AI4SoilHealth

Data sampling methodology co-designed with pilots

D6.3

31.3.2024

Lead Authors: Riikka Keskinen (Luke) and Kari Ylivainio (Luke)

Participating Authors: Lur Epelde (Neiker), Thomas Gumbricht (Stockholm University), Cordelia Hughes (Soil Association), Kostas Karyotis (AUTh), Aidan Keith (CEH), Frida Keuper (INRAE), Peter Lehmann (ETH Zürich), Sonia Meller (DS), Robert Minarik (OGH), Trine Nørgaard (Aarhus University), David Robinson (CEH), Marija Romić (AGR), Katy Jo Stanton (Soil Association), Joe Taylor (CEH), Mendy van der Vliet (Planet), Nikolaos Tziolas (AUTh)

Reviewed by: Emmanuel Arthur (Aarhus University) and Grant Campbell (University of Aberdeen)

Action Number: 101086179

Action Acronym: AI4SoilHealth
Action title: Accelerating collection and use of soil health information using AI technology to support the Soil Deal for Europe and the EU Soil Observatory

HISTORY OF CHANGES		
Version	**Publication date**	**Changes**
1.0	31 March 2024	• Initial version
		•

Table of contents

Production specifications

Project ID: https://cordis.europa.eu/project/id/101086179

FOR INTERNAL USE ONLY

1. Executive summary:

This deliverable provides an overview of the 12 pilot sites selected as part of the AI4SoilHealth project. These pilot sites have been chosen because of the varied soil types, land use forms and climate zones across Europe and will vary in scale from experimental field parcel size up to country level. Each of these sites are dedicated to evaluating soil health indicators, specifically important for the pilots. This will also help towards providing evidence for a future evaluation of soil health as has been specified within the draft Soil Monitoring Law. Data obtained from these pilot sites will be used for validating existing soil health metrics. In addition to these basic soil indicators specified in the draft directive, novel soil indicators, such as spectroscopy, soil macrofauna and eDNA, will be assessed.

In 10 of the 12 pilot sites, physical soil sampling will take place, whereas in two pilot sites (in Germany and Italy), legacy data and remote sensing will be utilized. Methodologies for soil sampling, pre-treatment of soil samples, their storage and methods for analysing soil health indicators will be presented.

2. Introduction

To achieve healthy soils, prevalent for sustaining the basic needs of humans, the EU through the EU Soil Strategy 2030, has targeted that all EU soil ecosystems are in healthy conditions by 2050 (https://environment.ec.europa.eu/topics/soil-and-land/soil-health_en). To achieve this target, the EU has set following aspects for soil degradation: 1) Salinization, 2) Soil erosion, 3) loss of soil organic carbon, 4) Soil compaction, 5) Excess nutrient content in soil, 6) Soil contamination, 7) Reduction of soil capacity to retain water, and 8) Acidification. Some of these aspects covers all Member States, including absolute criteria for healthy soil conditions (1-4), whereas others are on a Member State level, including criteria ranges for healthy soil conditions (5-7) and aspects without criteria (8). The criterion for subsoil compaction is pitched for the EU level, whereas topsoil compaction has no criteria. This information can be summarised in **Table 1**.

Table 1: Criteria for defining healthy soils.

Criteria	Set Soil Descriptor	Area(s) Applied	Additional information
Absolute Criteria for Healthy Soils	Salinisation	All Member States (EU wide)	
	Soil Erosion		
	loss of soil organic carbon		
	Soil compaction		Subsoil Compaction (EU Level) Topsoil compaction (no criteria)
criteria range for healthy soil conditions	Excess nutrient content in soil	Member State Level	
	Soil contamination		
	Reduction of soil capacity to retain water		
Descriptors without criteria	Acidification		

Descriptors for healthy soils are commonly determined in the laboratory after collecting soil samples from or are determined in situ in the field. However, these are time consuming and commonly, only small areas can be evaluated. Furthermore, various methods for determining soil descriptors exist among countries and correlation between methods may be inconsistent. In the EU, the Land Use/Cover Area frame Survey (LUCAS) soil sampling campaign covers all Member States, having uniform soil sampling scheme and protocols for soil analysis, providing comparable results across the Member States.

LUCAS soil sampling scheme provides legacy data source background for the monitoring of soil health indicators in the EU. Correlation of remote sensing and LUCAS data will be evaluated for soil health monitoring at pan-European level, and this correlation is further validated comparing predicted values of the indicators to the field observations (ground truth data) in the 12 pilot regions included in the AI4SoilHealth project, covering different climatic conditions across the EU. Activities within each of these pilots will vary, depending on which soil health indicator(s), are being investigated as stated by the EU Soil Observatory (EUSO). In most of the pilot sites, testing of the different soil health indicators will be based on known field measurements. Some pilot sites (e.g. Italy and Germany) will be reliant on legacy data and remote sensing due to difficulties in conducting actual field measurements.

AI4SoilHealth is collecting a range of data and formulating this into the form of a "Soil Health Data Cube". This will consist of a time-series of images including spectral indices related to the bare soil, biophysical status of the vegetation and management practice (Tian et al., 2024), primary and secondary soil properties, land degradation indices, terrain parameters, and similar EO products at 30 m resolution and encompassing the period 2000–2023+ of available satellite spectral datasets across Europe, in order to survey soil properties and to further support developing a decision support system for soil management (WP5). Furthermore, these EO products spectral data will be compiled alongside existing on-ground data to look for relationships

between these proxies. The pilot sites will determine soil health indicators, either through conventional soil testing methods or by utilising novel approaches to determine the potential of new soil health indicators, indicated in previous WP3 deliverables. These new soil health indicators will be validated through measurements conducted at the pilot sites. This deliverable will present the methodology used for collecting legacy data and provide methodology for soil sampling in order to conduct the state-of-the-art, science-based methods required for effective soil health assessment. The pilot regions will cover different pedoclimatic regions in Europe and therefore all indicators indicated by the EUSO are not relevant throughout Europe and may be only appropriate in certain pilots. For instance, salinity is detected only in certain regions across Europe, and Croatia is the only pilot region evaluating this indicator.

3. Multi-actor engagement pilots

A total of 12 pilot regions for demonstrating soil health indicators across Europe were established in various climatic regions with varying soil properties. The aim of these pilot sites are to validate existing datasets-utilized by WP5 - o develop web app-based cyberinfrastructure termed "AI4SoilHealth". The following sub-sections provide a background of the pilot regions and methods used for evaluating soil health indicators in these areas.

3.1. Croatia

3.1.1. Description of the pilot: The pilot area is located in the Neretva River Valley, on the southeastern coast of the Adriatic Sea (43°00' N, 17°30' E). The Neretva River is the largest river of the Adriatic catchment area and, according to the annual water discharge, is one of the largest rivers of the Mediterranean Sea. The Neretva River, with its numerous tributaries in its lower course, forms an extensive low-lying delta, which influences the main river either directly at the surface or indirectly in the groundwater. The entire delta is under the strong influence of the Adriatic Sea, both through sea water intrusion through the main riverbed and groundwater. In the delta, the Neretva River overlies Quaternary deposits. Carbonate rocks, mainly limestone, form the edge of the valley, with its flanks, and smaller isolated hill hummocks located within the valley. These rocks are intensely fractured and deeply karstified. The Surface Quaternary sediments are characterized mainly of peat and clay (in the form of organic marshy deposits), which are underlain by clayey sands, sands of variable grain size, sandy clay, gravelly sands, sandy gravel, Holocene gravels (alluvial sediments), and Pleistocene conglomerates. Polder-like agricultural land is intensively used for the cultivation of citrus fruits and vegetables. Since the elevation of the polders within the ameliorated area is lower than in the surrounding area, continuous drainage of excess water is required through a dense network of drainage canals, gates, and pumping stations.

The soil and water monitoring network covers an area of 5815 ha (consider using equivalent in km2) and includes four different polders. Polders drain excess water independently into higher areas or to the sea, forming a system that requires observation of water levels in every polder so that proper management of the hydraulic structures that connect them is achieved. Optimization of the monitoring network in such an aspect is performed to include all water bodies that have different functions within the polder. The segmentation of the polder-type land by active land use with agricultural dikes, ditches, and roads strongly

influences the dynamics of saline/brackish/freshwater circulation in the floodplains, causing the complex interactions of ecological, hydrological, and water and soil quality factors.

3.1.2. Climate and weather: The area is semiarid with a Mediterranean climate, with dry, hot summers and humid, mild winters. The mean annual rainfall for the main weather station of the area (Ploče) is 1077 mm (1988–2020), with most of the rainfall occurring in the period from October to April. The minimum rainfall occurs in July (28 mm) and the maximum occurs in November (153 mm). The mean annual air temperature is 15.9 °C, with January being the coldest (7.0 °C) and July the warmest month (25.7 °C). The annual Penman–Montheith reference evapotranspiration is 1217 mm, with the highest value of 197 mm occurring in July.

3.1.3. Overall description of relevant Soil Health Indicators: Specific soil health indicators to assess the extent of salinization and its effects on agricultural productivity are monitored. Firstly, **electrical conductivity** (EC) will be measured to assess the level of salinity in the soil. Monitoring **pH** levels relates to soil alkalinity, which will affect availability of nutrients and microbial activity. **Sodium adsorption ratio (SAR)** and **exchangeable sodium percentage (ESP)** will indicate the risk of soil dispersion and structural degradation in the pilot site. This will affect water infiltration and plant root growth, as well as the level of sodicity, which affects soil structure and permeability. **Soil moisture content** is relevant in saline-affected areas, as excess accumulation of salt can hinder water uptake by plants and lead to water stress. **Soil texture** affects water movement and drainage, influencing salt distribution and leaching. Salinity can affect nutrient availability and uptake by plants, so monitoring **essential nutrients levels** is needed to assess soil fertility and potential nutrient deficiencies. Salinity can alter microbial community composition and activity in the soil, impacting nutrient cycling and organic matter decomposition, and therefore assessment of **microbial diversity** is needed, but currently not performed.

3.1.4. How this indicator is evaluated:

Legacy data on soil health

- Surface soil (0-25 cm) 246 samples, 500 m grid
- pH, Electric Conductivity (Ece), ionic composition of the saturated paste, Cation Exchange Capacity (CEC), particle size distribution (psd), Cd, Co, Cr, Cu, Mo, Ni, Pb, V, Zn, Mn, P, Na, Al, Ca, Fe, Mg, K (mg/kg)
- Subsurface soil: samples taken at depths 25-50 cm; 50-75 cm; 75-100 cm from 63 locations

Soil salinity monitoring (2010-2023)

- 7 soil profiles: (samples taken at depths (0-25 cm; 25-50 cm; 50-75 cm; 75-100 cm))
- Sampling frequency: twice a year
- **Continuous sensor measurements (2021-2023);** Frequency-domain (FDR) sensors (FDR)
- 2 locations
- Depths: 0-25 cm; 25-50 cm; 50-75 cm; 75-100 cm
- Volumetric water content, temperature, Electric Conductivity of bulk sample

3.2. Denmark

3.2.1. Description of the pilot: Denmark is located in the north temperate climate zone and has a temperate coastal/sea climate (Cfb - Köppen climate classification) with mean annual temperatures of 8.3 degrees and annual precipitation of 746 mm.

3.2.2. Danish pilot site 1: This pilot covers a subset of the national peat soil sampling in Denmark. The national peat sampling in Denmark covers 1000 points, however, the pilot covers 130 of these points selected randomly from the national sampling. The 130 points are evenly distributed across Denmark (see **Figure 1**). The 130 points are sampled from peat soils in Denmark. The dataset contains in-field analyses, bulk soil analyses and intact/100 cm3-ring analyses (see table 2). In each of the sampling points, there are up to four depths for the measures of C and N.

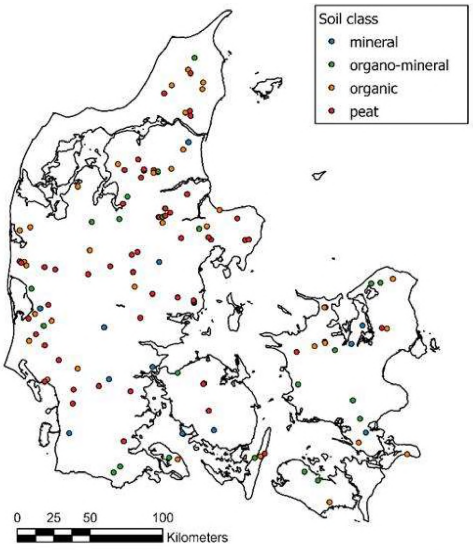


Figure 1. Map depicting sampling points in the Danish pilot site, differentiated by soil class.

3.2.3. Overall description of relevant Soil Health Indicators: As part of the national peat sampling programme, the indicators covered in this data set are included to assess the peat health status of Danish soils. Therefore, **Total Organic Carbon (TOC)** is the most critical indicator for this pilot, though several variables have been assessed.

3.2.4. How these indicators are evaluated: This dataset is unique as it covers a large database of peat soils with intensive sampling setup and analyses on bulk and intact soil samples. Sampling has been completed and all analyses are completed for this dataset which could be used to validate the accuracy of potential Soil Organic Carbon (SOC) maps in peat soil areas. Also, there is still soil left from the sampling points for future extension of the dataset if further analyses are required.

3.2.5. Listed parameters and methods used for the evaluation: See **Table 2** below.

Table 2. Methods used to evaluate soil health indicators in Denmark.

Parameter	Unit	Methodology
In-field analyses:		
Groundwater depth		Auger
Bulk soil analyses:		
TOC (Total organic carbon)	% DS	Dry combustion; Nelson and Sommers (1996)
TC (Total carbon)	% DS	Dry combustion; Nelson and Sommers (1996)
TN (Total nitrogen)	% DS	Dry combustion; Bremner, J. M. (1996)
C:N ratio (Carbon:nitrogen ratio)	-	
pH (in water and $CaCl_2$)	-	Electrode;Thomas (1996)
EC (Electrical conductivity, log10 transformed)	µS	Electrode;Thomas (1996)
Liquid surface tension (water repellency)	N/m	MED;Roy and McGill (2002)
Particle density	g/cm^3	Flint and Flint (2002)
NIR spectra		vis-NIR bench top and portable spectrometers
MIR spectra		FTIR bench top spectrometer
Extracellular enzyme activity	pmol/min	SEAR: Soil enzymatic activity reader
Microbial diversity index (Shannon, Simpson)	-	16S rRNA sequencing; Woese (1987)
Intact soil/100 cm^3-ring analyses:		
Bulk density	g/cm^3	Clarke and Ferré (2002)
Total porosity	cm^3/cm^3	Clarke and Ferré (2002)
Soil water retention (at h=-30, -50, -100, -300, -500, -1000 cm H_2O)	cm^3/cm^3	Clarke and Ferré (2002)
Air permeability (at h=-30, -50, -100, -300, -500, -1000 cm H_2O)	μm^2	Mass flow; Schjønning and Koppelgaard (2017)
Relative oxygen diffusivity (at h=-30, -50, -100, -300, -500, -1000 cm H_2O)		Single chamber; Schjønning (1984)
Saturated hydraulic conductivity	cm/h (log10 transformed)	
Air filled porosity (at h=-30, -50, -100, -300, -500, -1000 cm H_2O)	cm^3/cm^3	Calculated from total porosity and vol. water contents
CO_2 emissions at in situ water content and at pF2	mg C/h/kg DS	Incubation; Petersen et al. (2012)

3.2.6. Danish pilot site 2: This pilot will cover 10-20 fields selected based on gradients in land use practice, soil type/texture classes and management practice in Denmark (see location of the selected field in the **Figure 2**).

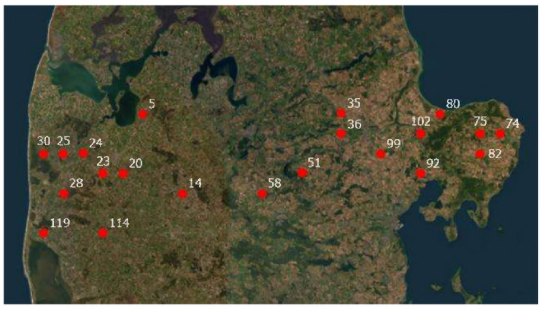


Figure 2. Map depicting locations of selected fields for Danish pilot 2.

3.2.7. Overall description of relevant Soil Health Indicators in the pilot: It is not necessarily one indicator that is relevant in this pilot. However, across large gradients in land use practice, soil type and management practices several indicators can and will be tested, and more importantly, the indicators will be tested in a spatial context. We will cover all the basic indicators, and some of the novel indicators can be tested at these sites.

3.2.8. How these indicators are evaluated: Within each field the sampling will be carried out in a spatial context, using an intensively sampling grid constructed around each of the selected points. From each field, we will sample 9 points. The landowners will be contacted in the spring of 2024, and sampling will be conducted after permission from the landowners.

3.2.9 Methods used for the evaluation:

Temporary list of soil analysis for the points in this pilot:

- In-field analysis: Near-Infrared spectroscopy (NIR), Mid-Infrared spectroscopy (MIR), earthworms, soil profile description, visual evaluation of soil structure (VESS)
- Bulk top-soil analysis: Texture, bulk density, total organic carbon (TOC), total nitrogen (TN), total phosphorus (TP), electrical conductivity (EC), cation exchange capacity (CEC), pH, NIR, MIR, liquid surface tension (water repellency), DNA

3.3. Finland

3.3.1. Description of the pilot: Finland belongs to the boreal climate zone with high regional variation in climate (see Heikkinen et al. 2013). Finland has a continental climate, with typically cold winters and relatively cool summers. Average temperatures from 1991-2020 have varied from 6.3 °C (Helsinki, south Finland) to -1.7 °C (Enontekiö, north Finland). Average annual precipitation during these years in Finland has been around 600-700 mm. Temperatures in July average from 8°C to 18°C depending on the region. February is usually the coldest month, with average temperatures ranging from -14°C to -2°C. Due to the cold winters, soils are typically frozen to a depth of between 10-60 cm, from November until up to June in northern part of Finland.

Due to the decreasing length of the growing season the further north you go, the most favourable conditions for agricultural production exist in the south, especially in the coastal area and in the southwest.

Finnish pilot is in Southwest Finland in Jokioinen, which includes fields adjacent to the Jokioinen Manor. These fields are owned by the state of Finland and are about 600 ha in total area. These fields are maintained using conventional farming techniques and crops (grass, cereals) are mainly utilized as fodder for the animals on the Luke´s experimental farms or used for field trials conducted by Luke. In this region, clay is the main soil textural class.

3.3.2. Overall description of relevant Soil Health Indicators in the pilot: For ensuring productivity of the fields, the chemical, physical and biological properties need to be at the optimal level. Out of the chemical properties, availability of plant nutrients needs to be sufficient for effective crop production. Nitrogen and phosphorus are the main nutrients potentially causing reduction of yields on the pilot region. However, concurrently, the overuse of these nutrients may cause eutrophication of surface waters. Phosphorus is also one of the soil health indicators, determined using the Olsen-P method. However, this method may overestimate soluble P content in slightly acidic soils, commonly found in the pilot site.

Clay soils in this pilot region are prone for soil compaction. Soil bulk density can be used as a proxy for soil compaction. Furthermore, compacted soils are more prone to soil erosion and restricting nutrient uptake by crops. Although agricultural soils in Finland have a higher than average organic carbon content compared to other countries in Europe, carbon content has also been known to have declined in Finnish agricultural soils (Heikkinen et al. 2013). This will influence soil physical properties and potentially on current and future agricultural productivity.

3.3.3. How this indicator is evaluated: The total area of the selected fields from the fields of Jokioinen manor will be decided at a later date after conducting of random sampling has completed (OGH). This will include fields with extreme soil pH, organic carbon and clay contents. Soil samples from these randomized sampling points will be collected during the autumn of 2024.

3.3.4. Listed indicators used for the evaluation: Bulk soil density, Olsen-P, soil pH, organic carbon content (total C), soil texture, NDVI during the growing season (autumn of 2024).

3.4. France

3.4.1. Description of the pilot: The INRAE ACBB (Agro-ecosystems, biochemical cycles and biodiversity; the long-term observation and experimentation system) long-term experimental research site on arable cropping systems is located at Estrées-Mons, Northern France (49°52'25.7"N 3°01'54.1"E, 22 ha in size). It was established in 2010 with the goal to monitor the environmental impacts and performance of arable cropping systems relevant to regional agriculture and to increase the understanding of agro-ecological processes.

Six-yearly crop rotations are implemented for eight different treatments (T1 – T8) to distinguish different agricultural management practices taking place. The main experiment consists of 32 plots and consists of six treatments (T1 – T6, n = 4) laid out in four blocks (11 ha). Treatment T5 and T6 are additionally applied to the flux footprint area of an eddy covariance tower (8 ha). In 2016, two treatments relevant to organic farming practices (T7 - T8, n = 3) were added to the experiment (3 ha).

Five specific treatment comparisons allow for in-depth analysis of the effects of **individual drivers**:

1) tillage type (T1 vs T2; 20 cm depth vs. 7 cm depth);

2) residue management (T2 vs T3; restitution/export);

3) N fertilization (T1 vs T4; 100% and 35% of reference treatment);

4) perennial (biomass) crop frequency (T3 vs T6);

5) legume frequency (T4 vs T5; low or high frequency of legumes in the rotation, as main crops or cover crops).

Three treatments resemble a **system approach,** i.e., multiple factors altered simultaneously, in which inputs (e.g., mineral nitrogen, pesticides) are reduced (T5), or eliminated (T7, T8) and associated management practices (e.g., tillage, mechanical weeding) are adapted.

Collected data are stored in an online database (AIDA: PostgreSQL relational database and web interface) managed by INRAE 1158 UMR BioEcoAgro. Raw data are digitized and can be extracted by queries fit to needs (upon request).

3.4.2. Overall description of relevant Soil Health Indicators: Soil health indicators such as soil organic carbon (SOC) stock and SOC change over time, soil moisture, soil nitrogen status can be assessed with the data collected.

3.4.3. How this indicator is evaluated: Following the research site's general sampling strategy, key variables to assess changes in biomass production, losses to the environment (i.e. N_2O, CO_2, NO_3^-, pesticides) and carbon storage in the soil are monitored both manually (i.e. yield, plant growth characteristics, soil and water chemistry, soil biodiversity) and with > 600 permanent sensors for continuous data acquisition (weather data, soil moisture, gas exchange).

More specifically, see below list of routine measurements:

Soil organic C and N stocks:

Measurement campaign every 6-7 years:

- C and N content on 5 soil layers: (depths: 0-10, 10-20, 20-35, 35-40, 40-60 cm)
- Bulk density per 5 cm layer, up to 40 cm depth.
- Calculation of stocks at equivalent soil mass.

N_2O emissions:

- Daily measurements with automatic chambers since 2012.
- The number of plots and treatments monitored has increased since 2012.
- 20 plots are currently (2022) monitored with 60 automatic chambers: T1, T2, T3, T6, T7, T8 on 3 blocks (3 chambers per plot) and T4 and T5 on 1 block (3 chambers per plot).
-

Soil water and mineral nitrogen stocks:

3 sampling dates:

- Post-harvest,
- Autumn (in some cases)
- After winter

Moisture, ammonium, nitrate

- Depths: 0-150 cm, 5 layers (0-30 cm, 30-60 cm, 60-90 cm, 90-120 cm, 120-150 cm)

Water quality:

Sampling with porous cups during the winter period (~ October to May)

- at 45 cm
- at 200 cm
- Nitrate analysis (among others)

Plants, main crops:

Samples taken at a young stage (e.g. 1 cm ear stage for cereals):

- Above-ground biomass, 4 sub samples per plot
  - of crops: organ ratio (green leaves, senescent leaves, and stems); No. of plants/m²; LAI – what does this stand for?
  - of weeds
- C and N content of plant samples, averaged per plot

Samples taken at flowering stage:

- Aboveground biomass, 4 sub samples per plot
  - of crops: organ ratio (green leaves, senescent leaves and stems); No. of ears/m²; LAI
  - of weeds
- C and N content, averaged per plot
- From 2018 to 2024, identification of weed species in the framework of the System-Eco+ project.
- Root biomass at 40 to 60 cm depth, on specific treatments (T1, T3, T4 since 2018).

Samples taken at harvest:

- Above-ground biomass, 4 subsamples per plot
  - of crops, organ ratio (grains, straws + chaff), No. of ears/m²
  - of weeds
- C and N contents averaged per plot

Plants, cover crops: samples taken before destruction:

- Above-ground biomass, 4 subsamples per plot

- of each species of the cover crop
- of weeds

- C and N content averaged per plot

3.5. Germany

3.5.1. Description of the pilot: Terrestrial Environmental Observatories (TERENO) is an EO (Earth Observatory) network that extends from the North German lowlands to the Bavarian Alps (**Figure 3**). This unique large-scale project aims to catalogue the long-term ecological, social, and economic impact of global change at regional level. Most of Germany has temperate brown and deep brown soils with their formation dependent on relief, hydrologic conditions, vegetation, and human intervention. The sandy soils in the north are mainly podzols.

The AI4SoilHealth project aims to connect with TERENO to obtain data both from relevant and available legacy time periods, and up to and including project lifespan (2023 - 2026).

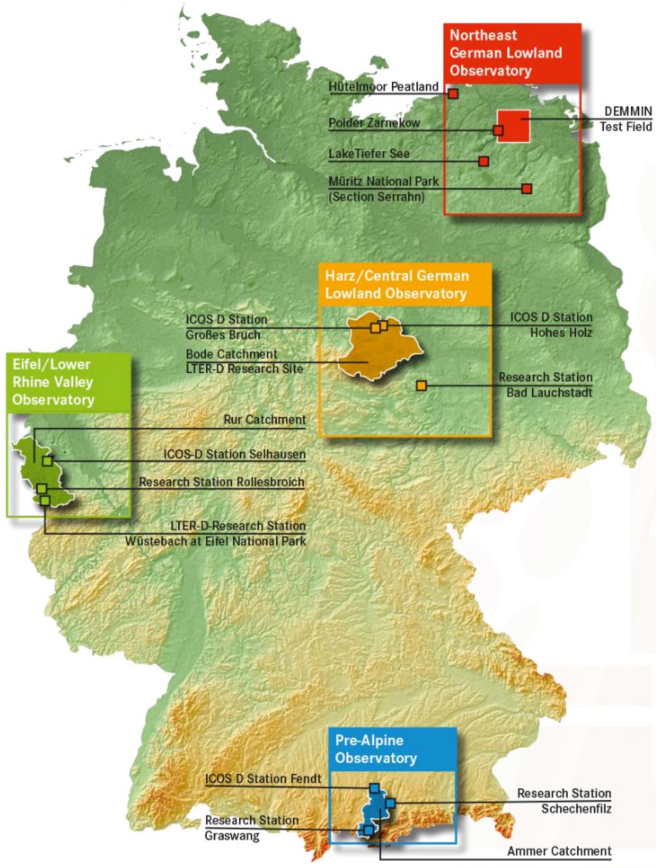


Figure 3. Location of Terrestrial Environmental Observatories in Germany.

3.6. Greece

3.6.1. Description of the pilot: Activity will focus on evaluating the impact of land degradation on soil health, farmland productivity and soil carbon sequestration potential, in Central Macedonia. It should be noted that the Region of Central Macedonia aims to create modern competitive farms based on new international data, towards the production of high-quality products that meet consumer demands, as reflected in its Regional Innovation Strategy. The Greek pilot aims at supporting the development of evidence-based conservation recommendations for policies and sustainable services for relevant economic operators tested at a modern winery "*Ktima Gerovassiliou*" – 70 ha in size. The selected test bed has been prioritized since it allows the evaluation of diverse crop management options, considering the main crop and soil types. Historic soil data have also been preserved and curated in dedicated repositories as results from previous activities.

3.6.2. Overall description of relevant Soil Health Indicators: Key indicators identified for vineyards include soil pH, texture, organic matter content, and nutrient levels as these have been known to profoundly influence the performance of crops. Maintaining optimal soil pH within a specific range is dependent on the variety has proper nutrient availability for grapevines. This will in turn be important for the soil texture as this will influence the water retention and drainage, crucial for maintaining and improving the vine root health. Moreover, a healthy balance of organic matter will benefit microbial activity and enhance soil structure, promoting retention of key nutrients and improve root development. Regular soil testing to monitor nutrient levels allows adjustment of the fertilization practices, accordingly, ensuring grapevines receive the essential nutrients required for optimal growth and fruit quality.

3.6.3. Indicator evaluation: Regular soil samplings and analyses are conducted annually from the vine-grower by the owner of the vineyard. This practice is integral to the winery's operations, as it produces wine labels exported to over 40 countries, thus establishing a valuable archive of soil chemical analyses. Over the past years, with the contribution of Aristotle University of Thessaloniki (AUTh), this archive has been augmented to include spectral measurements, resulting in the development of a spectral library at the farm scale. Additionally, a meteorological station operates at the vineyard, offering readily available weather data upon request.

3.6.4. Methods used for the evaluation: Accredited laboratories provide chemical analyses that are coupled with spectroscopy data and Remote Sensing data to upscale existing point datasets to digital maps capturing within field variability of the monitored properties.

3.7. Italy

3.7.1. Description of the pilot: The Italian pilot area, aims to cover thousands of square kilometres (km^2), rather than specific site(s) where soil samples are taken as explained in previous examples. Pilot area activity will test object-oriented dynamic soil erosion monitoring/modelling methods for future EU monitoring systems.

- ITI43 Provincia di Roma (IT)
- Potential pilot area size: 150,000 ha

- The approach will combine Land Parcel Identification System (LPIS), Copernicus and LUCAS data to compute an enhanced cover and management factor (C) for the revised universal soil loss equation (RUSLE) at monthly and event base timescales.

3.8. Netherland

3.8.1. Description of the pilot: The Boermarke Zeijen area is part of a farmer cooperation consisting of 1200 hectares (ha) of land of which 1000 hectares (ha) of farmland (e.g., grass, corn, wheat, potato, onion). The cooperation consists of 12 members of which there are five dairy farms, four arable farms, two mixed farms and one chicken farm. The pilot area is characterized by two sand ridges surrounding the stream valley of the Grote Masloot (**Figure 4**). The soil structure is characterized by a humus-containing A-horizon and a mostly tight transition to the parent material consisting of white/yellow sand. For the grasslands, the level of compaction is dependant on the location with respect to the stream valley. It is low along the Grote Masloot, with the A-horizon only 10cm thick and below it, the soil structure is moderate to poor with poor rooting and low availability of nutrients as a result. Higher on the flanks, the thickness and richness of soil activity and nutrients of the A horizon increases. For two arable plots examined in 2020, a compacted layer (plow-soil) of 5 to 10 cm was found just above the transition to the parent material.

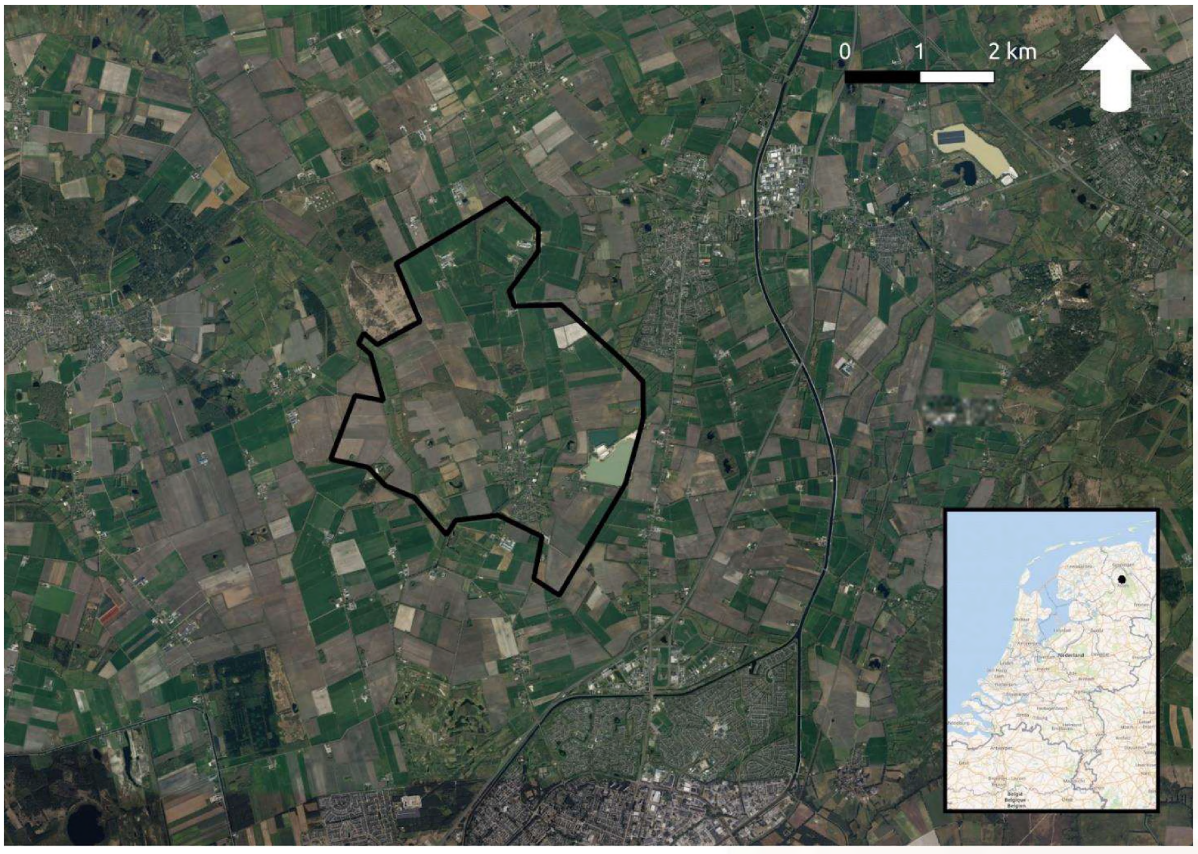


Figure 4. A map of the Dutch pilot area situated north of Assen, around the village of Zeijen (53°02'50.2"N 6°32'50.0"E). The black line is the contour of the Boermarke Zeijen area.

The pilot area has a temperate climate. On average, mean winter temperatures are around 3°C and mean summer temperatures are around 17°C, with monthly precipitation totals varying from 40 to 80 mm. In the nearby city of Groningen, the monthly average time with cloud cover is between 50-70% from October to March and 40-50% in the summer months.

There is also the Naober group active in the area, consisting of local and provincial governmental bodies, the local water board, industry, and citizen groups. In 2018, the local Water Board's groundwater measuring equipment in the area was expanded with a weather station and additional rain gauges. Furthermore, sensors were installed on a number of agricultural plots. This monitoring network now collects very detailed, location-specific, real-time information about the weather outlook, the level of surfacewater and groundwater, the moisture content of the topsoil and the water quality. Other legacy data includes in-situ samples and soil profiles (compaction and resistance) taken at 6 parcels by RMI. Within these parcels, samples were taken, and information recorded included: organic matter, texture, CEC (cation exchange capacity), occupation, Acidity, acid buffering, and lime advice, immediately available and reserves of nutrients (N, P and more), oxalate extraction (P, Fe, Al), microbial activity, microbial biomass, ratio fungi/bacteria. Furthermore, pH, EC, and composition of six groundwater samples are available.

3.8.2. Overall description of relevant Soil Health Indicators: For the arable agriculture crops, challenges exist with subsoil compaction alongside soil crusting (cementation of soil surface by small soil particles) which will reduce infiltration capacity as a result. A positive organic matter balance and as much year-round greenery as possible will help to prevent and minimise soil crusting. For dairy farms with plots near to the Great Masloot, phosphate is the biggest challenge. The phosphate status of the soil is low to neutral which provides sufficient space to supply phosphate. The phosphate soil balance, however, is negative because phosphate inputs are limited by the amount of manure on the farm and because of the ban on the supply of phosphate fertilizer on derogation farms. Because little to no buffering occurs from deeper soil layers, as well as poor soil structure, a higher supply of effective organic matter and phosphate (e.g. via livestock manure with a high P content) would contribute to better soil quality and rooting.

The **water infiltration capacity** will be studied as a soil health indicator related to compaction and soil crusting. A bare soil indicator will be tested to monitor the year-round greenery. Furthermore, the phosphorus and nitrogen levels will be measured to evaluate the nutrient balance.

3.8.3. Indicator evaluation: Bulk soil density, soil texture and soil moisture (continuous sensors + Planet Soil Water Content 100 m product) will be studied to evaluate water infiltration capacity. A bare soil indicator will be studied to evaluate land cover.

3.8.4. Methods used for the evaluation: Bulk soil density, soil texture and soil moisture (sensor + satellite) will be measured in connection to water infiltration capacity. Total nitrogen and total phosphorus will be measured with respect to the nutrient balance. Furthermore, organic carbon content (total C), cation exchange capacity and pH will be measured. Based on the remaining budget, other indicators will be measured (e.g. earthworms, soil enzymes, eDNA, NIR).

3.9. Spain

Spanish pilot site 1

3.9.1. Description of the pilot: This pilot site will explore the experimental pastures of Neiker in the province of Alava. The grazing area spans 4.5 hectares (ha) of semi-permanent pasture (42º51'11.41" N, 2º37'27.20'' W). The mean annual temperature and total precipitation in this area are 12 ºC (mean maximum of 17.9 ºC and mean minimum of 6.4 ºC) and 855 mm, respectively. The elevation averages 567 ± 4 meters, with a slope of 6 ± 3%.

In a completely randomized block design, a regenerative rotational dairy sheep grazing management system has been implemented since 2013, in comparison to a conventional grazing system. In the first treatment of free grazing, ewes freely roam and graze the entire plot for 6-10 days, followed by a rest of approximately 15 days. In the second treatment of rotational grazing, plots are divided into seven areas, with ewes grazing each area for just 1-2 days, followed by a longer rest of about 24 days. The experimental flock consists of 135-140 Latxa breed dairy ewes, split into two groups based on various factors (e.g., age, daily milk yield, live weight, and body condition scores).

3.9.2. Overall description of relevant Soil Health Indicators: Over a six-year period, rotational grazing showed significantly higher springtime grass production (30%) and topsoil carbon storage (3.6%) than free grazing. Rotational grazing supported more homogeneous pasture by the ewes and avoided the negative consequences of over and under grazing.

Furthermore, nine years after the establishment of the trial, there was a general tendency for higher relative abundances of functional genes involved in the soil cycles of C, N, P and S under rotational grazing. Five of these genes (i.e., alkaline phosphatase D, sulfite reductase α subunit, methanol/ethanol family PQQ-dependent dehydrogenase, nitrogenase iron protein and nitrite reductase) showed statistically significant differences under rotational grazing compared to free grazing.

3.9.3. Indicator evaluation: Within AI4SoilHealth, a temporal study will be performed throughout the grass growing period (from April to October 2024), where samples will be gathered every three weeks. The objective is to study these temporal dynamics of soil health descriptors, and to examine the relationships from these to the climatic and grass conditions under rotational versus free grazing.

3.9.4. Methods used for the evaluation: Physicochemical descriptors (e.g., texture, pH, organic carbon, carbonate content, total nitrogen content, extractable potassium content, phosphorus content, cation exchange capacity, electrical conductivity, metals, nitrates, ammonium, water holding capacity, POM/MAOM, bulk density, soil moisture, penetrometer, Slakes app); biological descriptors (e.g., metabarcoding of bacteria and archaea-16S rDNA, metabarcoding of fungi-ITS, metabarcoding of other eukaryotes-18S rRNA, soil respiration in situ with IRGA and in the lab, microbial biomass carbon, potentially mineralizable nitrogen; enzyme activities Digit Soil and traditional method; qPCR of fungi and bacteria, microBIOMETER, earthworm abundance); plant descriptors (e.g., grass biomass, RapidScan for NDVI, SPAD for chlorophyll content; total N, plant diversity); remote sensing-Planet (e.g., vegetation greenness-3m resolution, biomass-10m resolution).

Spanish pilot site 2

3.9.5. Description of the pilot: In this pilot site, four extensive and commercial livestock farms from the provinces of Alava and Bizkaia are being examined. This is due to the fact that these livestock farms are starting to apply regenerative rotational grazing regimes adapted to their agro-climatic conditions.

3.9.6. Overall description of relevant Soil Health Indicators: The objective is to assess the impact of land use disturbance (i.e. woodland, rotational grazing, overgrazing and non-permanent cropping) on various soil health indicators and their associated functions. While the primary focus is to compare rotational grazing with overgrazing, forests and non-permanent crops are included as references for low and high disturbance, respectively.

3.9.7. Indicator evaluation: Within AI4SoilHealth, soil health indicators are being measured in at least three sites for each land use and farm (samplings in 2023 and 2025).

3.9.8. Methods used for the evaluation: Physicochemical descriptors (e.g., texture, pH, organic carbon, carbonate content, total nitrogen content, extractable potassium content, phosphorus content, cation exchange capacity, electrical conductivity, metals, nitrates, ammonium, water holding capacity, POM/MAOM, bulk density, soil moisture, penetrometer, Slakes app); biological descriptors (e.g., oil respiration, microbial biomass carbon, potentially mineralizable nitrogen; enzyme activities, microBIOMETER, earthworm abundance).

3.10. Sweden

3.10.1. Description of the pilot: The Swedish pilot site is located in the Lönnstorp research station in Southern Sweden and is operated through the Swedish University of Agricultural Sciences. The pilot site is associated with studies of cropping systems ecology, withits focus on the design, sustainable development and assessment of arable cropping systems. The total area of the pilot site is 80 ha, of which 75% is under conventional farming and 25% is under organic farming. The main soil type associated in this landscape is loamy soils with about 15 % clay and 3 % organic material. The research station is open for research within ecology, agronomy, environmental science, agroecology and other disciplines. At present, there are around 40 active research experiments at the pilot site. From these, a handful of the experiments collect data that is directly relevant for the soil characterization tasks in AI4SH (see below). Some are also relevant for researching the information needs and responses of different stakeholders.

3.10.2. Overall description of relevant Soil Health Indicators: The Lönnstorp pilot site is part of The Swedish Infrastructure for Ecosystem Science (SITES), and is a key component for researching the application of optical remote sensing for supporting sustainable agriculture. This is referred to as SITES-spectral. SITES-spectral is an infrastructure for collecting spectral data for ecosystem monitoring. Through the infrastructure, SITES can offer data for research related to climate change, carbon and greenhouse gas balances, phenology, general ecology and biodiversity, and plant science. At Lönnstorp, spectral data is collected with both fixed and drone carried sensors. Two towers equipped with sensors monitor the soil and crops continuously and drones are flown seasonally. This data is freely available for research use.

The pilot site is also part of the EU-framework Generic bio-inventory of functional soil microbial diversity in permanent grassland ecosystems across management and climate (BIOINVENT). BIOINVENT was developed under the current European Common Agricultural Policy (CAP) and European Habitats Directives pointing

towards increased importance of permanent and extensively managed grassland systems (PEGS). The agro-ecological scheme of PEGS primarily aims at reducing fertilizer input and at the same time enhancing above- and below-ground biodiversity to profit from their support functions, while decreases in food and feed provision are only accepted to a certain extent.

Other relevant ongoing research projects at Lönnstorp include:

- SITES Agroecological Field Experiment,
- Nature-based perennial grain cropping to safeguard functional biodiversity,
- Intercropping and soil organic carbon pools,
- To capture and sequester carbon dioxide in perennial cultivation systems, and
- Diversification through Rotation, Intercropping and Multi-species cover crops.

The data collected by the projects run at Lönnstorp, including the SITES remote sensing data, adhere to the FAIR (Findable, Accessible, Interoperable, Reusable) principles and are available in open online repositories.

3.10.3. Indicator and methods for evaluation: As part of the research programs and projects at Lönnstorp, soil, water and crop data are regularly monitored and collected. Hitherto, this data has not been used for soil health research. The main reasoning for using the Lönnstorp pilot site within AI4SH has been to access comprehensive field data for evaluating the novel in-situ methods of AI4SH to be tested. The continuous spectral monitoring and seasonal drone spectral data is a valuable data source for researching temporal dynamics. Statistical regressions and machine learning modelling will be the principle methods used for evaluating the data.

3.11. Wales/UK

3.11.1. Description of the pilot: The Plynlimon research catchments have been intensively studied since the 1960s and have been leading in hydrological and hydro chemical research within the UK and internationally. The two adjacent catchments are the headwaters of the rivers Severn and Wye; roughly 19.25 km^2 (1925 ha) combined with contrasting land uses of moorland and plantation forest. Upland sites have traditionally been exploited for livestock grazing, timber production and salmon fisheries, but also are noted for areas of high conservation and amenity value as well as be important for water resources and carbon storage. The climate in the Plynlimon catchments is traditionally wet and cool with annual rainfall around 2500 mm and the mean annual temperature around 7°C. The hydrochemical record consists of more than 35 years of uninterrupted deposition and stream water samples analysed for a range of constituents including pH, alkalinity, nutrients, major cations, anions, trace metals, dissolved organic carbon (DOC) and dissolved organic nitrogen (DON). In addition to the routine monitoring, the experimental catchments have provided an outdoor laboratory for detailed plot and small catchment-scale studies e.g., biogeochemical responses of upland catchments to acid deposition, forest harvesting, agricultural management and climate change.

3.11.2. Overall description of relevant Soil Health Indicators: Within the Plynlimon catchments, the objective is to understand the state and change of physical, chemical and biological properties in order to track overall soil health. The likely measurements which will be taken in these catchments will include 1) the physical condition of soils assessed by the bulk density, aggregate stability and texture, 2) Acidity and nutrient status of soils through the measurement of pH, nitrogen and phosphorus, 3) Contaminant levels including

heavy metals and persistent organic pollutants, 4) the biodiversity status including maps of bacterial and fungal diversity as well as mesofauna abundance and diversity, 5) Soil functions including potential nitrogen mineralisation, nitrification rates, basal respiration rates, carbon substrate utilisation rates and water holding capacity. This will be assessed through moisture content at both time of sampling and field capacity, 6) Dried and frozen samples which will be archived to enable the emergence of new contaminants to be traced back and new molecular techniques to be applied as they become available, 7) integrated analysis to investigate changes in land use and vegetation composition.

3.11.3. Indicator evaluation: Collection of soil samples and field observations of habitat type and features like erosion and compaction.

3.11.4. Methods used for the evaluation: Bulk density (stone corrected), pH, conductivity, Loss on Ignition, Total Carbon-Nitrogen-Phosphorus, Olson-P, Laser particle size distribution, eDNA

3.12. Switzerland

3.12.1. Description of the pilot: The Swiss pilot site called "Lens" is a south-faced forest slope in the Valais region, located at 1060 m above sea level. The objective with this pilot site is to quantify the soil water balance in a forest during the dry season. The site is part of the Swiss Long-term Forest Ecosystem Research Programme LWF of the WSL research institute (https://www.wsl.ch/en/forest/forest-development-and-monitoring/long-term-forest-ecosystem-research/). The site is located within a natural, coniferous forest stand which consists of 150–170-year-old Scots pine (Pinus sylvestris) trees. The site is equipped with a meteorological measurement station and several soil profiles to measure soil moisture time series. Starting in 2024, a new profile will be instrumented as part of the new national soil moisture network. The soil is a Haplic Calcisol and the following properties were measured previously in 1996 (method on parentheses; Walthert et al., 2002):

- Soil texture (sedimentation)
- Skeleton content (field assessment or gravimetric)
- Bulk density
- pH (H_2O, $CaCl_2$, KCl)
- Cation-exchange-capacity including measurements of individual cations and base saturation (ammonium -chloride-extraction)
- Organic C (dry combustion) and inorganic C (sulfuric acid)
- Total nitrogen (dry combustion), organic and inorganic P (ashing, colorimetric)
- Chemical elements (HNO_3-extraction): Al, Ba, Ca, Cd, Co, Cr, Cu, Fe, K, Mg, Mn, Na, Ni, P, Pb, S, Si, Zn
- Chemical elements (NH_4Cl-extraction): Al, Ca, Fe, K, Mg, Mn, Na, Pb, Zn

3.12.2. Overall description of relevant Soil Health Indicators: The focus is on soil structural properties and their changes during the dry season. In addition to soil water content and soil matric potential (i.e., plant available water), the infiltration capacity, soil structure quality, hydrophobicity, and enzyme activity will be measured. In case of severe drought, the soil may become hydrophobic, limiting the water availability in the topsoil.

3.12.3. Indicator evaluation: The indicators are measured every month to quantify the change in structural properties during the dry season. At each sampling campaign, the measurements and the sampling are conducted at five locations.

3.12.4. Methods used for the evaluation: Infiltration capacity: the Beerkan method is used; it requires the measurement of the infiltration rate as function of time, texture information, and estimates of initial and saturated water content (Lassabatère et al., 2006)

Soil structure quality: The soil structure will be graded with values between 1 and 5 following the procedure described in VESS2020 (Johannes et al, 2020, https://ira.agroscope.ch/en-US/Page/Einzelpublikation/Download?einzelpublikationId=46489)

Hydrophobicity: The contact angle will be measured with the sessile drop method

Enzyme activity: The extracellular enzyme activity is measured using the method developed by DigitSoil (https://www.digit-soil.com/)

4. Field sampling framework

The overarching approach is that the pilot sites will support the testing and development required by WP's 3,4 and 5; mostly for upscaling. To achieve this, a matrix has been compiled by WP6 that has all the work-package work-streams as rows and then the pilot sites as the column headers (**Table 3**). Pilot sites can contribute towards 3 major outcomes. They may collect samples for use by the AI4SoilHealth project, they may contribute legacy data to test indicators, or they may be thematic sites where they focus on addressing specific issue(s). As the different work-package work-streams evolve and develop the pilot sites will be able to ask for the workstream and determine how best they can contribute or help. Soil sampling may take several different forms, some of which are described in this document following the LUCAS procedures. This will ensure that new, novel indicators are developed consistently and in a compatible way that aligns with LUCAS protocols. Pilot sites are responsible for collecting samples, compiling legacy data, and testing new tools as they emerge. They are also responsible for costs of any routine indicator analysis they may conduct. The novel indicator analysis is the responsibility of the organization developing and testing that indicator. The role of the pilots, in this case is to collect and provide the samples to work on that will form part of a centralised data set. Any analysis on single metrics conducted on this centralised data set will be performed at a single organisation to ensure analytical consistency and at the expense of that organisation.

Table 3. The overarching matrix design used by WP6 to coordinate sampling and data provision activities. (The following is a simplified, example version:)

Pilot Site:			Finland	Sweden	UK	Denmark	Germany...
	Indicators	Information					
WP6	Soil Organic Carbon	Sampling method found in D6.3	Y	Y	Y	Y	N
	...	...	N	N	N	N	N
WP3	...	...	N	N	N	N	N
	...	...	Y	Y	Y	Y	N
WP4	...	...	Y	?	Y	Y	N
WP5	...	...	N	N	N	N	N

Field sampling harmonisation on selected pilot sites will be conducted at sampling points predetermined via stratified random sampling (see 5.1.). The sampling framework is built upon previous work applied for the EU-wide LUCAS Soil monitoring (Orgiazzi et al. 2018) and involves 1) point description, 2) collection of composite samples from top- and subsoil, 3) collection of undisturbed samples from top- and subsoil, 4) in situ soil spectroscopy measurements, and 5) in situ macrofauna observations (**Figure 5**). Detailed description of the sampling, sample handling and following analyses are given in Chapter 5.

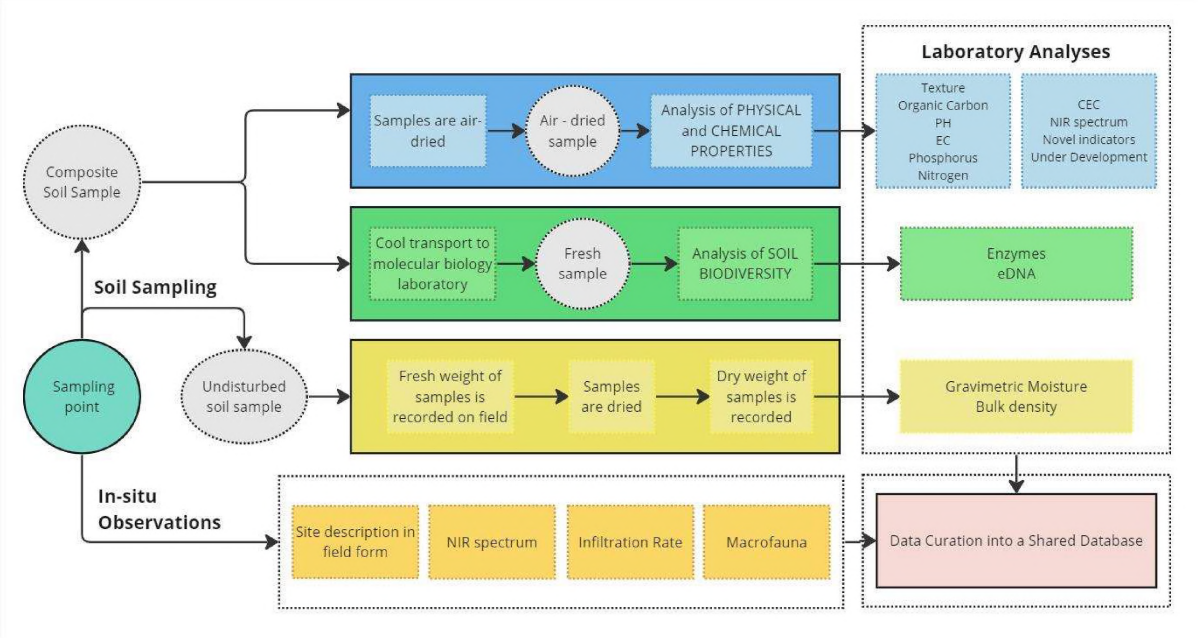


Figure 5. Schematic field sampling framework. The extent of observations and sampling carried out will vary between pilot sites.

5. Field sampling protocols

5.1. Probability based stratified random approach

5.1.1. Introduction

The selection of sampling approaches should be consistent with the aim of the survey. The following aims have been distinguished from previous works (e.g., Brus and de Gruijter, 1997; de Gruijter et al, 2006; Brus, 2022):

- estimating parameters for the population;
- estimating parameters for several subpopulations; and
- mapping the study variable.

The probability sampling (design-based approach) is required when the aim of the survey is to estimate population parameters. This is because it gives an objective assessment of the uncertainty of the estimated mean and that the coverage of confidence intervals is correct. Non-probability (model-based approach) sampling is more suitable for mapping where predictions of the target variable are calculated by the model in every pixel. The parameters of the subpopulations can be estimated using both approaches. Besides the aim, the second most important consideration involved in the selection of sampling algorithms is the presence/absence of auxiliary data (i.e. environmental covariates layers) selected from the soil forming factors such as terrain, climate, vegetation cover including cropping and human impact (McBratney et al, 2003). The presence of the auxiliary data makes it more attractive for selecting non-probability sampling as

there will be no need to estimate population means. However, when the goal is to estimate the population and to map at the same time, the probability sampling is the only relevant approach, because model-free design-based estimation of the population parameters is impossible, and therefore mapping is the only option (Brus, 2022).

Considering the above-mentioned statistical theory having universal sampling design for all aims (estimations and mapping), and the soil health law requirement of applying Bethel algorithm (Bethel, 1989) for optimal allocation of sample size to strata, a probability based balanced stratified sampling is adopted in the project. When using probability sampling, a positive probability of being included in the sample is known for any unit and the units are selected by (pseudo) random number generator. For example, when using single random sampling any unit has the same probability to be selected.

5.1.2. Probability based balanced stratified sampling

The proposed algorithm distributes the optimal sampling locations over geographical and feature space (doubly balanced) of the covariates with respect to the precision constraint (maximum allowable coefficients of variation) on the input data. The basic idea/assumption is that carefully selected environmental covariates reflect the variability of the target variables. Therefore, the precision constraints set on the precision of the estimates of the covariates should reflect the precision constraint of the survey target variables (Minasney and McBratney, 2006; Grafström and Tillé 2013; Ballin and Barcaroli, 2013; Brus 2022).

To begin this process, an optimising approach from R package '*Sampling Strata*' (Barcaroli, 2014) is applied to optimise stratification and allocate the number of samples in the strata in order to get the maximum advantage by the available auxiliary information. The approach considers each pixel having a unique combination of covariate values as an individual within a population. The algorithm iteratively forms strata from the pixels and calculates the minimum size of the sample in each stratum by optimising the minimum sample cost, sufficient to the precision constraints set on the precision of the estimates of the covariates. Because of implementing genetic algorithm searching iteratively for a suboptimal solution, the resulting sample size is significantly lower compared to the Bethel algorithm (Ballin and Barcaroli, 2013). The same allocation algorithm was used when designing sampling networks of LUCAS 2018 and LUCAS 2022 soil module (Ballin et al. 2022). The strata sizes, the stratum sample sizes and the membership of pixels to strata are used for computing the approximately equal inclusion probabilities for all population units. The equal inclusion probabilities are used for preserving the randomness of sampling which is required for estimating the (sub)population parameters (area-based estimates).

The second stage of the process involves doubly balancing the algorithm from R package '*Balanced Sampling*' (Grafström and Tillé, 2013). This is used to select the optimal sampling locations over geographical and feature space with respect to the previous stratification represented by the inclusion probabilities. The algorithm selects the random sample that is balanced on the covariates and is well spread in the geographical space. The proposed algorithm is valid both for mapping the target variables (e.g. Soil Organic Carbon density using predictive modelling) and for estimating the subpopulation parameters such as mean SOC stock in each stratum.

5.2. Field and sampling information

Information on the sampling point location, management and sampling conditions and implementation are to be completed from each point in situ in an electric or printed field form.

We need to agree, which information is included.

- Point ID
- Date, temperature, other weather information
- Land use, vegetation cover
- Topography
- Management: current and previous e.g., tillage, fertilisation, crop rotation?
- Sampling equipment
- Deviations from the protocol

5.3. Collection of soil samples

5.3.1. Composite samples

Composite soil samples are collected from each sampling point according to the scheme applied in LUCAS Soil (Fernández-Ugalde et al. 2017) with minor modifications. Topsoil (0-20 cm) samples are bulked from 5 subsamples taken from the predetermined central point and at two metre (2 m) distance from it to North, East, South, and West (Figure 2a). In subsoil sampling (20-50 cm), the number of subsamples may be reduced to 1-3 for saving labour. In case of only one subsample, it should be taken from the central point. If reducing the number of subsamples, the subsampling points should be indicated in the field form.

A gouge auger is preferred for sampling, but a spade may be used if a proper auger is not available (Fernández-Ugalde et al. 2020). Irrespective, the type of sampling equipment used is to be recorded in the field form. Moreover, care should be taken with accurate sampling depth and in avoiding mixing of the depth layers.

All subsamples from the same depth are gathered into a clean bucket (or large plastic bag), thoroughly mixed and thereafter packed into two plastic bags (air dried sample and fresh sample) labelled according to instructions given in **section 5.5**. Approximately, 0.5 l of soil is needed for the air-dry sample and 0.2 l for the fresh sample. At specific points (approximately 10 per pilot site representing different soil types), an additional topsoil sample of 1 l volume is taken in stock for later use in the development of novel indicators by WP3. These samples will be processed via air drying (**see section 5.4.1.).**

5.3.2. Undisturbed samples

Undisturbed soil samples for determining soil gravimetric moisture and bulk density will be collected into metal cylinders or rings of known volume from the middle part of the topsoil layer (0-20 cm) and subsoil layer (20-50 cm) in 1-3 replicates depending on the resources available. In case of only one subsample, it should be taken from the central point. The positions of any additional subsampling locations should be indicated in the field form.

For taking the sample, a layer of soil is first removed by a spade to reach the appropriate depth, which is to be adjusted according to the dimensions of the cylinder used. The bottom of the hole is gently levelled, and the cylinder is pressed or hammered into the soil. A block of wood or other suitable adapter should be used between the cylinder and the mallet to avoid compacting the soil. The ring is removed with a clump of soil at the bottom by digging with a spade or trowel. Finally, excess soil is removed using a knife. Instructions of the sampling with photographs are given in the LUCAS Soil Sampling instruction manual (Fernández-Ugalde et al. 2017). Please remember to record differences in sampling depth.

In the field, the cylinders are sealed with lids or placed/emptied in plastic bags to avoid loss of moisture and soil material.

5.3.3. Cases of difficult access

If the predetermined sampling point is difficult to access, it is possible to select a more suitable location within 10 m of the original central point. This should be clearly marked in the field form and accurate coordinates of the new position recorded.

In case of difficulties with the subsample locations, a more suitable position may be selected closer than 2 m from the central point or along the circumference of the sampling circle of 4 m in diameter (**Figure 6**) as instructed in the LUCAS Soil sampling (Fernández-Ugalde et al. 2017). Furthermore, a linear sampling pattern (**Figure 5**) may be adopted e.g., if the circular pattern would cause damage to crop rows.

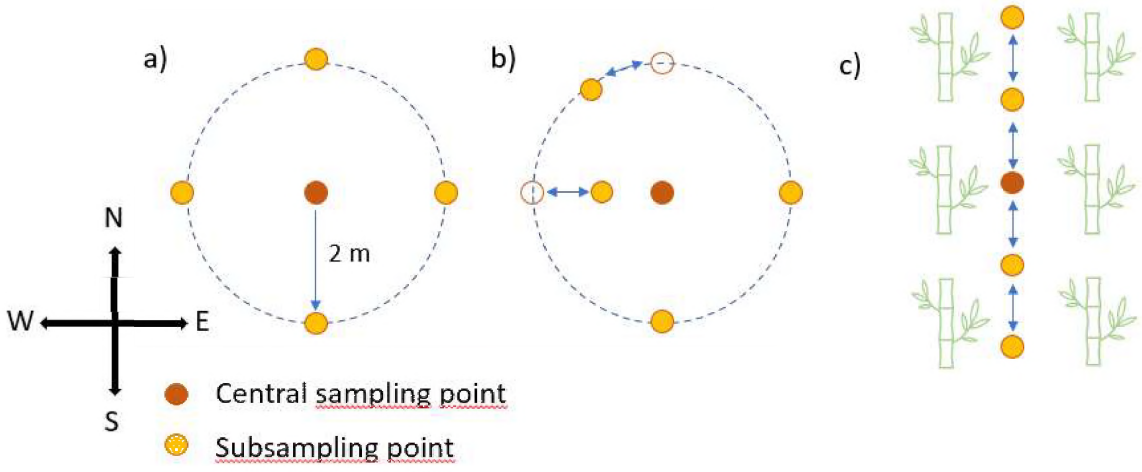


Figure 6. Soil sampling scheme in a) normal situation, b) case of restrictions with subsampling locations, or c) row cropping.

5.3.4. Repeated sampling

If the same point is sampled repeatedly during the project, there is no need to repeat the analyses of variables stable at this time scale (e.g., soil texture, total organic carbon, CEC) if the exact point can be relocated accurately. For this purpose, the central point can be marked by a stick placed on the site or measured in the field edges. A precision locator (GPS location) will also be sufficient.

5.4. Sample handling

5.4.1. Air-dried soil samples

The sample bags containing soil for the analysis of texture, organic carbon, pH, electrical conductivity, phosphorus, nitrogen and CEC should be air-dried as soon as possible after sampling. The samples can be dried at room temperature in the sample bags rolled open. Very wet samples may be spread on a tray or gently oven-dried at a temperature below 40°C. Prior to analyses, the samples are ground to pass a 2-mm sieve.

5.4.2. Fresh soil samples

The fresh soil samples are used for the analysis of soil extracellular enzymatic activity (EEA) (measured with Digit Soil SEAR) and eDNA. These samples should be placed in cold storage (-20°C), analysed or sent to be analysed without delay.

5.4.3. Undisturbed soil samples

The fresh mass of the soil in each cylinder is recorded and a sample of 20-40 g is oven-dried at 105°C to constant weight for determining soil gravimetric moisture content. Finally, the soil bulk density is calculated by dividing the dry mass of the soil by the volume of the cylinder.

5.5. Labelling of soil samples

Each sample should be labelled with an individual sample point code consisting of the country abbreviation (HR, DK, FI, FR, EL, NL, ES, SE, UK, CH), sampling year, site abbreviation (in case of several sites within the country), and an individual four-digit sample point number assigned during the randomization process by OpenGeoHub (see section 5.1). In addition, the sampling depth (topsoil or subsoil), sample type (air dry (AD); fresh (F); undisturbed (BD)) and sampling date need to be written in every sample.

5.6. Storage of soil samples

The air-dried samples can be stored in a dry place at room temperature. The fresh samples should be (deep) frozen if stored for longer periods.

6. Protocols for soil health indicators

Ensuring transparent evaluation of the validation data across pilot regions, alongside protocols for soil health indicators, with respect to those mentioned in the Soil Monitoring Law (basic indicators) as well as novel indicators (Tasks 4.3-4.5), need to be agreed among pilots.

6.1. Basic indicators (Annex II/LUCAS)

The basic indicators are to be analysed locally in an appropriately accredited laboratory using the standard methods outlined below. The cost of the analyses falls on the pilot site unless otherwise agreed. A common reference sample may be shared between pilots to assess variation between the testing laboratories.

6.1.1. Soil texture

In the EU soil monitoring law proposal, the preferred method for analysis of soil Particle Size Distribution (PSD) is based on sieving and sedimentation (**ISO11277:1998**). Alternatively, laser diffraction can be used (ISO13320:2009), which is the method currently applied in the LUCAS Soil monitoring. In AI4SoilHealth, both methods can be used though the methods are not directly comparable.

The PSD is determined for the fraction of soil passing through a 2 mm sieve. The standards do not specify cut-off limits used;

- clay < 0.002 mm,
- silt 0.002 – 0.063 mm,
- very fine sand 0.063-0.125 mm,
- fine sand 0.125 – 0.200 mm,
- medium sand 0.200-0.630 mm,
- coarse sand 0.630 – 1.250 mm,
- very coarse sand 1.250 mm – 2mm.

In brief, the **ISO11277:1998** involves optional (depending on the soil characteristics) removal of 1) organic matter by hydrogen peroxide, 2) soluble salts by washing with water, 3) carbonates by hydrochloric acid, and 4) iron oxides by sodium dithionite - sodium acetate solution. Thereafter, the soil is dispersed. Equivalent spherical diameters of the fractions exceeding 0.063 mm is determined after wet sieving and drying via dry sieving. Fractions smaller in size are determined either with a specific sampling pipette (preferred) or by less precise hydrometer analyses with periodical sampling / meter reading according to specific sedimentation times.

In the laser diffraction method (**ISO13320:2009**), dispersed soil suspension is passed through a beam of light in a laser diffraction instrument resulting in a scattering pattern that can be transformed via optical models and calculations to proportions of particles in various size classes.

6.1.2. Soil organic carbon

Soil organic carbon is to be determined after dry combustion (**ISO 10694:1995**). The method involves oxidising the total carbon in the sample into carbon dioxide by heating and determining the amount of CO_2

released via a selection of methods. For instance, carbonates can be removed from the sample beforehand by treatment with hydrochloric acid. Alternatively, organic carbon content can be calculated from the total carbon by subtracting the content of carbonates determined separately. In soils which have a pH below a value of 6.5, presence of carbonates is unlikely.

6.1.3. Soil pH

Soil pH is determined in a 1:5 (volume-to-volume) suspension both in water and in 0.01 M calcium chloride solution (**ISO10390:2005)**. The suspension is shaken for 60 min and then left to stand between 1 and 3 hours before stirring to homogeneous suspension. The pH is measured from the suspension immediately after or while stirring by a pH meter with a glass electrode.

6.1.4. Soil electrical conductivity (EC)

Soil electrical conductivity (EC) is determined in an aqueous extract of 1:5 (m/V) (**ISO11265:1994)**. In the method, the soil-water suspension is shaken for 30 min and then filtered through low ash filter paper. The specific EC of the filtrate is measured with a conductivity meter with the temperature corrected to 25 °C.

6.1.5. Soil phosphorus status

Soil phosphorus soluble in 0.5 M sodium hydrogen carbonate (pH 8.5, extraction ratio 1:20 w/V) is measured from the extract spectrometrically either as an antimony-phosphate-molybdate complex or as a phosphate-molybdate complex both reduced with ascorbic acid (**ISO11263:1994**). The extraction procedure involves addition of activated carbon and shaking for 30 min followed by an immediate filtration through a phosphorus-free paper.

6.1.6. Soil total nitrogen

Soil total nitrogen content (ammonium-N, nitrate-N, nitrite-N and organic N) is determined using a modified Kjeldahl method (**ISO11261:1995**). In brief, a soil sample of 0.2-2 g is first heat-digested in salicylic/sulfuric acid with sodium thiosulfate. Next, the digestate is cooled, whereafter a catalyst containing potassium sulphate, copper (III) sulphate and titanium dioxide is added, and the mixture is further boiled for 2-5hours. The cooled digestate is alkalized with sodium hydroxide and distilled in a distillation apparatus converting the distillate into boric acid. Finally, the distillate is titrated with sulphuric acid to a pH value of 5.0 and the total content of nitrogen is calculated based on the consumption of the sulphuric acid.

6.1.7. Cation exchange capacity

Effective cation exchange capacity (CEC)- that is the CEC at the pH of the soil- is determined using **ISO11260:1994**. In the method, the soil is first saturated with barium through leaching a sample three times in 0.1 M barium chloride ($BaCl_2$) solution (1:12 (m/V) and shaking for 1 hour. Thereafter, the soil is equilibrated overnight with approximately 0.01 M $BaCl_2$ solution. This is done by adding 0.0025 M $BaCl_2$ solution to the soil containing the remains of the 0.1 M $BaCl_2$. Finally, the soil is shaken overnight in a solution containing a known amount of magnesium sulphate, which precipitates barium as barium sulphate and saturates the exchange sites with magnesium. The excess magnesium remaining in the solution is analysed and the CEC is calculated from the magnesium concentration retained by the soil.

6.1.8. Soil bulk density

The protocol for soil bulk density determination is described above under sections of Undisturbed soil samples (**see section 5.3.2.**)

6.2. Novel health indicators

6.2.1. Spectroscopy tools for estimating Soil Health metrics (T4.3)

Introduction

Soil spectroscopy is a non-destructive analytical technique that utilizes the interaction between electromagnetic radiation and soil samples to characterize their chemical, physical, and biological properties. This technique involves measuring the reflectance, absorbance, or emission of electromagnetic radiation across different wavelengths, typically ranging from ultraviolet (UV) to infrared (IR) regions.

The significance of soil spectroscopy lies in its capacity to rapidly and cost-effectively produce comprehensive information on soil properties. Firstly, soil spectroscopy enables comprehensive soil characterization, concurrently analysing diverse parameters such as soil organic matter content, soil texture, nutrient levels, pH, moisture content, and mineral composition. This holistic approach offers valuable understanding of soil fertility, health, and productivity as well as facilitating rapid analyses compared to traditional methods, which are often laborious and time-consuming. Additionally, soil spectroscopy offers cost-effectiveness by reducing the need for expensive reagents and specialized equipment, thus lowering the overall cost per sample analysed, and increasing the capability to simultaneously assess multiple soil properties, enhancing cost efficiency as a result. Another key benefit of soil spectroscopy is that it isnon-destructive for sampling. This will help to preserve the integrity of the soil profile and enable repeated measurements over time without harming the ecosystem.

Standardized protocols are indispensable in ensuring replicability and accuracy across scientific research endeavours, including soil spectroscopy. By providing systematic guidelines for experimental procedures, data collection, and analysis, standardized protocols promote consistency in methodologies, minimizing variability between measurements. Through the inclusion of quality control measures, these protocols enable monitoring and maintaining the precision and accuracy of measurements, thereby ensuring the integrity of the data generated.

Measuring spectra in a laboratory setting offers distinct advantages rooted in controlled conditions. Laboratories enable precise control over environmental variables like humidity, temperature and lighting, minimizing external influences on spectral measurements and enhancing reliability. Equipped with suitable instrumentation optimized for high-resolution measurements, laboratories provide wider spectral ranges and greater sensitivity compared to portable field spectrometers, facilitating detailed analysis of soil properties, ensuring accuracy and reproducibility. On the other hand, measuring spectra in situ provides immediate results, valuable in applications such as agricultural management and environmental monitoring. Crucially, in situ measurements preserve the spatial integrity of the soil profile, allowing for repeated assessments over time without disturbing the natural state of the soil. By capturing spectra directly from the soil surface, in situ spectroscopy enables high-resolution spatial mapping, revealing fine-scale variations in

soil properties across landscapes. When measuring spectra in field the effects of confounding factors affect the spectral signature of the soil sample (**Figure 7**).

Figure 7. Dry and moist soil samples have visible differences that also affect the spectral signatures.

AI4SoilHealth adopts two proposed protocols that are currently developed by wide research groups:

- European Joint Programme (EJP) soils protocol for in situ measurements
- IEEE P4005: Standards and protocols for soil spectroscopy, which will be used for laboratory settings.

The protocols step by step

Naming conventions

Each measurement is assigned a unique identifier that provides information about the sensor used, the soil sample id, the soil treatment that was applied, the replication number and the depth of the soil sample. The naming conventions for the soil treatments have been adopted from the EJP SOIL 'Probefield' field protocol ('Field protocol, final version 2023'). The full dictionary of the abbreviations is shown in Error! Reference source not found.:

Table 4. Naming conventions table.

Sensor
LS – Labspec
NS – Neospectra
SE – Spectral Evolution
XV – xSpectre VIS, XN – xSpectre NIR
Treatment
NO – raw soil surface (25x25 cm)
MX – mixed field moist top and subsoil
DS – Air-dried and 2-mm sieved
Depth
T – Topsoil (0-20cm)
S – Subsoil (20-50cm)

For instance, the measurement LS_13_NO_2_T (according to the above mentioned conventions) corresponds to the second replication, a topsoil measurement with sample id 13, and is taken with Labspec spectrometer under no soil treatment.

Only spectrometers that are available to the AI4SoilHealth partners were taken into consideration to the sensor naming conventions, but it can be expanded under no restriction according to the sensor availability.

Field and laboratory protocol

This protocol has been established based on the work done by the EJP SOIL 'Probefield' project (Field protocol, final version 2023).

Measurement

When you arrive to the site of sampling and spectroscopic measurement, please note down the following:

ID of partner	**Date**	**Country**	**City**
Land use	**Crop/soil surface status**	**Weather (cloudy, sunny, …)**	**Soil temperature (°C)**
Soil moisture	**Scan type (Lab: L, Field: F)***	**Instrument***	**Pretreatment(s)***

1. Register the coordinate.
2. Take pictures of the landscape towards North (N), East (E), South (S) and West W). Further take a vertical picture downwards (D), illustrating the point for sampling/spectral measurement. Take the pictures in the order they are listed and name the pictures by point ID and direction (N, E, S, W or D): "ID"_"direction".
3. If it has not been done already, turn on spectrometer and let it heat up the required time specified by the manufacturer (see also "Annex A: Spectrometer warm-up").
4. Calibrate against white reference and scan Internal Soil Standard "Lucky Bay". "Lucky Bay" should be completely clean and can be easily polluted by touching a dirty contact probe.
5. Scan the soil according to the outlined pretreatments below.

Field Treatments

NO – Raw soil surface (25 x 25 cm)

An area of approximately 25 x 25 cm is required for the measurement.

1. If there is a plant cover, remove the plants and smoothen the topsoil surface. If the soil is bare without plants, the spectrum can be recorded directly from the original soil surface.
2. Perform five replicate scans (**Figure 8a**). Please avoid large stones and organic residues, unless you are scanning an organic soil. Aim for performing the scans at positions where it is possible to establish good contact between the soil and the probe.
3. Utilize a soil auger to take a 1 kg soil sample from 0 to 20 cm and from 20 to 50 cm, respectively. Save the soil sample in a sealed plastic bag and keep it cool until arrival at the laboratory (approximately 5 °C).

Laboratory Treatments

MX – Mixed field moist top- and subsoil

At arrival to the lab (preferably the same day as the sample was collected from the field):

1. Homogenize the field-moist sample thoroughly with e.g. a spatula.
2. Take a subsample for moisture content measurement:
    a. Note down the weight of the tray without soil (Tray)
    b. Weigh the tray and moist soil (Tray_wet)
    c. Put the sample in the oven at 105 °C for 24 hr.
    d. Take the sample out of the oven at let it cool down in a dessicator
    e. Weigh the oven-dry sample (Tray_dry)
    f. Calculate gravimetric moisture content from the following equation:
        i. w = (Tray_wet – Tray_dry)/(Tray_dry – Tray)
3. Take a representative subsample of the field-moist sample and place it in a dish.
4. Perform five replicate scans (**Figure 8b**) from the homogenized sample.

DS – Air-dried and 2-mm sieved

1. Clean a small amount of soil (a subsample of approximately 100 g) carefully from possible vegetation, roots, stones, etc.
2. Place the subsample in an open container and leave the soil sample for air-drying for approximately 2 to 3 weeks at room conditions.
3. If necessary, crush the soil sample, and thereafter pass it through a sieve with a mesh size of 2 mm.
4. Place the soil sample in a dish and perform five replicate scans directly after sieving (**Figure 8c**).
5. Archive the sample in a sealed plastic container for later (repeated) spectral measurement.

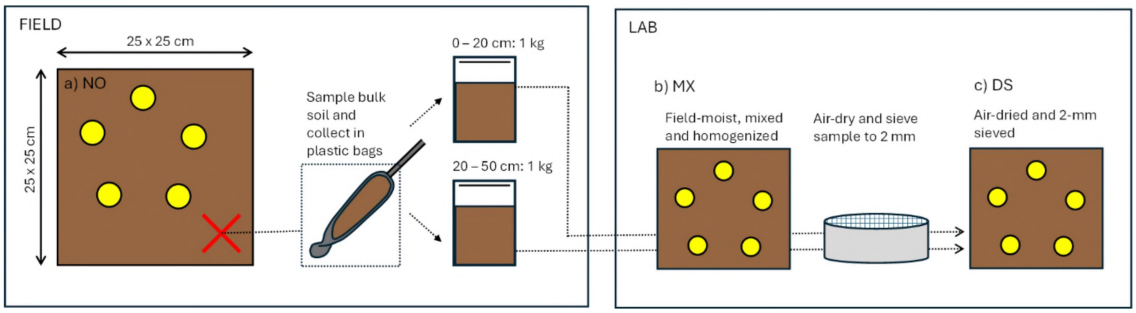


Figure 8. Workflow for obtaining spectral measurements from a) raw soil surface (NO), b) mixed field-moist top- and subsoil (MX) and c) air-dried and 2-mm sieved soil samples (DS) in the field and laboratory, respectively. The figure has been partly adopted and modified from the EJP SOIL 'Probefield' project (Field protocol, final version 2023).

Equipment preparation

Spectral measurements are prone to effects from ambient factors such as high temperature, source of illumination or humidity. To this end, it is highly suggested to control the environment of the laboratory (i.e. set room temperature to 22°C, or relative humidity to 50%), while each spectrometer requires to follow a warmup routine according to the manufacturer (or **Annex A: Spectrometer warm-up**). The warmup period varies between different instruments, as for benchtop it ranges from 30 to 90 minutes, while for portable it might be not required. In any case, each user is suggested to either refer to the manufacturer or conduct a small exercise that can be found at "**Annex A: Spectrometer warm-up**".

Measurement

White calibration

Calibration typically includes white and dark reference calibrations, where the white reference serves as a baseline for reflectance measurements, and the dark reference accounts for instrumental noise and stray light. Reflectance (R) is then calculated using the formula:

$$R = \frac{S_{sample} - S_{dark}}{S_{white} - S_{dark}}$$

Where S is the signal from the sample, the white calibration panel, and the dark calibration panel. Some sensors do not require a dark calibration panel as it is calculated internally by integrating a dark current measurement directly into the sensor's electronics.

Spectralon is suggested to be used as a white calibration medium as it is a type of highly reflective material commonly used as a standard reference in spectroscopy and radiometry. It is a diffuse reflectance standard, meaning it reflects light equally in all directions, making it ideal for calibration and standardization purposes.

White calibration is a very important measurement as any deviations captured during itpass to all subsequent measurements. To this end, it is suggested that every laboratory should produce a reference measurement every time they purchase spectroscopy equipment, or they perform any kind of periodic maintenance (i.e. calibration), in order to monitor the instrument's condition. Also, as mentioned, spectrometers are affected by environmental or internal instabilities (mainly temperature) resulting to deviations from reference measurements. To this end, it is suggested that after a predefined set of measurements, which is five soil samples, the user must measure the white reference and if the result is not close to 100% reflectance to each wavelength (except the bounds of the spectral range), then white calibration must be performed again.

Standardization

Two main factors can influence soil spectra: non-systematic and systematic. Non-systematic effects arise from uncontrollable phenomena, such as random noise and uncertainties, and can lead to noisy and inconsistent soil spectra if not minimized through consistent protocol maintenance. This involves keeping instrumentation factors and sample preparation constant using an agreed-upon protocol. On the other hand, systematic effects result from controlled responses that may vary between instruments but remain constant within a selected protocol.

To address systematic effects, factors like the white reference sample, spectral configuration, and environmental conditions must be kept constant or monitored. As these effects can vary between laboratories, efforts to establish measures for alignment between different spectral libraries are essential, especially given the lack of a standardized method and the rarity of cross-calibration between laboratory infrastructures. To minimize this effect, the use of internal soil standards is essential for standardizing spectral measurements across different setups, as they serve as reference materials with known spectral properties, allowing for consistent calibration and normalization of spectral data acquired from diverse sources. As proposed at Ben Dor, et al. (2015) and adopting the idea of Pimstein et al. (2011, an inexpensive spectrally featureless material that can be found in abundance and has similar spectral characteristics with soil, and is stable can be used as a standard for the abovementioned scope. To this end, two samples from sand dunes in Australia, Wylie Bay (WB) and Lucky Bay (LB) can be used. LB solely or both the two standards are measured after the white calibration step (**Figure 9**). The acquired spectral signature is then compared to their reference spectral signature, and a correction factor is calculated per wavelength which is the ratio of the acquired reflectance to the reference reflectance. Then, each of the subsequent soil spectral measurements are corrected by calculating the elementwise product of the correction factors and reflectance for each wavelength.

The need for this spectrum correction is also observable from

Figure as the same soil sample was measured under the exact same conditions and resulted to observable reflectance values among different spectrometers.

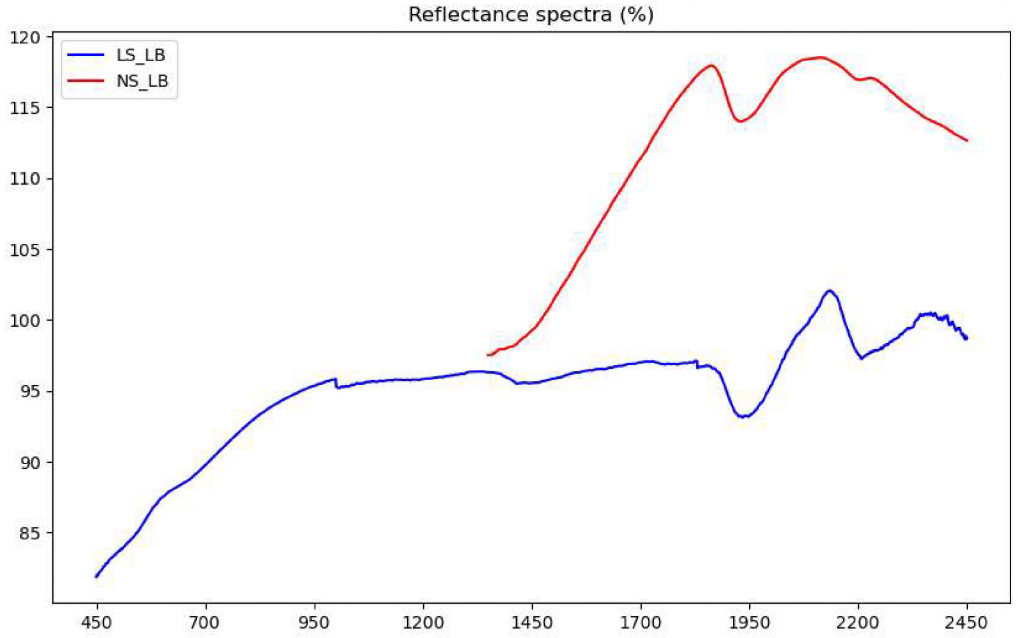


Figure 9. Reflectance spectra from LB standard for two different spectrometers

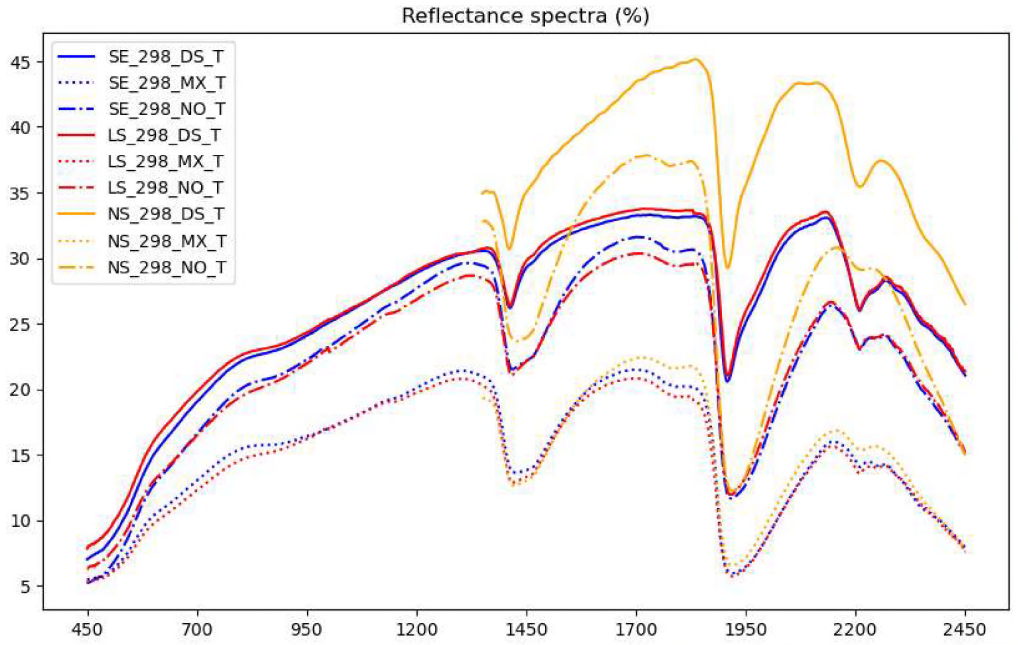


Figure 10. Reflectance spectra of a single soil sample under different treatments and as captured by different sensors. The continuous lines correspond to the DS treatment, the dotted lines correspond to MX and dashed-dotted lines correspond to NO treatment. Yellow lines are the spectra captured from NS, blue from SE and red from LS.

Annex A: Spectrometer warm-up

The warm-up time for VIS-NIR-SWIR (Visible-Near Infrared-Shortwave Infrared) spectrometers can vary based on several factors inherent to their design and operation. One key factor influencing warm-up time is the type of detector employed within the spectrometer. Whether it's a charge-coupled device (CCD), photodiode array, or specialized sensor like InGaAs for SWIR, each detector type may have distinct warm-up requirements based on its operational principles and material properties. Additionally, the precise temperature control necessary for optimal performance can affect warm-up time; instruments with more sophisticated temperature regulation systems may achieve stable operating conditions more quickly. Optical components such as lenses, filters, and gratings also require time to reach thermal equilibrium, particularly in spectrometers with complex optical setups or higher precision components, thus contributing to warm-up time disparities. Furthermore, the electronics within the spectrometer, including signal processing circuits and data acquisition systems, must stabilize, with more advanced electronics potentially prolonging warm-up time. Moreover, power consumption and calibration procedures may influence warm-up duration, with spectrometers designed for portable or low-power applications typically having shorter warm-up times. Additionally, achieving an optimal signal-to-noise ratio (SNR) is crucial for accurate spectral measurements, and the warm-up time may also include considerations for SNR stabilization, particularly in spectrometers where longer warm-up times can lead to improved SNR.

To achieve the optimal warm-up time for your spectrometer, the user is suggested to conduct the following exercise.

- Select two soil samples under the DS treatment: dark and bright.
- Turn on your spectrometer and lamp for 20 min.
- Configure your spectrometer to a large number of measurements (i.e. 40) – if it is supported by the spectrometer.
- Calibrate with white reference (WR) and convert all measurements to reflectance.
- Measure the soil sample the number of times set before at the same geometry each time with the same configuration with WR calibration.
- For the same soil samples, repeat all measurements after an extra 20, 40, 60 or more-time increments.

For each warm-up period, calculate the signal stability with your preferred metric (e.g. standard deviation per wavelength, coefficient of variance, etc.). Set a threshold according to your desired accuracy and when this is achieved, find the minimum warm-up period that achieves the desired stability. For the SE it was found that the optimal warm-up period is 90 minutes.

6.2.2. Method development for DNA sequencing (T4.4)

Purpose of this protocol

This protocol is designed to be used for collecting samples for molecular analysis (i.e eDNA, microbial community profiling). It is recommended that samples are paired to as many other measurements as possible, particularly nutrients and enzyme activities. Soil samples should be as representative as possible- soil cores should be homogenized, and sub-samples taken and preserved.

This protocol also makes suggestions for downstream analysis should partners wish to do the DNA extraction and sequencing themselves.

Protocol date and version.

13/03/2024 Version 1

Revise March 2025

Authors

Joe Taylor UK Centre for Ecology and Hydrology (UKCEH) joetay@ceh.ac.uk

Lur Epelde (NEIKER) lepelde@neiker.org

Soil sampling and preservation

Materials and Reagents:

- Zymo DNA RNA Shield (Catalogue Number: R1100-250) https://zymoresearch.eu/collections/dna-rna-shield
- Sterile collection tubes (2 mL microcentrifuge, 5mL Bijou tubes)
- Gloves
- Sterile spatula or spoon- sterilise with ethanol or IMS
- Marker for labelling
- GPS device or location recording tool
- Plastic bags for transporting samples
- Disposable wipes or tissues for cleaning tools
- Ethanol or IMS for cleaning
- Pipette to add Zymo DNA RNA shield

Procedure:

Preparations:

a. Wear gloves throughout the entire process to avoid contamination.

Sample Collection:

From a homogenised soil sample, using a sterile spatula or spoon, collect approximately 0.25-0.5 g of soil. Use the spatula to scoop the soil into the labeled collection tube.

Preservation with Zymo DNA RNA Shield

a.Add Zymo DNA RNA Shield to each soil sample at a ratio of 1.5-3 times the weight or volume of the soil. For example, if you collected 0.25 g of soil, add 0.75 ml of Zymo DNA RNA Shield. -Note if an accurate pipette is not available volumes can be approximate- soil samples should be fully submerged in solution.

b. Mix the soil and Zymo DNA RNA Shield thoroughly by shaking the collection tube.

c. Incubate the samples at room temperature for at least 5 minutes to ensure complete stabilization of nucleic acids.

d. Zymo DNA RNA shield preserves the samples for short term storage at room temperature, we would not advise longer than two weeks storage at room temperature-they should be frozen and then could be thawed for shipping.

Transporting samples:

a. Seal each collection tube tightly to prevent leakage during transportation.

b. Place the tubes in a plastic bag to further protect against potential leaks.

Shipping samples:

If the number of samples is not too large, it is recommended that samples are shipped to Lur Epelde at NEIKER for DNA extraction (then NEIKER will be able to make a collective shipment to the selected sequencing company, obtaining a more competitive price for everyone). Samples can be shipped at room temperature using next day or 48 hour courier. Please email Lur lepelde@neiker.org in advance of sending samples.

DNA extraction

It is recommended to use the Zymo Quick-DNA Fecal/Soil Microbe Kits

https://zymoresearch.eu/collections/quick-dna-fecal-soil-microbe-kits

Qiagen DNeasy PowerSoil Pro Kit will also work- but is less compatible with Zymo DNA RNA shield (https://www.qiagen.com/us/products/discovery-and-translational-research/dna-rna-purification/dna-purification/microbial-dna/dneasy-powersoil-pro-kit).

PCR and sequencing

While sequencing providers can vary it is recommended to use a service that uses the Illumina sequencing platforms, either the MiSeq, NextSeq or NovaSeq. Example costs are available from the sequencing provider here: https://imr.bio/pricing.html

It is approximately 20 € per sample for each target- i.e., Bacteria, Fungi, and Eukaryotes – so 60 € per sample total cost. We would recommend a minimum of 30,000 sequence reads per sample. Illumina sequencing can be 2 x250 bp or 2 x300 bp as the sequencing provider recommends.

PCR primer sets used are recommended to be the following targets - if money is only available for one or two targets then Bacteria and Fungi are the priority groups. Also, if money is only available for one of the sampling depths, analyzing the topsoil 0-20 cm sample is recommended (biota is more abundant in the topsoil):

Bacteria (Also amplify Archaea- but Archaea are usually low abundance)

16S Rrna gene V4 region

It is recommended to follow Earth Microbiome Project protocol: https://earthmicrobiome.org/

515f Modified GTGYCAGCMGCCGCGGTAA

806r Modified GGACTACNVGGGTWTCTAAT

Walters, W., Hyde, E. R., Berg-Lyons, D., Ackermann, G., Humphrey, G., Parada, A., Gilbert, J. A., Jansson, J. K., Caporaso, J. G., Fuhrman, J. A., Apprill, A., & Knight, R. (2016). Improved bacterial 16S rRNA gene (V4 and V4-5) and fungal internal transcribed spacer marker gene primers for microbial community surveys. mSystems, 1(1), e00009–15. http://doi.org/10.1128/mSystems.00009-15

Fungi

ITS2 region

gITS7 GTGARTCATCGARTCTTTG (Ihrmark et al. 2012)

ITS4 TCCTCCGCTTATTGATATGC (White et al. 1990)

Ihrmark, K., Bödeker, I.T., Cruz-Martinez, K., Friberg, H., Kubartova, A., Schenck, J., Strid, Y., Stenlid, J., Brandström-Durling, M., Clemmensen, K.E. and Lindahl, B.D., 2012. New primers to amplify the fungal ITS2 region–evaluation by 454-sequencing of artificial and natural communities. FEMS microbiology ecology, 82(3), pp.666-677.

Total Eukaryotes

18S rRNA gene V4 region

E572F CYGCGGTAATTCCAGCTC

E1009R AYGGTATCTRATCRTCTTYG

Comeau, A.M., Li, W.K., Tremblay, J.É., Carmack, E.C. and Lovejoy, C., 2011. Arctic Ocean microbial community structure before and after the 2007 record sea ice minimum. PloS one, 6(11), p.e27492.

If global comparisons needed then could use the protocol from EMP https://earthmicrobiome.org/protocols-and-standards/18s/ but would recommend 18S V4 as it is a longer amplicon fragment.

Archaea

ARCH-349 GYGCASCAGKCGMGAAW

ARCH-806R GGACTACVSGGGTATCTAAT

Bioinformatics

Recommended analysis is DADA2 https://benjjneb.github.io/dada2/

UKCEH can offer analysis service and advice joetay@ceh.ac.uk

6.2.3. Development of soil macrofauna observation and measurement tools (T4.5)

Macrofauna are invertebrates from a variety of broad taxonomic groups (e.g. Lumbricina [earthworms], Isopoda [woodlice], Chilopoda [centipedes], Diplopoda [millipedes], Gastropoda [slugs and snails]) (see **Table 4**), and generally with body widths between 2 mm and 20 mm (Barrios, 2007). The abundance and composition of macrofauna in the litter and soil layers can provide useful information on the impacts of land use and management on the distribution of biodiversity, and indicators of functions.

Various methods are used to sample and enumerate macrofauna and this can differ when targeting different broad taxonomic groups. There have been significant international efforts to collate existing data (e.g. Philips et al. 2019, Mathieu et al. 2022, Lavelle et al. 2022). Subsequently, important considerations for macrofauna approaches taken in AI4SoilHealth are standardization of methods and the ability to integrate with these existing initiatives and databases (and contribute to them).

The core macrofauna sampling approach will build upon the standardized method described in Mathieu et al. (2024; https://zenodo.org/records/10479451) for the #GlobalSoilMacrofauna initiative. Briefly, this consists of manual sorting of macrofauna from the litter layer and the soil monolith below, and enumeration of 34 possible broad groups; many groups are not prevalent, and their presence may also be dependent on biogeography and climate (e.g. termites, scorpions, mantids, Embiopterans), so it is more typical to find 5-10 groups in a sample.

Table 4. Macrofauna groupings for enumeration in samples.

Macrofauna groupings		
Earthworms	Diplopoda (Millipedes)	Orthoptera
Ants	Chilopoda (Centipedes)	Gastropoda (Snails, Slugs)
Hymenopterans (non-Ants)	Pauropoda	Diplura (Bristletails)
Termites	Symphyla	Embioptera (Web spinners)
Coleoptera - Adults (Beetles)	Isopoda (Woodlice)	Hirudinea (Leeches)
Coleoptera - Larvae (Beetles)	Diptera - Adults (Flys)	Mantodea (Mantis)
Araneae (Spiders)	Diptera - Larvae (Flys)	Phasmida (Stick insects)
Amblypygi (Whip spiders)	Cockroaches	Protura

Opiliones (Harvestmen)	Heteroptera (True bugs e.g, shield bug)	Thysanoptera (Thrips)
Pseudoscorpiones	Homoptera (True bugs e.g. aphids)	Zygentoma (Silverfish)
Scorpiones	Dermaptera (Earwigs)	
Solifugae (Camel spiders)	Lepidoptera - Larvae (Butterfly/Moth)	

A key objective of Task 4.5 is to evaluate image-based sensor methods to provide fingerprints of soil macrofauna abundance and composition. Image-based approaches using a mobile phone camera provides opportunity to streamline data capture for end-users and engage the wider public. We are currently undertaking work to develop a method for this as part of AI4SoilHealth sampling; this includes testing existing online platforms with trained computer vision engines (iNaturalist, ObsIdentify) for their ability to identify different groups of macrofauna and develop models for image-based estimation of macrofauna biomass and abundance. Consequently, the macrofauna methods have been separated into 'Core Macrofauna'' and 'Macrofauna+', so that pilot sites (and any additional sampling sites) can decide whether to provide basic macrofauna data or also contribute to the development of image-capture approaches.

Core macrofauna method:

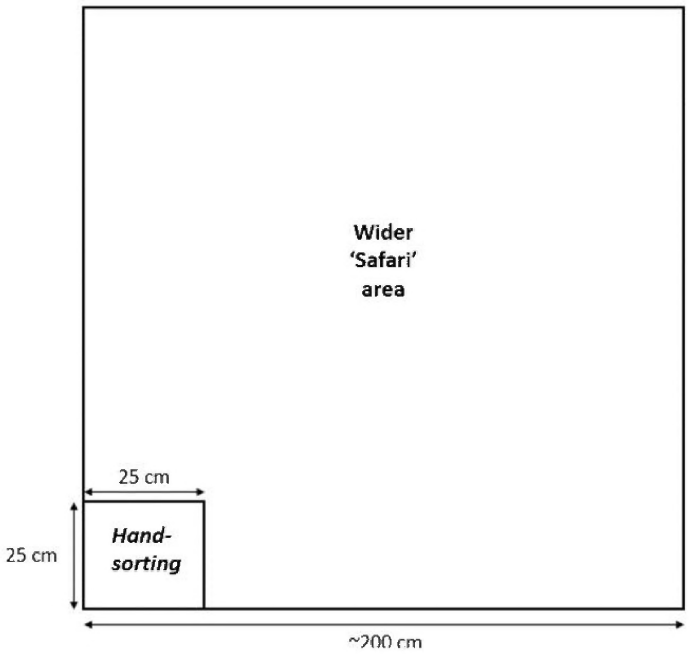
Materials and equipment

- 25 cm x 25 cm quadrat or flags (for marking out)
- Spade
- Ruler
- Plastic sheet
- Camera/mobile phone camera
- Sorting tray
- Specimen pots

Method

- The spatial unit of macrofauna sample is a 25 cm x 25 cm square.
- Corners of the sample are marked out accurately (using a quadrat or flags) and a monolith is excavated to at least 10 cm using a flat spade, with the depth recorded (ideally sampling should be 30cm depth).
- The monolith is placed on a plastic sheet, and a photograph is taken of the side view of the monolith with a ruler for scale.
- Hand-sorting begins by searching the litter layer, with different groups placed into separate specimen pots.
- The soil is then hand-sorted, with different groups placed into separate specimen pots.
- The number of individuals of the different groups in specimen pots for litter and soil layers should then be enumerated and recorded.

- The 25 cm x 25 cm sample should be replicated 5 times in a plot (e.g. representing a habitat or experimental treatment).
- OPTIONAL: A wider 200 cm x 200 cm area ('Safari' area) is searched, focusing on the litter layer, with the aim to determine presence of additional macrofauna groups not recorded in the 25 cm x 25 cm Hand-sorting square (see diagram below).



Macrofauna-plus methods (covering image capture*):

**These methods are still under development and may be altered depending on the outcome of testing and development*

Materials and equipment

- Specimen pots (from the 'Core Macrofauna' steps above)
- White sorting tray (A4 paper size)
- Weighing scales (accuracy to 0.1g is sufficient)
- Mobile phone camera
- Water
- Paper towels

Method

- There are two strands for image-capture, you can contribute to one or both:
    - A: Whole-group image capture

- B: Individual specimen image capture

- A: Whole-group image capture
  - Take an initial photo of the sample ID code (e.g. on paper), so that images taken after this can be linked to the sample using image metadata if necessary.
  - Record the mass of each group in the specimen pots.
  - Empty specimen pot containing a macrofauna group into the A4 white sorting tray, ensuring that most litter/soil material is (for earthworms, these can be washed in some water and lightly dried using paper towels).
  - Try to ensure that specimens are spaced out on the sorting tray.
  - Take photograph from above (approximately 25 cm) with the borders of the sorting tray at the edge of the image.
- B: Individual specimen image capture
  - Take initial photo of sample ID code (e.g. on paper), so that images taken after this can be linked to the sample using image metadata if necessary.
  - Add individual specimens from specimen pots to the sorting tray and take photograph including the entire body.
  - Ensure photograph is in focus on the specimen.
  - Continue until all specimens are photographed.

6.2.4. Soil enzymatic activity

DigitSoil - An easy way to assess soil extracellular enzymatic activity (EEA) of hydrolytic enzymes.

The measurement principle of SEAR (Soil Enzymatic Activity Reader) consists of the release of fluorescent compounds caused by hydrolysis from artificial substrates that are brought into contact with soil enzymes. An increase of fluorescence signal over time is registered by the reader that is equipped with excitation light matching fluorescent products and a sensitive detection system. Based on a standard curve containing reference chemicals in known concentrations, the registered values are converted to concentrations over time. As the standard curves for both reaction products (4-Methylumbelliferone - MUF; 7-Amino-4-methylcoumarin – AMC, Table 5) are incorporated in the patch, it accounts for soil specific artifacts. The reaction gel includes substrates targeting five different enzymes (**Table 6**). The measurements result in a release rate of products for a given period and area of contact (pmol/min).

Table 5: Targeted enzymes and their matching artificial substrates used in Digit Soil reaction gel as well as their natural substrate equivalents.

Abbreviation	Full chemical name of used substrate	CAS No
MUF	4-Methylumbelliferone	90-33-5 - 176
AMC	7-Amino-4-methylcoumarin	6093-31-2

Table 6. Targeted enzymes and their matching substrates as well as their natural substrate equivalents

Abbreviation	Full chemical name of used substrate	Reaction product	Target group of enzymes	Equivalent natural substrate
NAG	4-Methylumbelliferyl N-acetyl-β-D-glucosaminide (CAS 37067-30-4)	MUF	β-glucosaminidase	hydrolysis of chitin
GLS	4-Methylumbelliferyl β-D-glucopyranoside (CAS 18997-57-4)	MUF	β-glucosidase	beta-D-glucosides and oligosaccharides
MUP	4-Methylumbelliferyl phosphate (CAS 3368-04-5)	MUF	phosphatases (Phosphomonesterases)	compounds with phosphate-monoester
MUX	4-Methylumbelliferyl-β-D-xylopyranoside (CAS 6734-33-4)	MUF	β-xylosidase	hydrolysis of hemicellulose (beta-D-xylans and xylobiose)
LAP/ LEU	L-Leucine-7-amido-4-methylcoumarin hydrochloride (CAS 62480-44-8)	AMC	leucinaminopeptidase	hydrolysis of leucine at the N-terminus of polypeptides and proteins

Materials

- Digit soil reader (SEAR; see **Figure 11**) and compatible adaptors (cuvette, sieve; see **Figure 12**)
- Reaction gel (containing reagents)
- Fresh soil (about 15g per 1 measurement of set of enzymes with one reaction gel)
- Metal spoon

Method

- Sampling - general sampling protocols applicable
- Samples should be analyzed fresh, avoiding long term storage is recommended due to known effects of storage soil enzymatic activities (Dick 2011)

Sample processing:

1) (Optional) Sieve fresh soil (4 mm) to remove excess roots and stones that could prevent good contact of the soil with the reaction gel (can be sieved directly to the cuvette with accessory sieve)

2) Place your soil sample into the cuvette, approx. 10-15 g of soil is needed, depending on the weight, should fill up the cuvette resulting in a flat surface

3) Take one reaction gel and remove the protective foil on one side (indicated with a wing)

4) Place the reaction plate with the open side facing the soil and press gently

5) Put the cuvette with patch into the SEAR drawer, close it firmly and start the measurement

6) The measurement runs for 40 min

7) After measurement is finished, remove measured sample, clean the cuvette and insert a new sample

8) Once measurements are finished upload data to Digit Soil cloud where the data analysis is conducted and results are sent back

Figure 11. SEAR sensing system: reader body (right), cuvette with a reaction plate on top (middle).

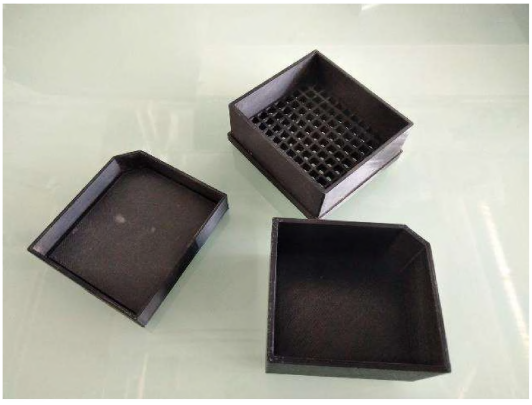


Figure 12. SEAR accessories: soil cuvette (left), press (right), sieve (back).

6.2.5. Infiltration rate

The infiltration rate provides information on the infiltration capacity (risk of ponding and erosion) and the hydraulic properties of the topsoil. Using the Beerkan method (Lassabatère et al., 2006), the infiltration time of a water layer < 1 cm in height within a ring inserted into the soil is measured until a constant rate can be observed. The amount of water is relatively small and depends on the size of the infiltration ring (less than 1 litres for a ring radius of 4 cm, almost 5 litres for a radius of 10 cm). The infiltration ring can be cut out from a can or can be bought.

Requirement to get the hydraulic properties: you need information on soil texture (sand, silt, and clay content) and bulk density of the topsoil.

Material

- Stopwatch to measure the infiltration time
- Metallic or plastic ring of 4 cm (this is, for example, the size of HYPROP cylinder) to 10 cm in radius
- 2-3 'Measuring vessel' (plastic bottles or cylinders) with marks for the amount to fill ≤ 1 cm water in the ring (for example, for a ring radius of 4 cm, 50 mL of water should be in one 'Measuring vessel'
- Canister of water to refill the 'measuring vessel'
- Plastic bags to take a sample of wet and a sample of dry soil (take a 'handful' of wet soil within the ring and a 'handful' of dry soil from the vicinity of the ring)
- Scissors to cut the grass within the ring
- Material to write down the time intervals (field laptop, pencil and paper)

Procedure

- Prepare a flat surface (e.g., cut the grass with scissors) and insert the ring 1 cm into the soil
- Prepare 2-3 'measuring vessels' with the right amount of water (for example, for a ring radius of 4 cm, 50 mL)
- Start the stopwatch and add one unit of water to the surface within the ring (dump the content of one 'measuring vessel' in the ring)
- When the water is infiltrated, note the time and add the next unit of water
- Repeat the addition of water until you have added 8 - 15 units or until the infiltration rate becomes constant (i.e., the time interval for the infiltration of one water unit is relatively constant)
- Take a sample from the wet soil within the ring and a sample from the dry soil close to the ring
- Measure the weight of the dry and wet samples before and after drying at 105 degrees in the oven
- Fill out the protocol shown below and send the data to peter.lehmann@env.ethz.ch to obtain the soil hydraulic properties

Table 6. Protocol sheet for determining infiltration rate.

Clay content [%]:		Sand content [%]:	
Silt content [%]:		Bulk density [g/cm3]:	
Mass inside the ring, before oven drying [g]:		Mass outside the ring, before oven drying [g]:	
Mass inside the ring, after oven drying [g];		Mass outside the ring, after oven drying [g]:	
Ring diameter [cm]:		Added water unit (volume of 'measuring vessel') [mL]:	
Time to infiltrate unit 1 [sec]:		Time to infiltrate unit 9 [sec]:	
Time to infiltrate unit 2 [sec]:		Time to infiltrate unit 10 [sec]:	
Time to infiltrate unit 3 [sec]:		Time to infiltrate unit 11 [sec]:	
Time to infiltrate unit 4 [sec]:		Time to infiltrate unit 12 [sec]:	
Time to infiltrate unit 5 [sec]:		Time to infiltrate unit 13 [sec]:	
Time to infiltrate unit 6 [sec]:		Time to infiltrate unit 14 [sec]:	
Time to infiltrate unit 7 [sec]:		Time to infiltrate unit 15 [sec]:	
Time to infiltrate unit 8 [sec]:		Note: It is more convenient to note the cumulative time	

6.2.6. Visual Estimation of Soil Structure Score

The quality of the soil structure can be quantified visually using a specific form of 'spade test', taking a soil sample of 25-35 cm depth with a spade. The analysis of the aggregate size and shape before and after breakage/opening provides a soil structural score (Sq-value) between 1 and 5. The method we apply for our study sites in Switzerland (Johannes et al, 2020) quantifies the score directly in the field and is a simplification of a more detailed visual assessment conducted in the lab. The score of the visual assessment can be linked to the ratio of clay and organic matter (Johannes et al. 2017) and is a good proxy for the potential of structure formation. The experimental procedure we conduct is described in detail in the following link.

https://ira.agroscope.ch/en-US/Page/Einzelpublikation/Download?einzelpublikationId=46489

The document includes examples of the main score classes shown below (**Figure 13**).

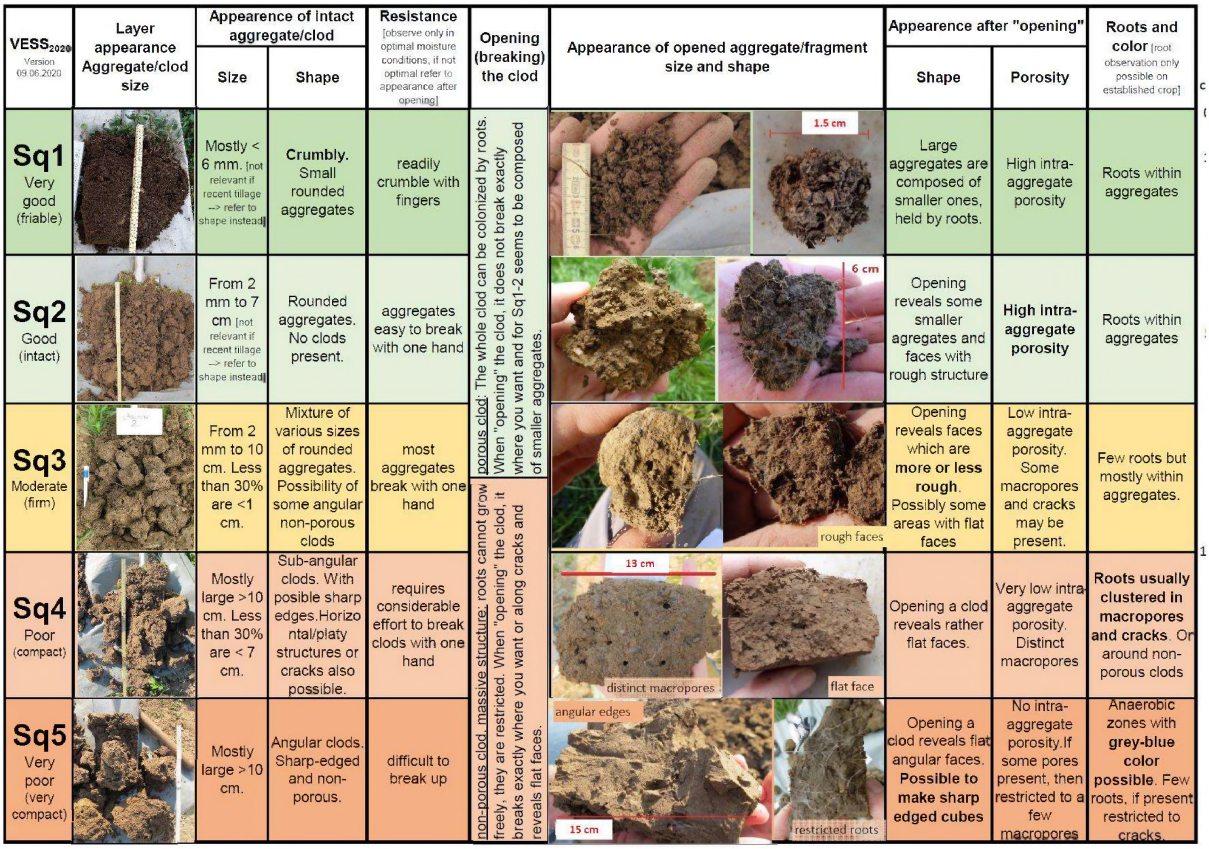


Figure 13. Scoring of soil structure based on visual assessment.

7. References

Ballin, M., Barcaroli, G. (2013). Joint determination of optimal stratification and sample allocation using genetic algorithm. Survey Methodology 39: 369–93.

Ballin, M., Barcaroli, G., Masselli, M. (2022). New LUCAS 2022 sample and subsamples design: Criticalities and solutions. https://ec.europa.eu/eurostat/web/products-statistical-working-papers/-/ks-tc-22-005.

Barcaroli, G. (2014). SamplingStrata: An R package for the optimization of stratified sampling. Journal of Statistical Software 61, issue 4. https://doi.org/10.18637/jss.v061.i04

Barrios, E. (2007) Soil biota, ecosystem services and land productivity. Ecol Econ 64: 269-285.

Bethel, J. (1989). Sample Allocation in Multivariate Surveys. Survey Methodology 15: 47–57.

Bremner, J. M. (1996). In D. Sparks, A. Page, P. Helmke, R. Loeppert, P. N. Soltanpour, M. A. Tabatabai, C. T. Johnston, & M. E. Sumner (Eds.), Methods of Soil Analysis: Chemical Methods (pp. 1085–1121, Vol. 3). John Wiley & Sons, Ltd.

Brus, D.J. (2022). Spatial sampling with R. https://doi.org/10.1201/9781003258940

Brus, D.J., de Gruijter, J.J. (1997). Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with discussion). Geoderma 80: 1-44

Clarke Topp, G. and Ferré, P.A. (2002). 3.1 Water Content. In Methods of Soil Analysis (eds J.H. Dane and G. Clarke Topp). https://doi.org/10.2136/sssabookser5.4.c19

de Gruijter, J. J., D. J. Brus, D.J., M. F. P. Bierkens, M.F.P., Knotters, M. (2006). Sampling for Natural Resource Monitoring. Berlin: Springer.

Dick, R.P. (2011). Methods of Soil Enzymology. Chapter 5

Dor, E.B., Ong, C., Lau, I. C. (2015). Reflectance measurements of soils in the laboratory: Standards and protocols. Geoderma 245: 112-124.

Fernández-Ugalde, O., Orgiazzi, A., Jones, A., Lugato, E., Panagos, P. (2017). LUCAS 2018 – SOIL COMPONENT: Sampling Instructions for Surveyors, EUR 28501 EN, doi 10.2760/023673.

Fernández-Ugalde, O., Jones, A., Meuli, R.G. (2020). Comparison of sampling with a spade and gouge auger for topsoil monitoring at the continental scale. European Journal of Soil Science 71:137-150.

Flint, A.L., and Flint, L.E. (2002). 2.2 Particle density. In Methods of Soil Analysis: Part 4 Physical Methods, 5.4 (eds J.H. Dane and G. Clarke Topp). https://doi.org/10.2136/sssabookser5.4.c10

Grafström, A., Tillé, Y. (2013). Doubly Spatial Sampling with Spreading and Restitution of Auxiliary Totals. Environmentrics 24: 120-131. http://dx.doi.org/10.1002/env.2194

Johannes A., Weisskopf P., Schulin R., and Boivin P. (2017). To what extent do physical measurements match with visual evaluation of soil structure? Soil and Tillage Research 173: 24-32.

Johannes A., Weisskopf P., Boivin P., Gondret K., Leopizzi P., Lamy F., Füllemann F., Boizard H., Baize D., Ball B., Cloy J., Munkholm L., Guimarães R. (2020). VESS2020 Visual Evaluation of Soil Structure.

Lassabatère, L., Angulo-Jaramillo, R., Ugalde, J. M. S., Cuenca, R., Braud, I., Haverkamp, R. (2006). Beerkan estimation of soil transfer parameters through infiltration experiments: BEST. Soil Science Society of America Journal 70: 521–532.

Lavelle et al. (2022) Soil macroinvertebrate communities: A world-wide assessment. Global Ecology and Biogeography 31: 1261-1276. https://doi.org/10.1111/geb.13492

Matthieu, J., Lavelle, P., Brown, G. (2024). Reportingdatatothe #GlobalSoilMacrofaunadatabasev.2.0. https://zenodo.org/records/10479451

Mathieu et al. (2022). sOilFauna - a global synthesis effort on the drivers of soil macrofauna communities and functioning. Soil Organisms 94: 111-126. https://doi.org/10.25674/so94iss2id282

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

Minasny, B., McBratney, A.B. (2006). A conditioned Latin hypercube method for sampling in the presence of ancillary information. Computers & Geosciences 32: 1378-1388.

Nelson, D. W., and Sommers, L. E. (1996). In D. Sparks, A. Page, P. Helmke, R. Loeppert, P. N. Soltanpour, M. A. Tabatabai, C. T. Johnston, & M. E. Sumner (Eds.), Methods of Soil Analysis: Chemical Methods (pp. 961–1010, Vol. 3). John Wiley & Sons, Ltd.

Pimstein, A., Notesco, G., Ben-Dor, E. (2011). Performance of three identical spectrometers in retrieving soil reflectance under laboratory conditions. Soil Science Society of America Journal 75(2): 746-759.

Petersen, S. O., Hoffmann, C. C., Schäfer, C.-M., Blicher-Mathiesen, G., Elsgaard, L., Kristensen, K., Larsen, S. E., Torp, S. B., and Greve, M. H. (2012). Annual emissions of CH4 and N2O, and ecosystem respiration, from eight organic soils in Western Denmark managed by agriculture, Biogeosciences 9: 403–422. https://doi.org/10.5194/bg-9-403-2012

Phillips et al. (2019). Global distribution of earthworm diversity. Science 366: 480–485.

Roy, J.L. and McGill, W.B. (2002). Assessing soil water repellency using the molarity of ethanol droplet (MED) test. Soil Science 167: 83–97. https://doi.org/10.1097/00010694-200202000-00001.

Schjønning, P. (1984). A laboratory method for determination of gas diffusion in soil. No. S 1773, Statens Planteavlsforsøg (Denmark), Beretning.

Schjønning, P. and Koppelgaard, M. (2017). The Forchheimer approach for soil air permeability measurement. Soil Science Society of America Journal 81: 1045-1053

SEAR: Soil enzymatic activity reader. https://www.digit-soil.com/

Thomas, G. W. (1996). In D. Sparks, A. Page, P. Helmke, R. Loeppert, P. N. Soltanpour, M. A. Tabatabai, C. T. Johnston, & M. E. Sumner (Eds.), Methods of Soil Analysis: Chemical Methods (pp. 475–490, Vol. 3). John Wiley & Sons, Ltd.

Tian, X., Consoli, D., Leandro, P., Ho, Y., Hengl, T. (2024). Landsat-based soil spectral indices for pan-EU 2000-2022: Long-term trend (2000-2022). https://zenodo.org/records/10776892.

Walthert, L., Lüscher, P., Luster, J., Peter, B., 2002: Langfristige Waldökosystem-Forschung LWF. Kernprojekt Bodenmatrix. Aufnahmeanleitung zur ersten Erhebung 1994–1999. Birmensdorf, Eidgenössische Forschungsanstalt WSL. 56 S. + Anhang

Woese C.R. (1987). Bacterial evolution. Microbiological Reviews, 51 : 221–271. https://doi.org/10.1128/mr.51.2.221-271.1987

Table x = will be determined, o = may be determined

	Basic indicators								Novel indicators			
Pilot	Texture	SOC	pH	EC	P	TN	CEC	BD	NIR	eDNA	Enzymes	Macrofauna
HR												
DK												
FI	x	x	x	x	x	x	x	x	o	o	o	o
FR												
DE												
EL												
NL	x	x	x	x	x	x	x	x	o	o	o	o
ES	x	x	x	x	x	x	x	x	o	x	x	x
SE												
UK												
CH	x	x	x		x	x	x	x			x	