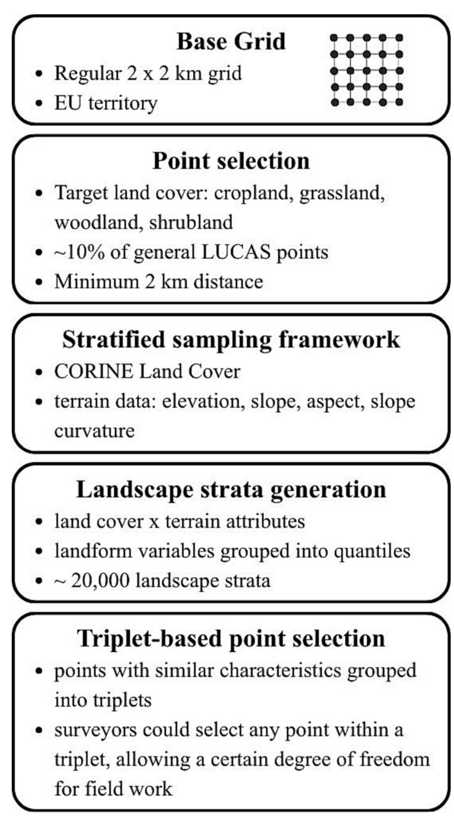

Fig. 1. Schematic representation of the first LUCAS topsoil sampling design.

extensive soil sampling campaigns based on land cover can be an effective and pragmatic solution when direct information on soil types is limited or unavailable.

The most widespread soil classifications systems are the Soil Taxonomy (Soil Survey Staff, 2022) and the World Reference Base (IUSS Working Group WRB, 2022). Several countries have developed local classification systems to complement WRB and ST, enabling more detailed characterization of their soils under specific regional conditions.

LUCAS represents one of the world's largest and harmonized continental-scale soil legacy dataset, freely-available to everyone with prior registration to the European Soil Data Centre (ESDAC), providing standardized measurements of soil properties collected using consistent protocol. While its sampling design has proven effective for continental-scale soil monitoring, the growing availability of legacy soil data and maps offers new opportunities to assess how well land cover driven sampling strategies capture soil type variability across different spatial scales, particularly at national and sub-national levels.

The representativeness of digital soil mapping methods, including several exercises carried out using the LUCAS topsoil dataset, in capturing real-world soil variability has been critically examined in previous studies (Rossiter et al., 2022; Rossiter & Poggio, 2025). In this study, such representativeness is explicitly addressed in terms of proportional correspondence between the areal distribution of soil map units and the density of sampling points. To evaluate to what extent the LUCAS soil module captures soil variability, Legacy Soil Maps (LSMs) represent an invaluable reference. These historical polygon maps, though heterogeneous in methodology and resolution, provide spatially explicit information on the distribution of mapped soil classes at the level of soil map units (SMUs). In most products, an SMU is an association or complex of several soil typological units (STUs); the legend reports the dominant or co-dominant soil class(es), which we use here as a proxy for the main mapped soil type in that polygon, without assuming internal homogeneity. Through this paper, "soil type variability" therefore refers to variability in mapped soil classes at map-unit level (e.g. dominant WRB Reference Soil Group), rather than to the full STU composition within each SMU (Soil Science Division Staff, 2017; Van Wambeke and Forbes, 1986).

The aim of this work is twofold:

(i) to analyze how well LUCAS topsoil sampling locations reproduces the distribution of mapped dominant soil classes in a set of legacy soil polygon maps (LSMs), and

(ii) to propose a simple, map-based methodology to evaluate the capacity of continental soil surveys (e.g., LUCAS) to capture mapped soil class variability at different spatial scales, recognizing that SMUs are often internally heterogeneous

In doing so, we provide recommendations for improving soil survey design and offer insights to support the potential extension of LUCAS to adjacent Mediterranean regions, particularly the Near East and North Africa (NENA), through harmonized methodologies.

2. Materials and methods

2.1. Study area

This study focuses on the Mediterranean region, including several southern EU Member States (Italy, Spain, and Greece) and associated islands (e.g., Sicily, Crete, Cyprus), where soil data surveys are often scattered and less harmonized compared to Northern Europe (Fig. 2). These areas exhibit pronounced soil heterogeneity due to their unique biophysical conditions and historical land use patterns. The insular territories were selected because they constitute naturally bounded systems with well-defined ecological and pedological gradients, with natural sharp borders. The delimitation of the study area in this way is a natural constraint and not an administrative choice.

2.2. Soil dataset

For this study, we used the three publicly available LUCAS datasets from 2009/12, 2015, and 2018, retrieved from the European Soil Data Centre (ESDAC) online repository. Each campaign followed a standardized sampling protocol, generally targeting around 20,000-30,000 locations, with plans for the 2022 survey aiming for 41,000 (Tóth et al., 2013; Panagos et al., 2022). A set of legacy soil maps (LSM) was compiled, covering continental, national, and sub-national scales. These maps differ in terms of resolution, classification system, and thematic focus. Sources included national soil inventories, FAO (FAO 1974) and WRB-based thematic maps, and ESDAC's EUDASM archive. Table 1 lists all maps used, ranging from the Soil Map of Europe (1:1,000,000, WRB 1998/2006; FAO/ISRIC/ISSS, 1998; IUSS Working Group WRB, 2006)) to detailed regional maps such as the Soil Map of Sicily (1:250,000) and the Soil Map of Viotia prefecture, Greece (1:100,000, Soil Survey Staff, 1975; USDA, 1975)). All maps were converted to vector format and georeferenced to the ETRS89 LAEA projection (EPSG: 3035) for spatial analysis. All maps used are either official national/regional products or ESDAC/EUDASM datasets currently maintained by the competent authorities.

2.3. Data extraction

The continental-scale Soil map covers more countries than those involved in LUCAS, so it was cropped to cover the LUCAS countries only, using the administrative country borders (ESTAT, NUTS 2021 at 01 M scale). Croatia was included in the LUCAS soil survey only from the second survey (2015) and for this reason was not included in the first campaign computation (LUCAS 2009–12).

LUCAS sampling points were overlaid on each LSMs to assign the soil type classification: WRB-Reference Soil Groups, Soil Taxonomy-Orders or other National soil classifications. For each soil map unit (SMU) and for each comparison, we calculated: i) the proportion of the total area it occupies in the study region; ii) the proportion of LUCAS location falling within it.

2.4. Metrics

To quantify the ability of the given soil points data to capture soil type variability at different scales, we applied two metrics: the Average Absolute Coverage Difference (AACD) and a proportional mean index. The unit of comparison is the soil map unit (SMU), here considered as a georeferenced polygon that delineates areas with certain defined soil characteristics, in particular the same soil type as defined in the respective classification (WRB, ST, local classifications). However, we must point out that in the European WRB maps, the dominant Reference Soil Group (RSG) for each SMU corresponds to the soil type that represents the largest share within the unit, which does not necessarily imply exclusivity. In the European WRB maps, the dominant RSG for each SMU corresponds to the soil class that occupies the largest share of the polygon area, which may be well below 50% in associations or complexes (IUSS Working Group WRB, 2022). This must be kept in mind when interpreting discrepancies between point-based observations and mapped units.

In this study we define representativeness as the agreement between (i) the areal proportions of mapped soil classes in a given legacy soil map and (ii) the proportions of LUCAS soil sampling points falling in SMUs labelled with those classes, within the same spatial domain. The AACD and mean index therefore evaluate how well the LUCAS network reproduces the distribution of mapped dominant soil classes, rather than the within-polygon frequency of all STUs or the vertical extent of the soil body.

For each *i*-th SMU we compute "points – surface Δ", i.e. the difference between the proportion of LUCAS points falling within the *i*-th SMU

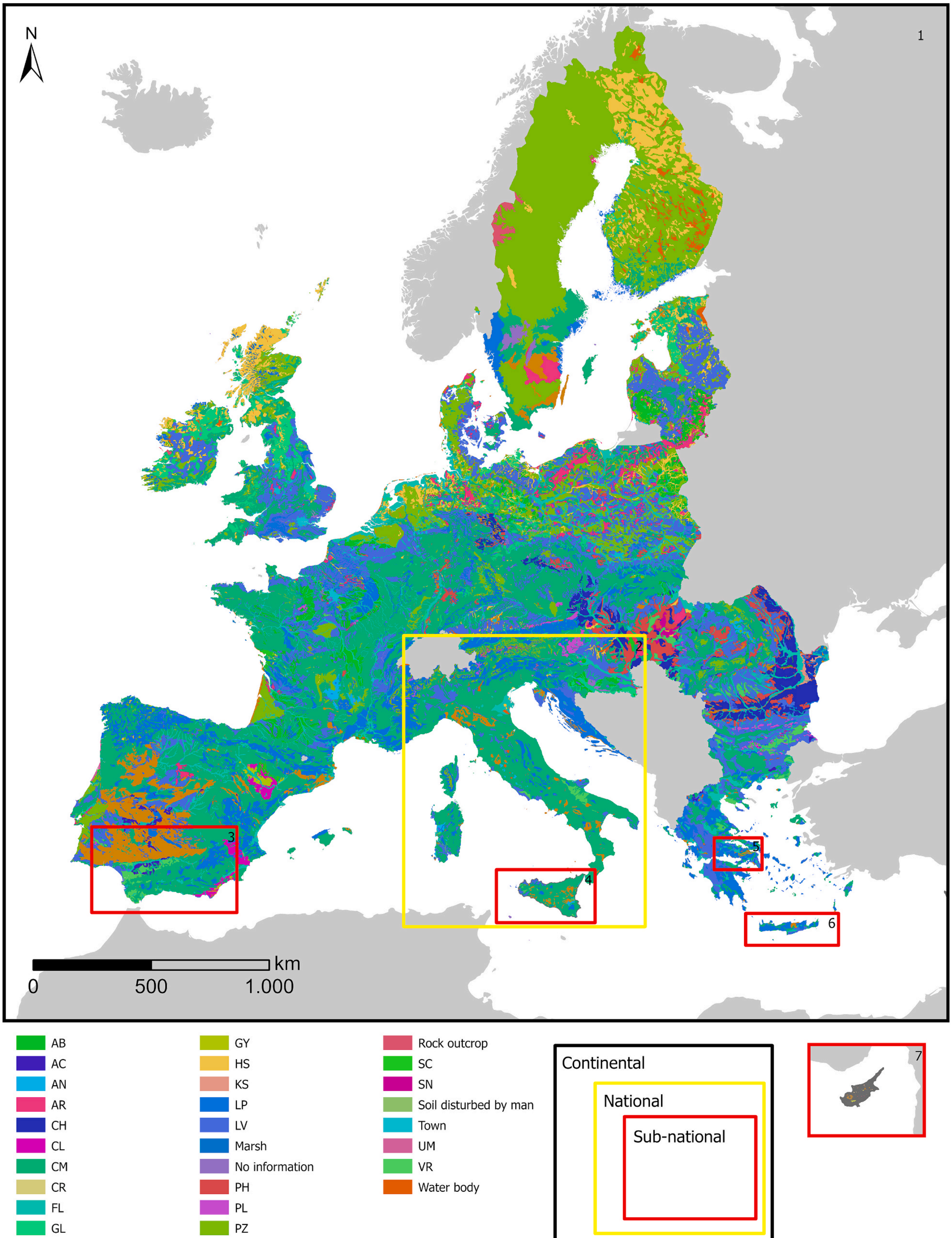

Fig. 2. Overview of the soil-legacy maps used in this study. The continental base layer corresponds to the WRB LV1 (RSG) soil map at 1: 1 000 000 scale. Coloured bounding boxes outline the areas for which more detailed legacy maps were examined: (1) Europe (continental extent, black); (2) Italy (national extent, yellow); and the sub-national domains (red) (3) Andalusia, Spain; (4) Sicily, Italy; (5) Viotia, Greece; (6) Crete, Greece; (7) Cyprus.

Table 1
Compilation of soil maps in the Mediterranean: scales and classification systems.

Map	Year	Area	Geographic scale	Nominal scale	Soil classification system
[a]Europe	1998	Europe	continental	1:1,000,000	WRB 1998
[2]WRB-FULL	2024	Europe	continental	1:1,000,000	WRB 2006
[3]WRB-LEV1	2024	Europe	continental	1:1,000,000	WRB 2006
[4]Soil Map of Italy	2012	IT	national	1:1,000,000	WRB 2006
[5]Soil Map of Crete	1986	Crete, GR	sub-national	1:1,000,000	FAO 1974
[6]Soil Map of Andalusia	1989	Andalusia, ES	sub-national	1:400,000	FAO 1974
[7]Soil Map of Sicily	1988	Sicily, IT	sub-national	1:250,000	FAO 1974
Reclassified Sicily Map	2024	Sicily, IT	sub-national	1:250,000	WRB 2022
[8]General Soil Map Map of Cyprus	1970	Cyprus	sub-national	1:200,000	FAO 1974
[9]Soil Map of Viotia prefecture	1992	Viotia, GR	sub-national	1:100,000	Soil Taxonomy 1975

[a] Soil map of Europe; [2, 3]https://esdac.jrc.ec.europa.eu/content/european-soil-database-v2-raster-library-1kmx1km European Soil Database v2 Raster Library 1 km x 1 km – Full soil code from the World Reference Base (WRB) for Soil Resources (2) and Soil reference group code (3); [4]https://esdac.jrc.ec.europa.eu/content/carta-dei-suoli-ditalia-soil-map-italy; [5]https://esdac.jrc.ec.europa.eu/content/soil-map-crete; [6]https://portalrediam.cica.es/descargas?path=%2F04_RECURSOS_NATURALES%2F02_GEODIVERSIDAD%2F04_SUELOS; [7]Fierotti et al. (1988); [8]https://esdac.jrc.ec.europa.eu/content/general-soil-map-map-cyprus; [9]Misopolinos et al., 2015.

$\left(\frac{numberofLUCASpointsinSMU_i}{totalnumberofLUCASpointsinthestudyarea}\right)$ and the share of the total surface covered by the i-th SMU $\left(\frac{surfaceareaofSMU_i\left[km^2\right]}{totalsurfaceareaofthestudyarea\left[km^2\right]}\right)$. A negative value indicates that LUCAS underrepresents that specific SMU, whereas a positive value indicates that the density of LUCAS points in that SMU is higher than the proportion of that SMU in terms of surface area.

Then, for each comparison and for each LUCAS surveys (2009–12, 2015 and 2018) plus the overall situation (total, i.e. the sum of the three LUCAS surveys), we calculated a newly defined metric, the Average Absolute Coverage Difference, AACD (eq. (2), i.e. the mean of the absolute value of the previous metric Δ points – surface, to figure out a single number to represent the overall performance of the LUCAS surveys in representing soil type variability.

$$AACD = \frac{1}{n}\sum_{i=1}^{n}|\Delta_i| = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{P_i}{P_{tot}} - \frac{A_i}{A_{tot}}\right| \tag{2}$$

Where n = total number of SMUs in the map; P_i = number of LUCAS points in the i-th SMU; P_{tot} = total number of LUCAS points in the map; A_i = surface area of the i-th SMU; A_{tot} = total surface area of the map.

The closer the value is to zero, the better the LUCAS distribution represents the soil type. However, this metric can undervalue small variations for smaller map units. To avoid this, we calculated an average index (hereafter referred to as "mean index", eq. (3) derived from the ratio between the percentage of points to the percentage of surface for each SMU. In this case, the ideal value is 1, with which the LUCAS network perfectly represents the soil type variability. These metrics identify biases or discrepancies in sampling at different spatial scales.

$$meanindex = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{P_i/P_{tot}}{A_i/A_{tot}}\right| \tag{3}$$

Additionally, the Shannon diversity index (H′) was used to evaluate the diversity of SMUs, providing context for interpreting sampling representativeness in relation to the complexity of the landscape. Originally developed by Shannon (1948) in the context of information theory, the Shannon index – also known as Shannon entropy – has been widely adopted in ecology and soil science to quantify diversity. It accounts for both the abundance and evenness of categories (e.g., species, molecular classes, or soil types), with higher values indicating greater diversity. In soil science, it has been applied to evaluate pedodiversity (Vašát et al., 2023), soil microbial communities (Su et al., 2024; He et al., 2023), and soil organic matter complexity (Drosos et al., 2023), offering insights into ecosystem functioning, nutrient cycling, and long-term soil stability. Its versatility also extends to the design of soil information systems and digital soil mapping frameworks (Murphy, 2022). The Shannon diversity index H′ was calculated to assess the diversity of soil map units within the study area, considering both the number of map units and their proportional representation:

$$H' = -\sum_{i=1}^{N} p_i \bullet \ln(p_i) \tag{4}$$

where p_i = % surface/100, the proportion of the total surface area occupied by the i-th soil map unit; N = the total number of map units. The H' captures the balance and richness of SMUs. It is used here to complement AACD by providing insight into the overall spatial diversity of the soils within the study area. Higher H' values indicate greater diversity, with more SMUs that are more evenly distributed. Lower H′ values suggest dominance by fewer map units or a highly uneven distribution of areas across soil types. Together, these metrics offer a robust framework for assessing the suitability of the LUCAS dataset in capturing soil variability. AACD and Mean Index quantify different aspects of proportional agreement between soil maps and sampling density. However, neither index alone provides a complete assessment of representativeness, and both require critical interpretation. Firstly, we compared the results of LSMs at different spatial scales (continental, national, sub-national). Then, to better understand the relationships between Shannon Index, spatial scale and number of the number of map units, we focused on a case study in Sicily, the largest island in the Mediterranean Sea. We compared the performance of LUCAS with respect to four different soil class maps (Fig. 3): 1) Soil Map of Sicily (Fierotti et al., 1988, 1: 250,000), 2) the Sicily sheet of the soil map of Italy (1: 1,000,000), 3) a WRB 2022-based reclassified version of the Soil Map of Sicily (conversion table in Supplementary Information), 4) the Sicily part of the soil map of Europe LV1 (1: 1,000,000). Only for the traditional soil maps, we calculated the Average Size Delineations (ASD), the Minimum Legible Area (MLA), the index of maximum reduction (IMR) and the effective scale, following the procedures described by Forbes et al. (1987), to check whether the declared nominal scale accurately correspond to the real scale.

2.5. Software

Spatial analyses, including zonal statistics and point-in-polygon operations, as well as data visualization, were performed using QGIS 3.34.7 Long Term Release (LTR) and ArcGIS Pro 3.4.2. Statistical analyses were conducted in R (version R 4.3.1) utilizing base R and the tidyverse package collection.

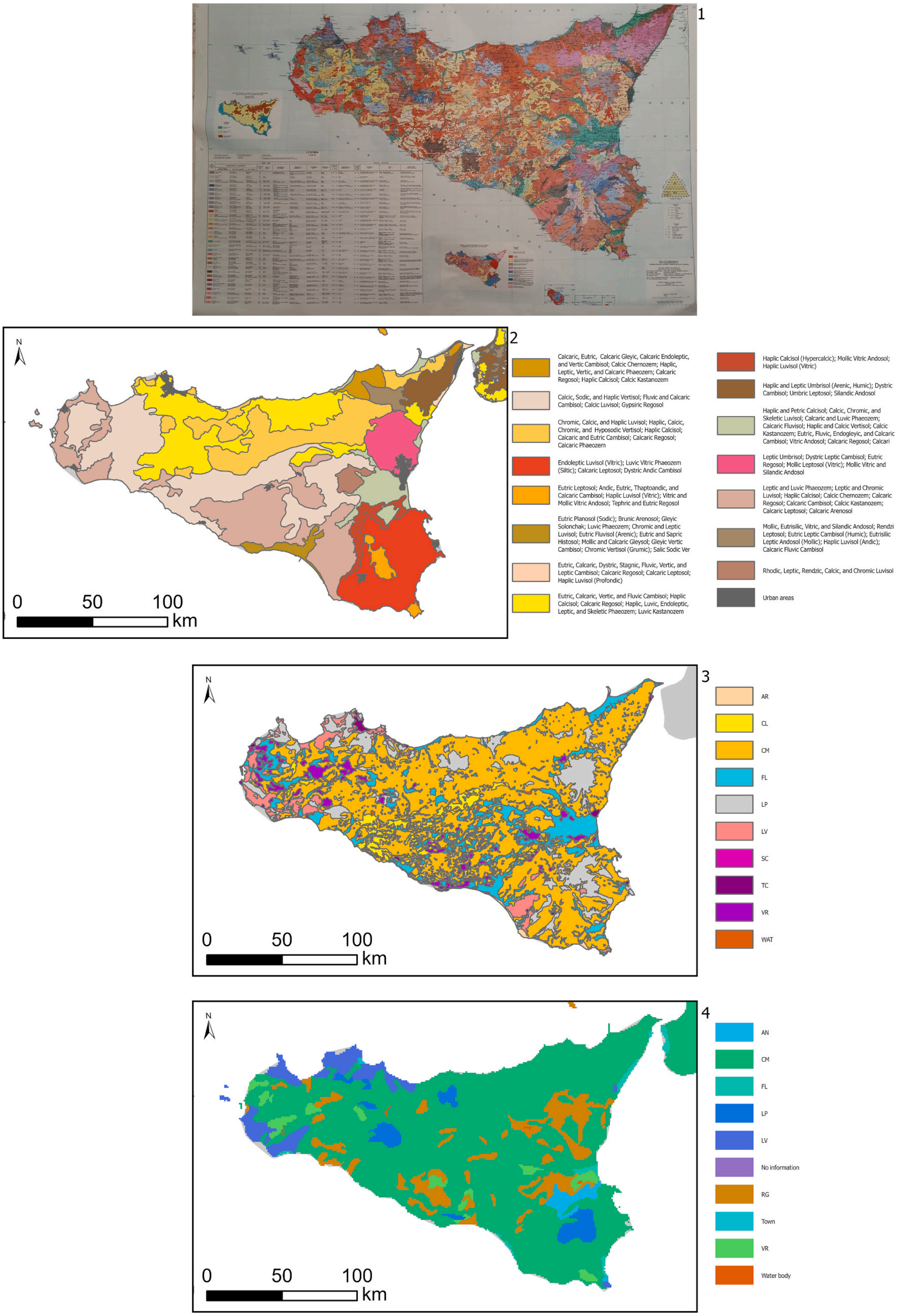

Fig. 3. Overview of the legacy map used for Sicily. From top to bottom: 1) Soil Map of Sicily; 2) Clip from Soil Map of Italy; 3) Reclassified Soil Map of Sicily; 4) clip from WRB LEV1.

3. Results

3.1. Δ points–surface differences across map scales

The difference between the percentage of LUCAS points and the percentage of area covered by each SMU – hereafter Δ points – surface – tends to increase as map scale increases. Broad-scale continental maps (small scale, 1:1,000,000) showed values tightly clustered around zero, indicating relatively LUCAS survey's good capacity to represent soil types at this scale. In contrast, large-scale maps displayed more scattered distributions. These trends are shown in the boxplots in Fig. 4, with maps ordered by increasing nominal scale. The most pronounced outlier was the Map of Crete (1986), which presented both interpretative and digitization challenges due to inconsistent SMUs definitions, duplication, and the presence of additional categories such as "stony". Furthermore, it is a small-scale map (1:1,000,000) covering a small area (~8000 km^2).

3.2. Average Absolute Coverage Difference (AACD)

The Average Absolute Coverage Difference (AACD) provides a synthetic measure of sampling representativeness across map scales, classifications, and campaigns (Figs. 5 and 6). Continental-scale maps such as the WRB-FULL and WRB-LEV1 consistently recorded low AACD values (<1%), indicating that LUCAS sampling effectively captured dominant soil types at this level. AACD tended to increase with higher resolution, smaller map extent, and lower soil diversity. For instance, sub-national maps of Viotia prefecture and Crete, both with fewer than 10 soil units, exhibited AACD values above 3%. In contrast, maps with high Shannon Index values (>2.5), such as the original Soil Map of Sicily (1988), performed well despite their higher complexity. A clear trend emerged when examining the AACD relative to: i) Map scale: coarser scales generally corresponded to higher AACD; ii) Map surface: larger areas showed better performance; iii) Number of SMUs: more units typically led to lower AACD; iiii) Soil diversity (H$'$): higher diversity (higher Shannon index) corresponds with better LUCAS representativeness. These patterns highlight the challenges of maintaining representativeness at finer resolution, especially in pedologically complex regions.

3.3. Sicily case study: comparing soil map representations

Comparing four soil maps of Sicily with varying complexity reveals how the capacity of LUCAS points to represent the mapped soil type is influenced by scale, classification systems, and soil diversity. The original Soil Map of Sicily (1:250,000) included 36 SMUs ("soil associations") and had the highest Shannon Index (H$'$ = 3.12) among the four maps considered. This indicates a rich variety of soil types. In contrast, the Soil Map of Italy (Sicily sheet) at 1:1,000,000 scale, which uses a more general classification system, has only 16 map units and a Shannon Index of 2.23. Both the Reclassified Soil Map of Sicily and the Soil Map of Europe LV1 are simpler, each having 10 SMUs. However, their diversity levels differ slightly, with Shannon Index values of 1.39 and 0.98, respectively. In the case of the Soil Map of Sicily, the AACD values range between 1.2% and 1.7% across the different LUCAS campaign years (Fig. 7). The Reclassified Soil Map of Sicily has slightly higher AACD values (around 2%), rather constant over time. The Soil Map of Europe, despite being the least detailed, has the lowest AACD values (0.5% to 1.4%). The scale of the map and the associated categorical level of the classification system used for the SMU, and whether the SMU are compound (associations) or simple (consociations) as is possible at more detailed scales, clearly affect the AACD. In the case of the two maps at 1:250,000 scale, the original version performs better than the reclassified one. However, at 1:1,000,000, the Soil Map of Europe performs better than the Soil Map of Italy, suggesting that scale alone doesn't determine the result and that other factors have a role, for example, the categorical level of the legend. Maps with higher Shannon Index values, indicating more soil diversity, generally show more stable AACD values over time. However, the Soil Map of Europe has the lowest diversity but still achieves better AACD values.

3.4. Mean index vs AACD

Examining the WRB Soil Map of Europe, with its 25 SMUs, the Δ points-surface varies between a maximum of + 3.18% (Calcisols, overrepresented) and −1.93% (Histosols, underrepresented). Most comparisons fall very close to 0% (see SI). The AACD is 0.56%, one of the best results observed. SMUs where Leptosols and Histosols are the dominant soil type, tend to be underrepresented compared to other soil types. Underrepresentation of Leptosols can be influenced by the initial decision of LUCAS to exclude points above 1000 m in elevation. The Leptosols are shallow, poorly developed soils typical of mountainous

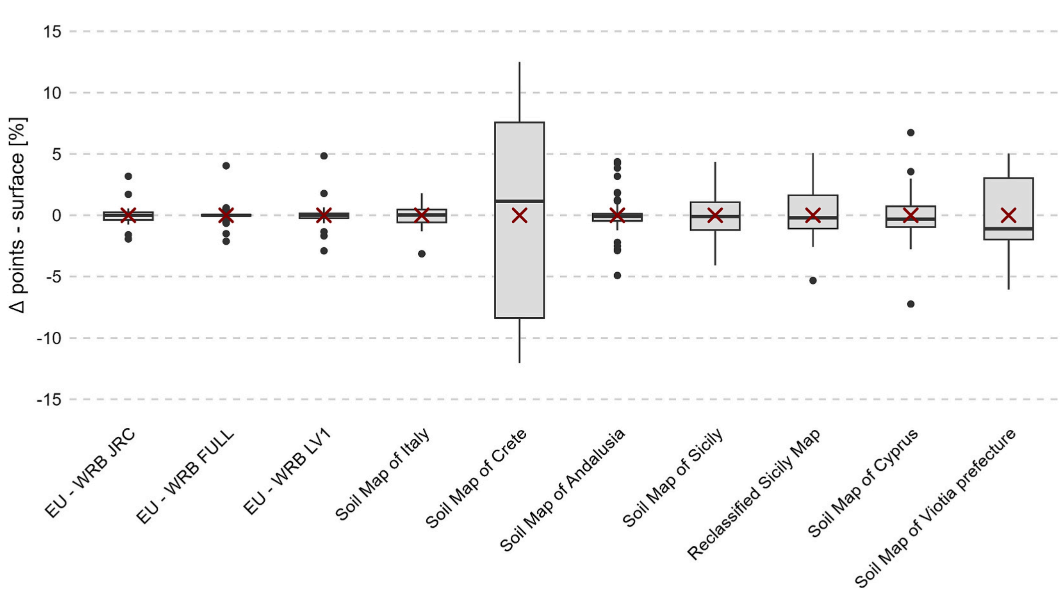

Fig. 4. Boxplots of Δ points – surface [%] across all soil maps analysed, ordered by increasing nominal scale.

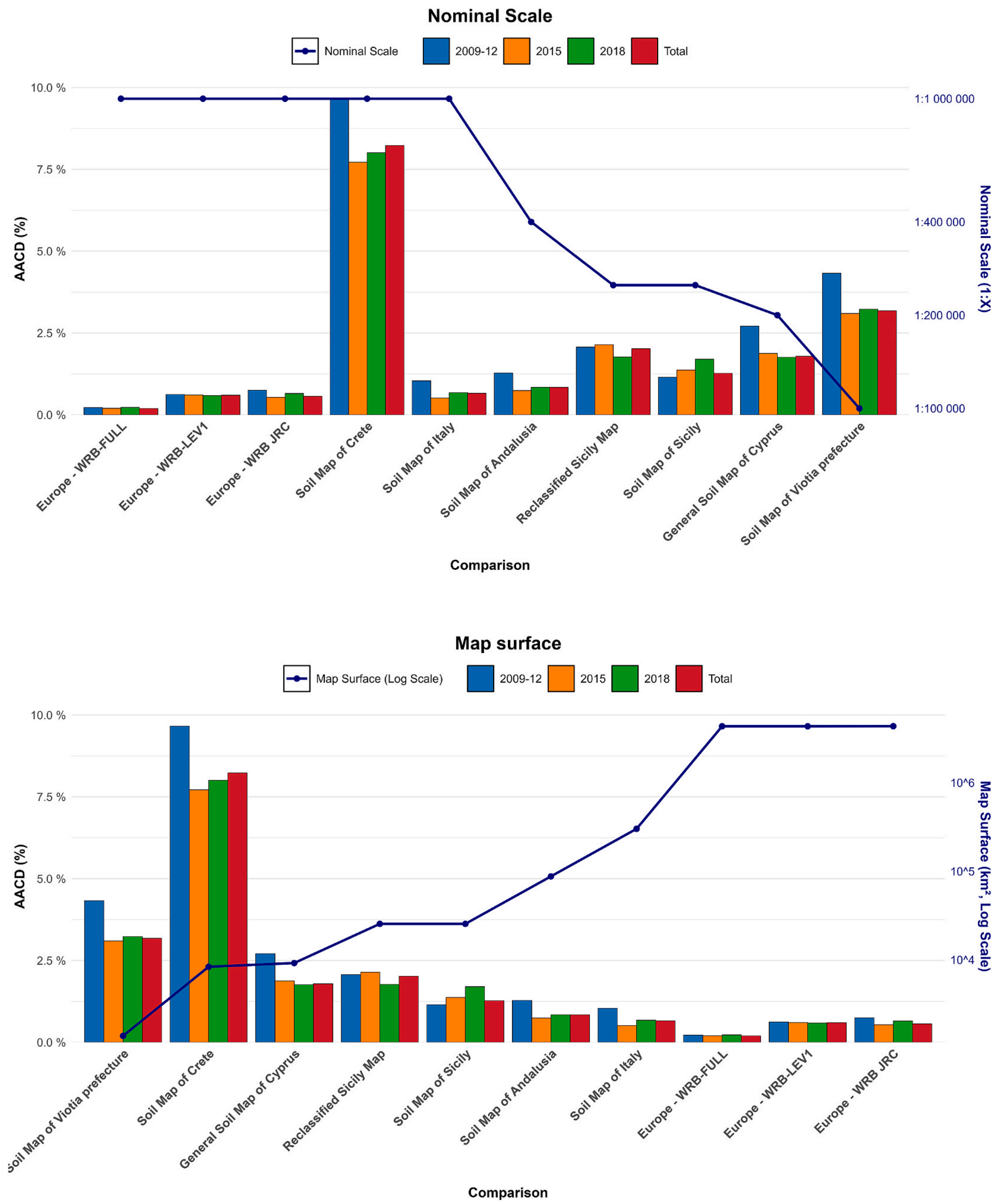

Fig. 5. Average Absolute Coverage Difference (AACD) for nine European soil maps, broken down by LUCAS survey campaign. Panel above: Grouped bars show AACD (%) for each map in the 2009/12 (blue), 2015 (orange), 2018 (green) campaigns and their overall total (red). Maps are arranged from the smallest-scale continental products on the left (Europe – WRB FULL, WRB LEV1, WRB) to the largest-scale local survey (Soil Map of Crete) on the right. The blue polyline superimposed on the bars tracks the nominal spatial scale (denominator of the representative fraction, 1: X) and is read on the secondary y-axis.

regions. In subsequent campaigns, with the inclusion of points above 1000 m, the representation of Leptosols improved significantly. This was not the case for Histosols, soils with a high level of organic carbon and concentrated only in specific areas, which remained slightly underrepresented in the other two campaigns as well. The Podzols, initially overestimated, became less sampled and statistically underrepresented in the 2015 and 2018 campaigns. A significant overestimation of Luvisols was observed in the first survey, decreasing in subsequent years. Only Calcisols remained consistently overestimated, with the degree of overrepresentation increasing in time along the surveys. Apart from these exceptions, the remaining major soil groups appear to be adequately represented. It should be noted that the AACD loses statistical power, particularly when comparing soil map units with reduced surface area. Since it is calculated as the mean of differences between percentages, it does not account for the fact that small variations can result in disproportionately large differences in smaller soil map units. For example, in the Soil Map of Europe, Gypsisols represent 0.28% only of the total surface area, while the corresponding LUCAS points for this soil map unit constitute 0.61%. Calculating their Δ points-surface yields + 0.33%, a low value not far from the optimal value (0%). However, the mean index (based on the average of ratios % points / % surface) for this specific case is 2.21, the most divergent value observed, precisely because it indicates that Gypsisols have been largely oversampled (twice as much as the proportion needed to balance their area). Table 2 summarises the evaluation metrics used to evaluate the quality and accuracy of the soil legacy maps. Comparing the proposed mean index with the AACD, there are no substantial differences in the magnitude of values (Table 2). The only notable difference is that the LUCAS points appear to

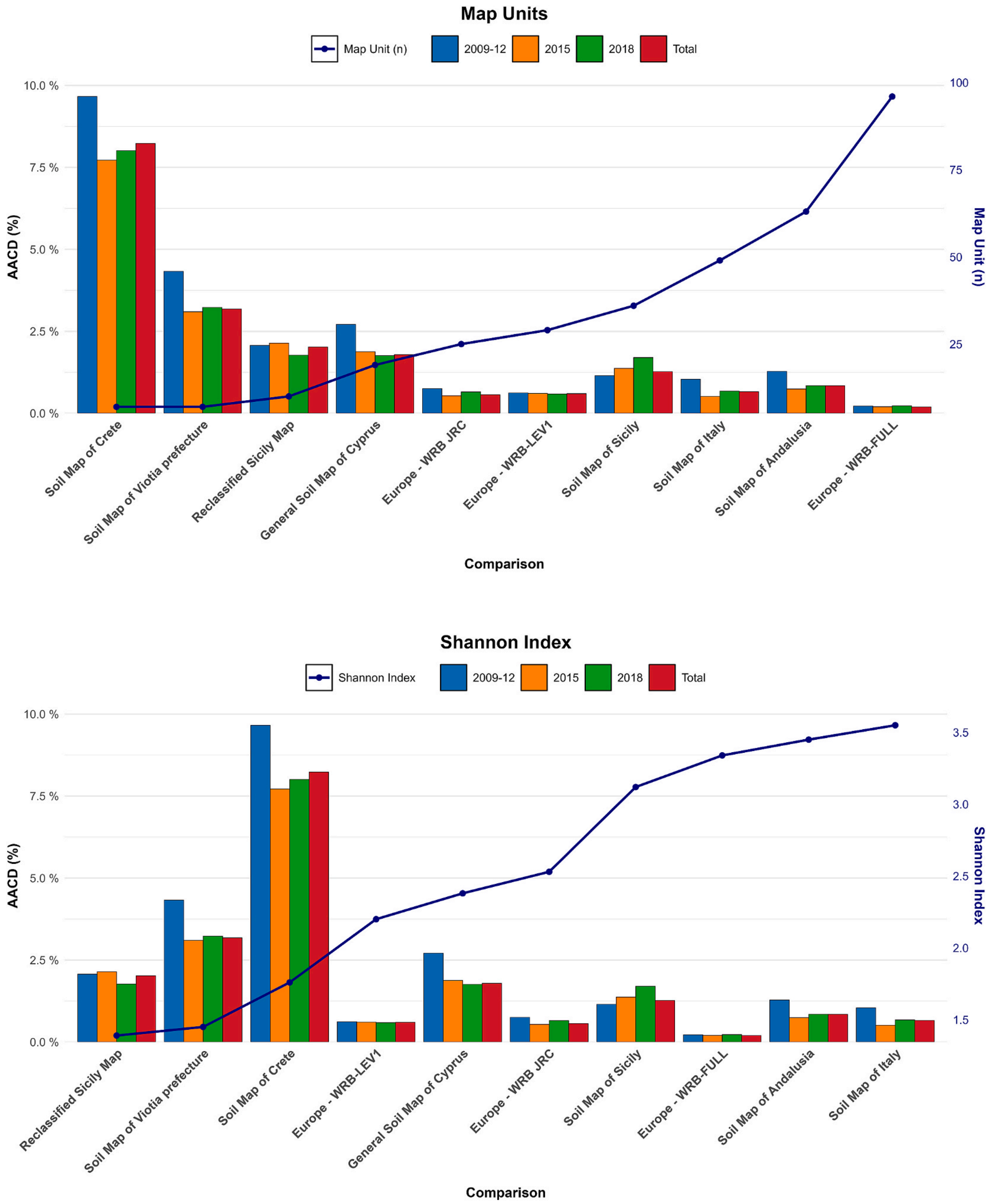

Fig. 6. AACD values for each LUCAS campaign (2009/12, 2015, 2018) and total across different soil maps. On top maps are ordered for increasing number of map units and on the bottom for increasing Shannon index.

be better distributed in the case of the Soil Map of Sicily with the “mean index” compared to the AACD.

To further diagnose the nature of the observed sampling biases, AACD and Mean Index were interpreted jointly. For this purpose, AACD values were grouped into three classes (low $\leq$ 1%, moderate 1–3%, high $>$ 3%), while deviations from proportionality were assessed using $|1 -$ Mean Index$|$, classified as low ($<$ 0.25) or high ($\geq$ 0.25). For each comparison, the two most over- and under-sampled soil units were then identified (Table 3). Several comparisons characterized by low AACD values also exhibit systematic over- and under-representation of specific soil units. At continental scale (e.g. Europe–WRB JRC, WRB-FULL, WRB-LEV1), widespread soil groups such as Cambisols, Calcisols, and Luvisols tend to be over-sampled, whereas less frequent or environmentally constrained soils, including Leptosols, Histosols, and Podzols, are consistently under-sampled. High AACD and high deviation values are observed for smaller and less diverse mapping units, such as Crete and the Viotia prefecture, where a small number of dominant soil types drives large proportional mismatches.

4. Discussion

4.1. LUCAS soil class representativeness

The assessment of soil class variability using LUCAS reveals several insights into the challenges and opportunities of representing soil characteristics at varying spatial scales (Fatichi et al., 2020). While for coarse-scale maps, like WRB-FULL and WRB-LEV1, the LUCAS topsoil survey, with its even coverage of the continental area, adequately captures the SMUs, as evidenced by low AACD values ($<$1%), the representativeness decreases significantly in finer-resolution sub-national

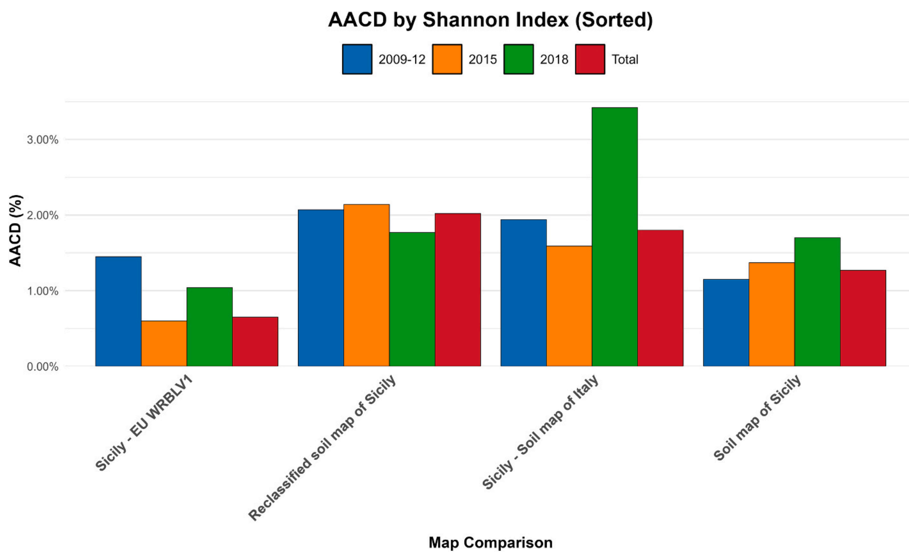

Fig. 7. AACD computed for four thematic soil maps of Sicily and three LUCAS sampling campaigns (blue = 2009–12, orange = 2015, green = 2018) plus a pooled total (red). The soil maps differ in thematic resolution (map-unit count) and spatial scale. Their Shannon diversity index rises from left to right, so the X-axis is sorted by increasing legend depth rather than by map name.

Table 2
Metrics for the soil legacy maps in the considered geographical areas in order of spatial scale coarser to finer resolution, (ASD = Average Size delineations; MLA = Minimum Legible Area; IMR = Index of maximum reduction; AACD = Average Absolute Coverage Difference).

Map	Nominal scale	Area [km^2]	Polygons [n]	ASD [ha]	MLA [ha]	IMR	Effective scale	Total AACD [%]	Mean Index
Europe – WRB JRC	1:1,000,000	4,351,933	–	–	4000	–	–	0.56	1.03
Europe – WRB-FULL	1:1,000,000	4,341,232	–	–	4000	–	–	0.20	0.83
Europe – WRB-LEV1	1:1,000,000	4,341,232	–	–	4000	–	–	0.60	0.89
Soil Map of Italy	1:1,000,000	301,449	1886	15,984	4000	2.00	1:1,000,000	0.66	1.03
Soil Map of Crete	1:1,000,000	8347	31	26,926	4000	2.59	1:1,295,000	8.23	1.54
Soil Map of Andalusia	1:400,000	87,606	3075	2849	640	2.11	1:422,000	0.84	0.79
Soil Map of Sicily	1:250,000	25,474	2281	1117	250	2.11	1:263,750	1.27	0.98
Reclassified Sicily Map	1:250,000	25,474	1419	1795	250	2.68	1:335,000	2.02	0.71
General Soil Map of Cyprus	1:200,000	9196	275	3344	160	4.57	1:457,000	1.79	0.77
Soil Map of Viotia prefecture	1:100,000	1384	374	370	40	3.04	1:152,000	3.18	0.71

maps, especially those with finer resolution (1:250,000 or better), smaller surface area, and fewer SMUs. For instance, Viotia (1:100,000) and Crete (1:1,000,000 but very small area) showed AACD > 3%, reflecting potential mismatches between sampling density and actual soil class distributions. Notably, some small-scale maps (e.g., WRB-LEV1) with low diversity outperform more complex sub-national maps. This underscores that map scale alone does not explain representativeness: the legend depth and thematic simplification play key roles. Simplified, well-structured legends seem to align better with LUCAS's land-cover-driven sampling grid. A clear pattern emerges when AACD values are plotted against the Shannon diversity index (H'): maps with higher H' tend to show better and more stable AACD performance. This suggests that pedodiversity improves LUCAS representativeness, likely because a diverse soil landscape increases the chance that well distributed sampling points hit the different soil types. However, there are some exceptions: Crete, for example, has a relatively low H' and a very high AACD, highlighting how map quality and legend clarity affect representativeness. Maps with a greater number of SMUs generally exhibited lower AACD values, suggesting that classification granularity contributes positively to representativeness by capturing finer soil variability. Similarly, maps covering larger surface areas tended to perform better, likely because extensive coverage increases the probability of aligning with the spatial distribution of LUCAS sampling points.

The combined use of AACD and Mean Index provides a diagnostic framework to distinguish between apparent and effective representativeness. While AACD captures the average agreement between sampled points and mapped soil proportions, it is insensitive to the direction and concentration of deviations. The identification of the most over- and under-sampled soil units demonstrates that systematic biases often affect specific soil groups rather than the overall distribution. These biases are consistent with known constraints of the LUCAS design, including its land-cover-driven stratification, limited sampling density, and historical exclusion of high-elevation areas, which disproportionately affect shallow, mountainous soils, or the organic-rich soils largely concentrated in northern latitudes, not taken into account in this study.

Low AACD values do not necessarily indicate an unbiased representation of soil types, as they may result from cancellation effects between strongly over- and under-represented soil units. This pattern is clearly illustrated by several comparisons in which low AACD coexists with substantial large positive and negative Δ values affecting specific soil groups (Table 3).

Conversely, the Mean Index is highly sensitive to the omission or misrepresentation of rare soil types. While this sensitivity is useful for detecting systematic deviations, it may obscure severe biases affecting dominant soil units, particularly when over- and under-sampling are unevenly distributed across soil groups.

In general, comparisons classified as “Low AACD – Low deviation” correspond to soil maps at continental or broad sub-national scales, characterized by a high number of map units (>25) and elevated pedodiversity, as indicated by Shannon index values exceeding 2.2. In contrast, cases falling within the “High AACD – High deviation” category are consistently associated with small spatial extents, very limited

Table 3
Joint diagnostic interpretation of AACD and Mean Index. For each comparison, AACD values were classified as low (≤ 1%), moderate (1–3%), or high (> 3%), and deviations in the Mean Index were assessed using |1 − Mean Index| and classified as low (< 0.25) or high (≥ 0.25). The table reports, for each map, the representativeness pattern derived from the joint classification and the two most over- and under-sampled soil units, expressed as the difference (Δ) between the proportion of LUCAS sampling points and the mapped areal proportion of the corresponding soil unit. Positive Δ values indicate over-sampling, while negative values indicate under-sampling.

Comparison	Rank	Oversampled 1°	Δ	Oversampled 2°	Δ	Undersampled 1°	Δ	Undersampled 2°	Δ
Europe – WRB JRC	Low AACD – Low deviation	Calcisols	3.18	Luvisols	1.71	Histosols	−1.93	Leptosols	−1.59
Europe – WRB-FULL	Low AACD – Low deviation	Calcaric Cambisol	4.04	Calcaric Fluvisol	0.61	Haplic Podzol	−2.12	Dystric Histosol	−1.50
Europe – WRB-LEV1	Low AACD – Low deviation	Cambisols	4.84	Luvisols	1.78	Podzols	−2.90	Histosols	−1.68
Soil Map of Italy	Low AACD – Low deviation	*1	1.79	*2	1.78	*3	−3.14	*4	−1.31
Soil Map of Andalusia	Low AACD – Low deviation	Regosoles Calcáreos y Cambisoles cálcicos con litosoles, Fluviosoles calcáreos y Rendsinas	4.37	Cambisoles cálcicos con Regosoles calcáreos, Fluvisoles calcáreos y Luvisoles Cálcicos	4.23	Regosoles éutricos, Litosoles y cambisoles eútricos con Rankers, sobre materiales metamórficos	−4.91	Litosoles, Luvisoles crómicos y Rendsinas con Cambisoles cálcicos	−2.88
Sicily WRBLV1	Low AACD – High deviation	Cambisols	1.74	Vertisols	0.72	Regosols	−1.95	Leptosols	−0.75
General Soil Map of Cyprus	Moderate AACD – Low deviation	Ochric Rendzinas	6.74	Vertic Cambisols	3.56	Xeric Vertisols	−7.24	Pellic Vertisols	−2.77
Soil Map of Sicily	Moderate AACD – Low deviation	Eutric Regosols – Eutric e/o Vertic Cambisols – Eutric Fluvisols e/o Chromic e/o Pellic Vertisols	4.35	Eutric Cambisols – Vertic Cambisols – Chromic e/o Pellic Vertisols	3.55	Eutric Cambisols – Orthic Luvisols – Eutric Regosols e/o Lithosols	−4.09	Calcaric Regosols – Lithosols – Eutric e/o Vertic Cambisols	−2.58
Sicily Soil Map of Italy	Moderate AACD – Low deviation	*5	6.30	*6	4.13	*7	−4.55	*8	−3.18
Reclassified Soil Map of Sicily	Moderate AACD – High deviation	Leptosols	5.07	Fluvisols	1.96	Cambisols	−5.32	Calcisols	−2.58
Soil Map of Viotia prefecture	High AACD – High deviation	Fluvisols	5.05	Vertisols	3.04	Leptosols	−6.06	Calcisols	−2.36
Soil Map of Crete	High AACD – High deviation	Calcaric Regosols	12.50	Calcaric Regosols, stony	8.20	Calcaric Lithosols, stony	−12.05	Calcaric Lithosols	−8.99

*1 = Chromic, Haplic, Gleyic, and Skeletic Luvisol; Calcic Skeletic Luvisol; Haplic Luvisol (Dystric); Calcaric Cambisol (Bathicalcic); Eutric Vertic Cambisol; *2 = Hypercalcaric Regosol (Humic); Calcaric Episkeletic Regosol; Episkeletic Calcaric and Calcaric Fluvic Cambisol; Haplic Luvisol (Cutanic); Calcaric Regosol (Escalic); *3 = Albic, Umbric, and Entic Podzol; Haplic Podzol (Skeletic); Dystric Cambisol; Umbric and Dystric Hyperskeletic Leptosol; Haplic Phaeozem; Fibric Histosol; *4 = Dystric Leptic and Eutric Leptic Cambisol; Eutric and Lithic Leptosol; Eutric Regosol; *5 = Calcic, Sodic, and Haplic Vertisol; Fluvic and Calcaric Cambisol; Calcic Luvisol; Gypsiric Regosol; *6 = Leptic and Luvic Phaeozem; Leptic and Chromic Luvisol; Haplic Calcisol; Calcic Chernozem; Calcaric Regosol; Calcaric Cambisol; Calcic Kastanozem; Calcaric Leptosol; Calcaric Arenosol; *7 = Chromic, Calcic, and Haplic Luvisol; Haplic, Calcic, Chromic, and Hyposodic Vertisol; Haplic Calcisol; Calcaric and Eutric Cambisol; Calcaric Regosol; Calcaric Phaeozem; *8 = Leptic Umbrisol; Dystric Leptic Cambisol; Eutric Regosol; Mollic Leptosol (Vitric); Mollic Vitric and Silandic Andosol.

thematic complexity (7 map units), and low pedodiversity (Shannon index < 1.8). Maps classified as “Moderate AACD – Low deviation” typically represent sub-national extents and display an intermediate level of thematic detail, with a moderate number of map units (16–36) and relatively high pedodiversity (Shannon index > 2.2). By contrast, both the “Low AACD – High deviation” category, characterized by the lowest observed pedodiversity (Shannon index < 1.0), a limited number of map units (n = 10), and sub-national extent, and the “Moderate AACD – High deviation” category, represented by the same region (Sicily) but using a more pedodiverse map (Shannon index = 1.39), contain only a single comparison each. This highlights the sensitivity of the joint AACD–Mean Index classification to changes in thematic resolution within the same geographic context.

Over-represented soil units are predominantly widespread and geomorphologically stable soils, such as Cambisols, Luvisols, Regosols, and Vertisols, which typically occur in extensive lowland or gently undulating landscapes that are well captured by a land-cover–based sampling framework. Conversely, under-represented soil types are frequently associated with spatially restricted or environmentally marginal conditions, including Histosols, Podzols, Leptosols, and Lithosols.

4.2. *Policy impact*

Due to the “Mission soil” (Panagos et al., 2024), studies about legacy soil data and their potential use to support sampling design are increasing in number and scope (Schillaci et al., 2019; Stumpf et al., 2016; Xia et al., 2024). The new Soil Monitoring Law (European Union, 2025), states that it is “appropriate to lay down criteria for sampling points that are representative of the soil units reflecting a certain degree of homogeneity of soil condition under different soil types, climatic conditions and land use", and indicates that "LUCAS is to be enhanced and upgraded to fully align it with the specific quality requirements to be met for the purposes of this Directive".

In response to that, the proposed approach has proven effective in capturing soil variability across diverse landscapes, offering a standardized and harmonized dataset that can significantly enhance soil monitoring efforts. Extending this methodology to the NENA region could present several strategic advantages including soil conservation aspects, harmonization of methodologies for sampling across large region, and economic feasibility (de Bruyn and Prager, 2024; Louis et al., 2014). The standardized data collection facilitated by LUCAS can significantly contribute to soil conservation efforts. By identifying areas at risk of degradation, targeted interventions can be designed to prevent soil erosion and loss of fertility (Ferreira et al., 2022). Proactive soil conservation management at all stakeholder levels is crucial in regions like NENA, where soil degradation poses a significant threat to sustainable land use.

Implementing a consistent methodology across the NENA region would allow for better comparison of soil data across countries, facilitating regional comparisons and policy development (UNEP(DEPI)/MED IG.22/28 decision). This can inform regional policies (Castellini et al., 2025; Puertas and Marti, 2023), encourage collaboration, and foster the development of cross-border strategies for soil management and conservation. Additionally, it aligns with international efforts to combat desertification as highlighted in the ECA special report 33 (Prigent et al., 2018). The use of a harmonized methodology like LUCAS can reduce costs associated with data collection and analysis due to the standardized protocols and centralized laboratory processing, whenever cross-border agreements make this feasible. This cost-effectiveness is crucial for regions with limited financial resources dedicated to soil monitoring (Tepes et al., 2021). Moreover, the initial investment in establishing such a framework can lead to long-term savings by minimizing the need for repetitive surveys and facilitating more precise agricultural planning. This economic aspect includes added value generation for Agricultural Management: Accurate soil data are essential for optimizing agricultural practices (Barão et al., 2024). By adopting the LUCAS methodology, the NENA region can develop a comprehensive understanding of soil properties, which is fundamental for improving crop yields, managing soil nutrients efficiently, and selecting suitable crops for specific soil types. This data-driven approach can lead to enhanced agricultural productivity and food security in the region. Another uncharted aspect that this methodology uncovers is the leveraging of local legacy soil maps and knowledge (e.g. Sicily, Crete and Cyprus). While extending the LUCAS methodology, it is vital to adapt the sampling design to the pedological diversity, landscapes and climate conditions of the Mediterranean and nearby NENA countries at least with a robust technique (e.g. stratified random sampling, random forest, neural networks) (Shao et al., 2022; Wadoux et al., 2019; Zhang and Hartemink, 2017). Incorporating local expertise and knowledge will enhance the relevance and applicability of the data collected, ensuring better monitoring through data analytics, the methodology is tailored to address regional challenges effectively, extending the legacy of LUCAS soil sampling approach from the EU to the NENA region offers a practical, economically feasible, and scientifically robust approach to improving agricultural management and soil conservation (Hendriks et al., 2019). By leveraging the strengths of the LUCAS framework, NENA countries can enhance their soil monitoring capabilities, contributing to sustainable agricultural practices and environmental protection.

4.3. Limitations

While the legacy maps provide knowledge of soil characteristics, some limitations must be acknowledged. This analysis is based on data points representing specific locations within broader delineated areas where variations within each area might be underrepresented. Given that SLD maps were produced at different times and under varying conditions, there might be gaps in the temporal record. Moreover, the context in which SLD was collected might differ significantly from current land-use conditions. Changes in land management and anthropogenic influences over time can limit the applicability of historical findings to present-day scenarios. Nevertheless, when assessing soil classification accuracy, considering taxonomic distance is essential to refine accuracy metrics to better reflect map utility (Rossiter et al., 2017). However, our approach doesn't consider it, simplifying the complexity of the system.

Advanced digital soil mapping products, such as SoilGrids 2.0, provide predictive soil maps with quantified spatial uncertainty, complementing initiatives like LUCAS (Poggio et al., 2021) so choosing an appropriate procedure to evaluate whether the LUCAS topsoil survey adequately captures soil type variability presented significant challenges. Our primary task involved comparing SMUs, in terms of surface areas, with their corresponding LUCAS sampling points. SMUs are geometric polygons delineated in specific locations, historically drawn by experts based on their understanding of soil-forming factors in the study area. As SMUs commonly represent combinations of STUs, our analysis cannot describe the within-unit mixture of soils. The metrics we propose must therefore be interpreted as measures of representativeness with respect to the polygon-level legend of the LSMs, not as an evaluation of how well the sampling network captures all STUs present inside each polygon. SMUs usually represent combinations of STUs, and the dominant class may not even occupy half of the polygon area. Our approach therefore cannot quantify representativeness at the level of individual STUs; it evaluates alignment only with the dominant mapped classes in the legacy maps. Evaluating the accuracy of soil classification is complex due to several interconnected factors. Multiple classification systems exist, each with different rules and hierarchies, and some soil classes can belong to multiple soil types. The process is inherently influenced by subjective interpretations, and all cartography represents a simplified version of reality, dependent on the observational scale.

Soil is a complex three-dimensional physical system whose properties are often hidden beneath the surface. Its characteristics vary spatially, both horizontally and vertically, and change over time. Like oceans, soils lack distinct physical boundaries, necessitating the establishment of artificial limits for studying their characteristics. A partial solution to this boundary problem is the concept of the polypedon, a three-dimensional soil body with lateral dimensions sufficient to represent variations in horizon shapes and soil composition (Johnson, 1963). The horizontal extent typically ranges from 1 to 10 square meters, depending on soil variability, while its volume depends on soil depth (Soil Survey Staff, 1999).

LUCAS sampling points represent specific spatial locations. According to the LUCAS field protocol, the spatial support of each sample can be considered equivalent to a cylinder with a base diameter of 4 m and a height (soil depth) of either 0.2 or 0.3 m, depending on the sampling campaign. Our research confronts the difficulty of comparing fundamentally different concepts. On one side, we have soil type map units, which are quasi-subjective classifications based on pedon observations. On the other hand, we have point sampling, which represents a three-dimensional support structure of lesser depth that doesn't reach the lower natural soil boundary, assimilable to a point in the space. The LUCAS topsoil surveys cannot describe pedon, but provide soil properties at the sampled points. Our comparison relies on the spatial attribution of soil classifications based on LSMs. These maps were constructed using different classification systems, created at varying scales, and developed during different time periods, adding further complexity to our analysis.

Some of the legacy maps used are relatively old and may not fully reflect recent land-use change or updates in national classification systems. We chose them because they remain the authoritative polygon products openly available at continental and sub-national scales, but their varying age and design inevitably limit the temporal representativeness of our assessment. Future work could complement this approach also using recent predictive grids (e.g. SoilGrids 2.0) and updated national surveys, where available.

5. Conclusions

Although the LUCAS spatial distribution was originally designed to monitor land use changes without considering soil types, the array of information layers used for the design captures key factors of soil formation from topography and climate to, more broadly, biota. As a result, and thanks to the large number of evenly distributed sampling points across the EU, the LUCAS sampling points effectively represented the different soil type in all study areas, independently of the scale of observation. This study provides a new way to assess the spatial representativeness of a continental soil sampling network (LUCAS topsoil) in relation to the distribution of mapped soil classes as encoded in legacy soil maps with varying scale, classification systems, and pedodiversity. Using the proposed Average Absolute Coverage Difference (AACD) and "mean index", we demonstrated that the effectiveness of LUCAS in capturing mapped soil type variability is highly dependent on map characteristics. Overall, the LUCAS topsoil survey shows a consistent ability to represent soil type distributions not only at the continental level but also at the sub-national scale, particularly in cases where legacy maps exhibit higher pedodiversity (as indicated by Shannon index), broader spatial extent, and a greater number of soil map units. While representativeness tends to decrease with increasing map scale (i.e., greater spatial detail), maps with high soil diversity and well-structured classifications—such as the Soil Map of Sicily (Fierotti et al., 1988) maintain low AACD values, demonstrating that spatial resolution alone does not limit sampling effectiveness.

These findings support the strategic use of LSM, and more generally of available soil data, not only for retrospective assessment but also for improving the design of future soil surveys. This could be important to achieve a stronger alignment of the LUCAS approach with what is prescribed by the new Soil Monitoring Law of Europe. Furthermore, as the LUCAS framework is considered for extension into the Near East and North Africa (NENA) region, integrating existing spatial soil information will be essential to ensure balanced representation, particularly in ecologically heterogeneous and data-scarce contexts. Based on the results achieved through the analysis of the LUCAS soil sampling design methodology compared with the legacy soil type maps, it is recommended the extension of this approach to the NENA region.

CRediT authorship contribution statement

G. Belvisi: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **C. Schillaci:** Writing – review & editing, Writing – original draft, Formal analysis, Data curation, Conceptualization. **L. Gristina:** Writing – review & editing, Investigation. **A. Delgado:** Writing – review & editing, Investigation, Data curation. **D. Triantakonstantis:** Writing – review & editing, Investigation, Data curation. **N. Lolos:** Writing – review & editing, Investigation, Data curation. **M. Batsalia:** Writing – review & editing, Investigation, Data curation. **A. Jones:** Writing – review & editing. **L. Ribeiro Roder:** Writing – review & editing. **C. Zucca:** Writing – review & editing, Investigation, Funding acquisition. **R. Scalenghe:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

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

Acknowledgments

The authors express their gratitude to the PRIMA Program, supported under Horizon 2020, for funding the SOILS4MED project (Grant Agreement: 2212). They also extend their sincere appreciation to all project partners for their valuable contributions. Special thanks go to David G. Rossiter for his helpful insights throughout the process of writing this research paper. His contributions were much appreciated.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.geoderma.2026.117735.

Data availability

The data presented in this study refer to processed inputs derived from publicly available sources. A summary of the analytical outputs, including key indicators (Δ points–surface, AACD, Mean Index) for each map comparison and LUCAS campaign (2009–12, 2015, 2018, and total), along with the conversion table for the reclassification of the Soil Map of Sicily, is provided in the Supplementary Information, with a glossary and documentation of the formulas used. The original soil maps and LUCAS topsoil survey data are available from the respective public repositories as cited in the manuscript.

References

Arrouays, D., Leenaars, J.G.B., Richer-de-Forges, A.C., Adhikari, K., Ballabio, C., Greve, M., Grundy, M., Guerrero, E., Hempel, J., Hengl, T., Heuvelink, G., Batjes, N., Carvalho, E., Hartemink, A., Hewitt, A., Hong, S.-Y., Krasilnikov, P., Lagacherie, P., Lelyk, G., Rodriguez, D., 2017. Soil legacy data rescue via GlobalSoilMap and other international and national initiatives. GeoResJ 14, 1–19. https://doi.org/10.1016/j.grj.2017.06.001.

Barão, L., Ramos, T., Basch, G., Ribeiro, I., Teixeira, R., 2024. Improved agricultural management systems for soil conservation in the Mediterranean. In: Ferreira, C.S.S., Destouni, G., Kalantari, Z. (Eds.), Environmental Sustainability in the Mediterranean Region: Challenges and Solutions. Springer International Publishing, Cham, pp. 171–191. https://doi.org/10.1007/978-3-031-64503-7_8.

Cambule, A.H., Rossiter, D.G., Stoorvogel, J.J., Smaling, E.M.A., 2015. Rescue and renewal of legacy soil resource inventories: a case study of the Limpopo National Park, Mozambique. Catena 125, 169–182.

Carrubba, A., Scalenghe, R., 2012. The scent of Mare Nostrum: medicinal and aromatic plants in Mediterranean soils. J. Sci. Food Agric. 92 (6), 1150–1170.

Castellini, M., Castelli, C., Gusperti, C., Lupi, V., Vergalli, S., 2025. Balancing climate policies and economic development in the Mediterranean countries. Energy Econ. 145, 108452. https://doi.org/10.1016/j.eneco.2025.108452.

Cornu, S., Keesstra, S., Bispo, A., Fantappiè, M., Van Egmond, F., Smreczak, B., Wawer, R., Pavlů, L., Sobocká, J., Bakacsi, Z., Farkas-Iványi, K., Molnár, S., Møller, A.B., Madenoglu, S., Feiziene, D., Oorts, K., Schneider, F., Da Conceição Gonçalves, M., Mano, R., Chenu, C., 2023. National soil data in EU countries, where do we stand? Eur. J. Soil Sci. 74 (4), e13398. https://doi.org/10.1111/ejss.13398.

de Bruyn, L., Prager, K., 2024. How can soil data and information become actionable knowledge to advance sustainable land management? Soil Use Manag 40, e13115. https://doi.org/10.1111/sum.13115.

Dent, D.L., Ahmed, F.B., 1995. Resurrection of soil surveys - A case-study of the acid sulfate soils of the Gambia. 1. Data validation, taxonomic and mapping units. Soil Use Manag. 11, 69–76.

Drosos, M., Orlando, M., Cozzolino, V., Scopa, A., Piccolo, A., 2023. Deriving the Shannon Index from the soil molecular Humeome serves as a descriptor of soil organic matter stability under different cropping systems. Chem. Biol. Technol. Agric. 10 (1), 105. https://doi.org/10.1186/s40538-023-00473-w.

Ellis, E.C., Gauthier, N., Klein Goldewijk, K., Bliege Bird, R., Boivin, N., Díaz, S., Fuller, D. Q., Gill, J.L., Kaplan, J.O., Kingston, N., Locke, H., McMichael, C.N.H., Ranco, D., Rick, T.C., Shaw, M.R., Stephens, L., Svenning, J.-C., Watson, J.E.M., 2021. People have shaped most of terrestrial nature for at least 12,000 years. Proc. Natl. Acad. Sci. 118 (17), e2023483118. https://doi.org/10.1073/pnas.2023483118.

Fatichi, S., Or, D., Walko, R., Vereecken, H., Young, M.H., Ghezzehei, T.A., Hengl, T., Kollet, S., Agam, N., Avissar, R., 2020. Soil structure is an important omission in Earth System Models. Nat. Commun. 11 (1), 522. https://doi.org/10.1038/s41467-020-14411-z.

European Union, 2025. Directive (EU) 2025/2360 of the European Parliament and of the Council of 12 November 2025 on soil monitoring and resilience (Soil Monitoring Law). Off. J. Eur. Union L, 2025/2360, 26.11.2025, 1-56.

FAO/ISRIC/ISSS, 1998. World Reference Base for Soil Resources. World Soil Resources Report, 84. FAO, Rome, p. 88pp.

Ferreira, C.S.S., Seifollahi-Aghmiuni, S., Destouni, G., Ghajarnia, N., Kalantari, Z., 2022. Soil degradation in the European Mediterranean region: processes, status and consequences. Sci. Total Environ. 805, 150106. https://doi.org/10.1016/j.scitotenv.2021.150106.

Fierotti, D., Dazzi, C., Raimondi, S., 1988. Soil Map of Sicily (scale 1:250,000). Regione Siciliana. Assessorato Agricoltura e Foreste, Palermo, Italy.

Forbes, T., Rossiter, D., van Wambeke, A., 1987. Guidelines for evaluating the adequacy of soil resource inventories. In: SMSS Technical Monograph: Soil management support services, 4. Cornell University.

Gardi, C., Florczyk, A.J., Scalenghe, R., 2021. Outlook from the soil perspective of urban expansion and food security. Heliyon 7 (1), e05860. https://doi.org/10.1016/j.heliyon.2020.e05860.

He, L., Sun, X., Li, S., Zhou, W., Chen, Z., Bai, X., 2023. The vertical distribution and control factor of microbial biomass and bacterial community at macroecological scales. Sci. Total Environ. 869, 161754. https://doi.org/10.1016/j.scitotenv.2023.161754.

Hendriks, C.M.J., Stoorvogel, J.J., Lutz, F., Claessens, L., 2019. When can legacy soil data be used, and when should new data be collected instead? Geoderma 348, 181–188. https://doi.org/10.1016/j.geoderma.2019.04.026.

Hiederer, R., Durrant, T., 2010. Evaluation of BioSoil Demonstration Project - Preliminary Data Analysis. EUR – Scientific and Technical Research series EUR 24258 EN, 126 pp. Office for Official Publications of the European Communities, Luxembourg.

IUSS Working Group WRB, 2006. World reference base for soil resources, 2nd edition. World Soil Resources Reports No. 103. FAO, Rome.

IUSS Working Group WRB, 2022. World Reference Base for Soil Resources. International soil classification system for naming soils and creating legends for soil maps, 4th ed. International Union of Soil Sciences (IUSS), Vienna, Austria.

Jenny, H., 1941. Factors of Soil Formation: a System of Quantitative Pedology. Dover Publications, New York, p. 281 (Dover ed., 1994; unabridged republication of 1941 original).

Johnson, W.M., 1963. The pedon and the polypedon. Soil Sci. Soc. Am. J. 27 (2), 212–215. https://doi.org/10.2136/sssaj1963.03615995002700020034x.

Jones, R.J., Houšková, B., Bullock, P., 2005. In: Montanarella, L. (Ed.), Soil resources of Europe. Ispra, Italy.

Louis, B.P., Saby, N.P.A., Orton, T.G., Lacarce, E., Boulonne, L., Jolivet, C., Ratié, C., Arrouays, D., 2014. Statistical sampling design impact on predictive quality of harmonization functions between soil monitoring networks. Geoderma 213, 133–143. https://doi.org/10.1016/j.geoderma.2013.07.018.

McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117 (1–2), 3–52. https://doi.org/10.1016/S0016-7061(03)00223-4.

McKenzie, N.J., Webster, R., Ryan, P.J., 2008. Sampling using statistical methods. In: McKenzie, N.J., Grundy, M.J., Webster, R., Ringrose-Voase, A.J. (Eds.), Guidelines for Surveying Soil and Land Resources, Second Edition. CSIRO Publishing, Melbourne, Australia, pp. 317–326.

Misopolinos, N., Silleos, N., Bilas, G., Karapetsas, N., Dionisiou, N., Misopolinos, L., Pazarloglou, M., Srati, S., Cherif, I., Ovakoglou, G., et al. (2015). Soil Map of Greece 1:30.000, Development of a Universal SYSTEM of Handling geoinformation Data and Mapping Agricultural Areas of Greece, Report in Greek, Payment and Control Agency for Guidance and Guarantee Community Aid (O.P.E.K.E.P.E), Aristotle University of Thessaloniki (AUTH), Thessaloniki.

Morvan, X., Saby, N.P.A., Arrouays, D., Le Bas, C., Jones, R.J.A., Verheijen, F.G.A., Bellamy, P.H., Stephens, M., Kibblewhite, M.G., 2008. Soil monitoring in Europe: a review of existing systems and requirements for harmonisation. Sci. Total Environ. 391 (1), 1–12. https://doi.org/10.1016/j.scitotenv.2007.10.046.

Murphy, B. (2022). An Application of Information Theory to a Soil Map: Implications for Soil Information Systems (SSRN Scholarly Paper 4077504). Social Science Research Network. https://doi.org/10.2139/ssrn.4077504 (Accessed 11 February 2026).

Pacini, L., Yunta, F., Jones, A., Montanarella, L., Barrè, P., Saia, S., Chen, S., Schillaci, C., 2023. Fine earth soil bulk density at 0.2 m depth from Land Use and Coverage Area Frame Survey (LUCAS) soil 2018. Eur. J. Soil Sci. 74. https://doi.org/10.1111/ejss.13391.

Panagos, P., Borrelli, P., Jones, A., Robinson, D.A., 2024. A 1 billion euro mission: a Soil Deal for Europe. Eur. J. Soil Sci. 75. https://doi.org/10.1111/ejss.13466.

Panagos, P., Van Liedekerke, M., Borrelli, P., Köninger, J., Ballabio, C., Orgiazzi, A., Lugato, E., Liakos, L., Hervas, J., Jones, A., Montanarella, L., 2022. European soil data centre 2.0: soil data and knowledge in support of the EU policies. Eur. J. Soil Sci. 73 (6), e13315. https://doi.org/10.1111/ejss.13315.

Poggio, L., De Sousa, L.M., Batjes, N.H., Ribeiro, E., Rossiter, D., 2021. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. Soil 7 (1), 217–240. https://doi.org/10.5194/soil-7-217-2021.

Prigent, O., Wynn Owen, P., Homrich Hickmann, M., et al., 2018. Combating desertification in the EU: a growing threat in need of more action. In: Special Report No 33/2018. European Court of Auditors. Luxembourg, p. 65. https://doi.org/10.2865/425778.

Puertas, R., Marti, L., 2023. Regional analysis of the sustainable development of two Mediterranean countries: Spain and Italy. Sustain. Dev. 31, 797–811. https://doi.org/10.1002/sd.2420.

Rasaei, Z., Rossiter, D.G., Farshad, A., 2020. Rescue and renewal of legacy soil resource inventories in Iran as an input to digital soil mapping. Geoderma Reg 00262. https://doi.org/10.1016/j.geodrs.2020.e00262.

Rossiter, D.G., 2008. Digital soil mapping as a component of data renewal for areas with sparse soil data infrastructures. In: Hartemink, A.E., McBratney, A., Mendonça-Santos, M.d.L. (Eds.), Digital Soil Mapping with Limited Data. Springer, Dordrecht, pp. 69–80.doi. https://doi.org/10.1007/978-1-4020-8592-5_6.

Rossiter, D.G., Poggio, L., 2025. Representing soil landscapes from digital soil mapping products: let the map speak for itself. Egusphere. https://doi.org/10.5194/egusphere-2025-1896.

Rossiter, D.G., Poggio, L., Beaudette, D., Libohova, Z., 2022. How well does digital soil mapping represent soil geography? an investigation from the USA. Soil 8 (1), 559–586. https://doi.org/10.5194/soil-8-559-2022.

Rossiter, D.G., Zeng, R., Zhang, G.L., 2017. Accounting for taxonomic distance in accuracy assessment of soil class predictions. Geoderma 292, 118–127. https://doi.org/10.1016/j.geoderma.2017.01.021.

Sarmento, E.C., Giasson, E., Weber, E.J., Flores, C.A., Rossiter, D.G., Hasenack, H., 2014. Characterization of legacy soil maps: use of indicators in maps at different cales in Rio Grande do Sul (in Portuguese, English abstract). Rev. Brasi. Ciência Do Solo 38, 1672–1680. https://doi.org/10.1590/S0100-06832014000600002.

Schillaci, C., Acutis, M., Vesely, F., Saia, S., 2019. A simple pipeline for the assessment of legacy soil datasets: an example and test with soil organic carbon from a highly variable area. Catena 175, 110–122. https://doi.org/10.1016/j.catena.2018.12.015.

Schillaci, C., Scarpa, S., Yunta, F., Lipani, A., Visconti, F., Szatmári, G., Balog, K., Koganti, T., Greve, M., Bondi, G., Kargas, G., Londra, P., Kaya, F., Lo Papa, G., Panagos, P., Montanarella, L., Jones, A., 2025. Empirical estimation of saturated soil-paste electrical conductivity in the EU using pedotransfer functions and Quantile Regression Forests: a mapping approach based on LUCAS topsoil data. Geoderma 454, 117199. https://doi.org/10.1016/j.geoderma.2025.117199.

Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27 (3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

Shao, S., Su, B., Zhang, Y., Gao, C., Zhang, M., Zhang, H., Yang, L., 2022. Sample design optimization for soil mapping using improved artificial neural networks and simulated annealing. Geoderma 413, 115749.

Soil Science Division Staff, 2017. Soil Survey Manual. Ditzler, C., Scheffe, K., Monger, H. C. (Eds.), USDA Handbook 18. Government Printing Office, Washington, DC. Available at: https://www.nrcs.usda.gov/resources/guides-and-instructions/soil-survey-manual (Accessed 11 February 2026).

Soil Survey Staff 1999, Soil Taxonomy. Soil Use and Management, 17(1), 57–60. https://doi.org/10.1111/j.1475-2743.2001.tb00008.x.

Soil Survey Staff. 2022. Keys to Soil Taxonomy, 13th edition. USDA Natural Resources Conservation Service.

Soil Survey Staff, 1975. Soil Taxonomy: A Basic System of Soil Classification for Making and Interpreting Soil Surveys. U.S. Dep. Agric. Handb. 436. U.S. Government Printing Office, Washington, DC, pp. 754–pp.

Stumpf, F., Schmidt, K., Behrens, T., Schönbrodt-Stitt, S., Buzzo, G., Dumperth, C., Wadoux, A., Xiang, W., Scholten, T., 2016. Incorporating limited field operability and legacy soil samples in a hypercube sampling design for digital soil mapping. J. Plant Nutr. Soil Sci. 179, 499–509. https://doi.org/10.1002/jpln.201500313.

Su, H., Yang, Z., Liu, Z., Zhang, R., Wu, S., Li, Y., Yao, H., 2024. Applicability of soil health assessment dominated by biological indicators in facility agriculture. Sci. Total Environ. 957, 177346. https://doi.org/10.1016/j.scitotenv.2024.177346.

Tóth, G., Jones, A., Montanarella, L., 2013. The LUCAS topsoil database and derived information on the regional variability of cropland topsoil properties in the European Union. Environ. Monit. Assess. 185 (9), 7409–7425. https://doi.org/10.1007/s10661-013-3109-3.

Tepes, A., Galarraga, I., Markandya, A., Sánchez, M.J.S., 2021. Costs and benefits of soil protection and sustainable land management practices in selected European countries: towards multidisciplinary insights. Sci. Total Environ. 756, 143925. https://doi.org/10.1016/j.scitotenv.2020.143925.

USDA (1975) Soil Taxonomy. A Basic System of Soil Classification for Making and Interpreting Soil Surveys. Washington DC, 745.

Van Leeuwen, J.P., Saby, N.P.A., Jones, A., Louwagie, G., Micheli, E., Rutgers, M., Schulte, R.P.O., Spiegel, H., Toth, G., Creamer, R.E., 2017. Gap assessment in current soil monitoring networks across Europe for measuring soil functions. Environ. Res. Lett. 12 (12), 124007. https://doi.org/10.1088/1748-9326/aa9c5c.

Van Wambeke, A., Forbes, T.R., 1986. Guidelines for Using "Soil Taxonomy" in the Names of Soil Map Units. SMSS Technical Monograph No. 10. New York State College of Agriculture and Life Sciences (Cornell University), Ithaca, NY, 75 pp. Available at: https://edepot.wur.nl/479300 (Accessed 11 February 2026).

Vašát, R., Vacek, O., Borůvka, L., 2023. Studying the spatial structure of pedodiversity (Shannon's entropy) as related to the land area – an example from Czechia. Geoderma Reg. 32, e00607. https://doi.org/10.1016/j.geodrs.2023.e00607.

Wadoux, A.-M.-C., Brus, D.J., Heuvelink, G.B.M., 2019. Sampling design optimization for soil mapping with random forest. Geoderma 355, 113913. https://doi.org/10.1016/j.geoderma.2019.113913.

Xia, Y., Sanderman, J., Watts, J.D., Machmuller, M.B., Ewing, S., Rivard, C., 2024. Leveraging legacy data with targeted field sampling for low-cost mapping of soil organic carbon stocks on extensive rangeland properties. Geoderma 448, 116952. https://doi.org/10.1016/j.geoderma.2024.116952.

Yaalon, D.H., 1997. Soils in the Mediterranean region: what makes them different? Catena 28 (3–4), 157–169.

Zhang, Y., Hartemink, A.E., 2017. Sampling designs for soil organic carbon stock assessment of soil profiles. Geoderma 307, 220–230. https://doi.org/10.1016/j.geoderma.2017.08.013.

Zdruli, P., Kapur, S., Çelik, I., 2010. Soils of the Mediterranean region, their characteristics, management and sustainable use. In: Kapur, S., Eswaran, H., Blum, W.E.H. (Eds.), Sustainable Land Management: Learning from the Past for the Future. Springer, Dordrecht, pp. 125–142. https://doi.org/10.1007/978-3-642-14782-1_4.