CODEX PLANETARIUS

The Practical Considerations of Soil Health Measurement for *Codex Planetarius*

Ecometric WWF *Codex Planetarius* Paper

V. Rojas Moreno, G. Milner, and H. F. J. Evans

About *Codex Planetarius*

Codex Planetarius is a proposed system of minimum environmental performance standards for producing globally traded food. It is modeled on the *Codex Alimentarius*, a set of minimum mandatory health and safety standards for globally traded food. The goal of *Codex Planetarius* is to measure and manage the key environmental impacts of food production, acknowledging that while some resources may be renewable, they may be consumed at a faster rate than the planet can renew them.

The global production of food has had the largest impact of any human activity on the planet. Continuing increases in population and per capita income, accompanied by dietary shifts, are putting even more pressure on the planet and its ability to regenerate renewable resources. We need to reduce food production's key impacts.

The impacts of food production are not spread evenly among producers. Data across commodities suggest that the bottom 10-20% of producers account for 60-80% of the impacts associated globally with producing any commodity, even though they produce only 5-10% of the product. We need to focus on the bottom.

Once approved, *Codex Planetarius* will provide governments and trade authorities with a baseline for environmental performance in the global trade of food and soft commodities. It won't replace what governments already do. Rather, it will help build consensus about key impacts, how to measure them, and what minimum acceptable performance should be for global trade. We need a common escalator of continuous improvement.

These papers are part of a multiyear proof of concept to answer questions and explore issues, launch an informed discussion, and help create a pathway to assess the overall viability of *Codex Planetarius*. We believe *Codex Planetarius* would improve food production and reduce its environmental impact on the planet.

This proof-of-concept research and analysis is funded by the Gordon and Betty Moore Foundation and led by World Wildlife Fund in collaboration with a number of global organizations and experts.

For more information, visit www.codexplanetarius.org

The Practical Considerations of Soil Health Measurement for *Codex Planetarius*

Ecometric WWF *Codex Planetarius* Paper

V. Rojas Moreno[1], G. Milner[2], and H. F. J. Evans

[1] Corresponding author; [2] corresponding author

Abstract

Understanding soil health is crucial for agricultural productivity and ecosystem services, yet its complex physical, chemical, and biological interactions pose challenges for monitoring. Accurate soil health assessment traditionally relies on soil sampling, which provides precise data but is resource intensive. Sampling approaches must balance depth, density, and timing with logistical and financial considerations, while associated laboratory analysis requires robust standards and quality assurance.

This paper presents the practical considerations of soil health monitoring at large scales based on global experience with Soil Organic Carbon (SOC), the learnings from which can be applied to other *Codex* metrics such as Soil Organic Matter (SOM). These are key indicators for soil health, reflecting key functions such as nutrient cycling, water retention, structural stability, and microbial diversity. This paper evaluates the use of remote sensing and machine learning approaches to complement targeted soil sampling for effective and accurate soil health measurement. The strengths and limitations of these approaches are discussed, taking into consideration all relevant practicalities when implementing these approaches in the field. The paper concludes with recommendations for a global monitoring approach that integrates targeted soil sampling, remote sensing, and machine learning. This adaptive approach prioritizes high-risk areas, improves efficiency, and delivers scalable solutions for tracking soil health dynamics across diverse landscapes.

1 Soil Health Metrics

1.1 Soil Health

Soil health refers to the capacity of soil to function as a living ecosystem, sustain biological productivity, and support environmental quality. A soil is considered healthy when it consistently delivers ecosystem services such as nutrient cycling, water regulation, pollutant buffering, and carbon storage.

1.2 Key Indicators for Soil Health

Soil systems are highly complex due to the interaction of multiple dynamic physical, chemical, and biological attributes. Monitoring all these attributes individually to assess overall soil health can obscure overarching trends, complicate decision making, and often proves too costly in time and resources to practically implement. Key indicators simplify this process by providing an efficient method of measuring overall soil health through a selection of chosen attributes which reflect multiple key soil functions. This method is cost effective and supports easier stakeholder involvement by reducing complex measurement and analysis.

Soil Organic Matter (SOM) is a valuable metric for understanding nutrient cycling, microbial activity, and soil fertility. It is most commonly measured through loss-on-ignition (LOI) laboratory testing. In many regions, particularly sub-Saharan Africa, South Asia, and Southeast Asia, LOI remains the default laboratory method for soil testing, meaning that SOM is often the only parameter monitored. While SOM estimates provide useful insights, they are highly variable and method dependent, including kiln temperature differentials that can overestimate SOM by up to 200%. LOI results can also be influenced by factors beyond carbon content, including water loss, volatile minerals, and other organic compounds. They also tend to emphasize more labile fractions which can exaggerate short-term changes, and are sensitive to soil type, particularly in soils with high moisture retention. Despite these limitations, SOM monitoring can still help track improvements in nutrient retention, water-holding capacity, and microbial activity, provided results are interpreted with care.

In comparison, Soil Organic Carbon (SOC) is typically a more consistent indicator of soil health. High SOC levels are linked to better water retention and infiltration (Masood & Ali, 2023), and have been shown to improve soil structure by binding soil particles together into stable aggregates, reducing the risk of erosion and compaction (M. Zhou et al., 2020). Carbon-rich soils also support diverse communities of bacteria, fungi, and other organisms which drive nutrient cycling and availability (J. Zhou et al., 2023). Importantly, SOC can be measured more consistently and precisely in laboratories than SOM through methods such as temperature-ramping Dry Combustion, with sampling plans designed to ensure affordability. It also changes more gradually than SOM and is less affected by short-term biological activity, making it better suited for long-term monitoring and management. Due to its wide relevance and ease of measurement, SOC is often used as a key indicator for assessing soil health (Rao, 2019). Throughout this paper, learning from experience with SOC, particularly in terms of sampling, testing and measurement can, in most cases, be applied to SOM.

2 Factors Affecting Soil Health

The following section details key factors that influence soil health, with a specific focus on their impact on SOC levels.

2.1 Temporal

SOC levels naturally fluctuate over time even when soil management and environmental conditions remain stable due to factors such as moisture fluctuation and seasonal biological activity. These temporal changes, particularly on a month-to-month basis, can be significant enough to obscure underlying trends in carbon accumulation or loss. However, these fluctuations do not always follow predictable seasonal patterns, complicating efforts to account for them in routine monitoring.

As a result, single-time-point measurements may not accurately reflect long-term soil health or carbon status. It is recommended to conduct measurements at regular intervals to reduce uncertainty and improve confidence in SOC assessments, rather than relying on a snapshot approach (Wuest & Durfee, 2024). Long intervals such as five years or more are often insufficient for timely decision-making and may fail to detect degradation before it becomes severe. Annual sampling provides a practical frequency that allows the impacts of management changes to be tracked whilst smoothing out short-term variability. This is important when evaluating the effects of soil management or carbon sequestration initiatives.

2.2 Slope

Slope has a strong influence on soil health, primarily through its impact on water movement and erosion which directly affects SOC distribution. On steeper gradients, rainfall runoff accelerates erosion and the loss of carbon-rich topsoil, thereby degrading the soil structure. This reduces the soil's ability to retain nutrients and water, weakening its fertility and biological function (Nozari & Borůvka, 2023).

Vegetation cover can help reduce these losses. Grass cover, for example, slows down surface runoff, protects the soil surface, and helps maintain SOC (J. Li et al., 2024). Tillage can also affect the rate of soil erosion. Conservation tillage methods of zero or minimum tillage in conjunction with vegetation cover has been shown to significantly reduce soil loss on steep slopes compared to conventional cultivation practices (Salim et al., 2017). A combination of slope management and ground cover is essential for protecting soil carbon and maintaining healthy soils in sloped landscapes (De Baets et al., 2011; Vannoppen et al., 2015).

Gentler slopes or lower slope positions often act as deposition zones. These areas accumulate eroded material from higher elevations including carbon-rich particles, leading to higher SOC levels. However, flat, poorly drained areas can be susceptible to waterlogging. Waterlogged soil limits oxygen availability, impairs soil structure, reduces microbial diversity, and can also promote disease (Chhabra et al., 2025). As such, careful management decisions are required across sloped landscapes to reduce rainfall runoff, erosion, and waterlogging to promote healthy soils.

2.3 Rainfall

Rainfall is a key driver of soil health, directly influencing the processes that control SOC dynamics. The timing, intensity, and volume of rainfall impacts microbial activity, moisture availability, erosion processes, and nutrient dynamics. These impacts are particularly evident in regions with seasonal rainfall patterns or sloped and vulnerable landscapes.

Soil microbial communities respond quickly to changes in rainfall timing and frequency. For example, rainfall following extended dry periods can trigger a spike in microbial activity due to sudden nutrient availability, even though the broader microbial community structure remains stable. This was observed in Mediterranean-type grasslands in Northern California (Cruz-Martínez et al., 2012).

Intense rainfall events accelerate SOC and nutrient loss, particularly on sloped farmland. Research in karst regions of southwest China found that surface runoff is the primary pathway for nutrient loss, for example nitrogen, phosphorus, and potassium, during intense rainfall. This nutrient loss during intense rainfall events can be mitigated through strategies such as enhancing vegetation cover and reducing water-soil contact time (Yao et al., 2021). This highlights the importance of appropriate management practices and vegetation cover to protect SOC levels and overall soil health during intense rainfall events.

2.4 Soil Moisture

Soil moisture plays a vital role in regulating soil health, directly impacting the biological and chemical processes that govern SOC dynamics. Water availability within soil supports nutrient retention and microbial activity, influencing microbial decomposition processes and organic matter stabilization. Moisture levels also affect soil pH, nutrient concentrations, and microbial biomass dynamics. Vegetated soils retain significantly more moisture than bare soils, reducing evaporation and promoting microbial activity. In arid regions, such as the Nabid Desert, these effects were found to be more pronounced in the surface layers (B. Li et al., 2016).

Optimal moisture levels create a positive feedback loop by supporting microbial processes which build SOC, whilst high SOC levels enhance the soil's water-holding capacity, thereby supporting healthier, more resilient soils. Monitoring soil moisture can therefore support the measurement of SOC levels and wider soil health.

Modern remote sensing technologies such as Radio Detecting and Ranging (RADAR), specifically Synthetic Aperture Radar (SAR) data provided by the European Space Agency's Sentinel-1 satellite mission (Balenzano, Mattia, Satalino, Lovergine, Palmisano, & Davidson, 2021; Balenzano, Mattia, Satalino, Lovergine, Palmisano, Peng, et al., 2021) allows researchers to assess soil moisture conditions even across large agricultural landscapes. Integrating this data with physically based models and artificial neural networks has improved the precision of soil moisture estimates, helping detect subtle changes with relatively low error and provide a greater understanding of how soils function over time (Mirsoleimani et al., 2019). Due to the direct interaction between SOC and soil moisture, such techniques and data on large-scale soil moisture patterns can support improved SOC monitoring. Soil moisture is, however, too dynamic in nature to be used as a standalone indicator but when used in context to other covariate factors is a valuable part of a soil health toolkit.

2.5 Temperature

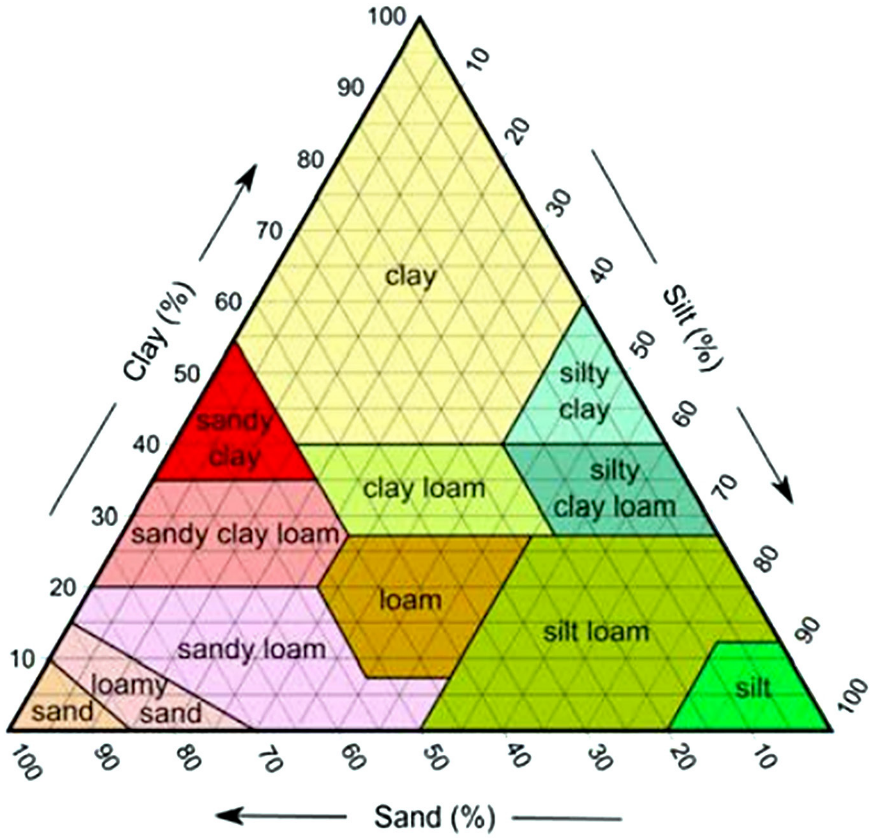
Soil temperature plays a crucial role in regulating core soil processes such as microbial activity, which directly impacts the rate of organic matter decomposition and nutrient cycling. Microbial and macro-organism activity is generally optimal between 10°C and 30°C, with micro-organism activity and survival declining sharply in sub-zero or high (> 50°C) temperature extremes (Davidson & Janssens, 2006; Kirschbaum, 1995). Since the rate of microbial activity and decomposition drives SOC levels, temperature strongly influences carbon sequestration potential and wider soil health.

High soil temperatures also reduce moisture retention and increase evaporation, leading to reduced water availability, poorer soil aeration, and impaired plant growth. Additionally, temperature influences soil chemistry. For example, increased decomposition at higher temperatures can lead to greater acid production, indirectly lowering soil pH and impacting soil fertility (Onwuka, 2018).

2.6 Soil Type

Soil type, defined by texture, structure, and inherent characteristics like clay or stone content, sets the foundational potential for soil health and productivity. These traits, shaped by parent material, landscape position, and climate, influence critical functions such as water retention, nutrient storage, root development, and SOC storage. While properties like pH or organic matter can be managed over time, soil texture is largely fixed and determines the soil's long-term potential (H. Li et al., 2020). Figure 1 (next page) presents an overview of soil types, characterized by relative percentages of clay, silt, and sand.

Figure 1: Soil Type Diagram (British Society of Soil, 2023)



Stable soil aggregates are essential for water infiltration, soil aeration, and resistance to degradation. Their formation depends on the interaction between SOC, clay minerals, metal oxides, and biological agents such as microbial by-products and root exudates, and is largely influenced by underlying soil type. Sandy soils often have poor aggregation and nutrient retention leading to low SOC levels, while clay-rich soils are more able to stabilize SOC and retain soil moisture, reducing risks of pollution, and improving resilience (Kibblewhite et al., 2008).

In practice, management decisions can influence these conditions but cannot change the fundamental soil type or texture. Understanding soil type is therefore important to identify realistic targets for SOC improvement. For example, soil types with a high sand percentage are typically limited to lower SOC values due to large particle size, low surface area, and poor aggregation, whereas soil types with a high clay percentage typically provide more favourable conditions for SOC accumulation due to fine particle size and high surface area allowing strong binding of OC to minerals and good aggregation.

2.7 Biome

Biomes such as forests, grasslands, deserts, and tundra shape soil health through distinct combinations of climate, vegetation, and microbial processes. Each biome develops soils with unique physical and biological traits, driven by long-term ecological conditions (X. Zhao et al., 2019).

The conversion of natural ecosystems to agriculture consistently alters soil microbial communities across all major biomes, including tropical, temperate, and arid systems. These biological changes often occur before any noticeable shifts in soil chemistry or structure, making microbial shifts a potential early warning signal of soil health degradation (Trivedi et al., 2016).

SOC plays a key role in regulating the balance between microbial diversity and biomass across biomes. In arid, carbon-poor soils, microbial communities tend to have high diversity but low biomass, whereas colder, carbon-rich soils support greater microbial biomass but lower diversity. As SOC declines due to land use change or climate pressures, these microbial dynamics and the ecosystem function they support can be significantly disrupted (Bastida et al., 2021). Overall, SOC provides a useful indicator for comparing soil health across biomes.

2.8 Vegetation

Vegetation is a primary driver of soil health and carbon sequestration. Plant cover helps to protect the soil surface from wind and water erosion, while root systems create channels in the soil which improve aeration and water infiltration. Vegetation is also responsible for supplying the majority carbon to the soil through decaying biomass and root exudates which feed soil microbes (Jones, n.d.).

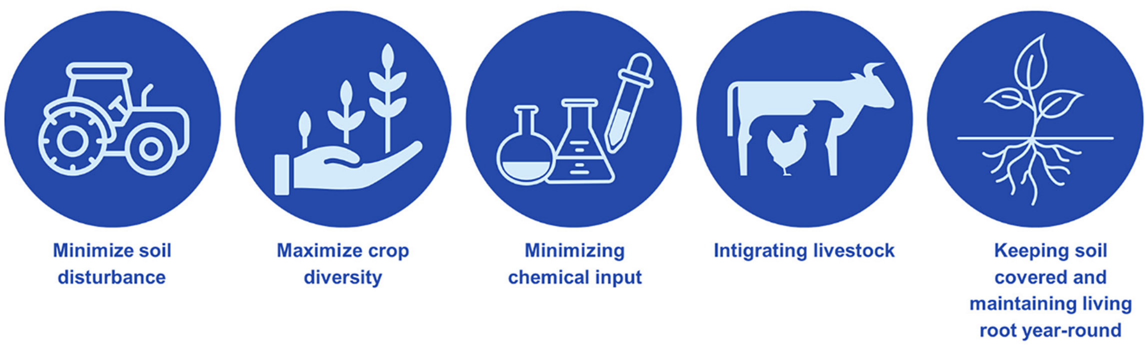
The diversity of vegetation is important as mixed-species systems such as multi-crop fields generally accumulate more SOC and have healthier microbial ecosystems than monocultures. This is due to different plant species typically having different root architectures and supporting distinct microbial communities (Jones, n.d.).

2.9 Farming Practices

Farming practices directly impact SOC and overall soil health. Conventional intensive practices such as repeated tillage, monocropping, and heavy fertilizer use generally reduce SOC, disrupt microbial communities, and accelerate erosion. Over time, this undermines soil fertility and resilience, leading to higher input dependence.

Conversely, regenerative agriculture practices such as cover cropping, crop rotation, and avoiding soil disturbance (no-till farming) can increase the diversity of soil microbes and boost organic matter while reducing the need for chemical inputs and the susceptibility of soil to erosion. Figure 2 (below) presents the main pillars of regenerative agricultural practices. Keeping the ground covered with living plants year-round helps maintain a steady flow of carbon from plant roots to the soil, feeding microbial communities that create soil structure, store nutrients, and improve water retention (Jones, n.d.). By allowing microbes, especially nitrogen-fixing bacteria and mycorrhizal fungi, to function naturally, soils can begin to regenerate their own fertility. This leads to healthier and more resilient soils over time (Laganière et al., 2022).

Figure 2: Regenerative Practices (Ecometric, 2025)



A meta-analysis across global croplands showed significant increases in topsoil (0-20 cm) carbon levels with the implementation of regenerative agriculture practices whereby SOC rose by approximately 11-12%. These benefits were most pronounced when multiple practices were combined, particularly when no-till was paired with livestock integration (Prairie et al., 2023). This demonstrates the importance of farming practices and management decisions and overall soil health.

It is important to note that the extent of soil carbon sequestration depends both on the chosen management approach, and the effectiveness of its implementation. For example, successful cover cropping helps to keep soil protected year-round, preventing SOC loss through erosion. Conversely, if the cover crops fail to properly establish, soils remain as exposed to measurable SOC decline as they would be in the absence of the practice. Regenerative Agriculture practice models risk false assumption here in allocating a guaranteed performance factor to each management practice while taking no account of the actual effectiveness of the practice. This highlights the importance of tracking measurement-based performance metrics to monitor actual rather than assumed outcomes, especially in the context of increasingly frequent weather extremes associated with climate change that exceed historic model parameters.

2.9.1 Case Studies

Increased SOC levels associated with regenerative agriculture practices are evident not only in academic research but also in field practice. Two examples at either end of the performance scale are shown below from the Ecometric 2024 UK SOCS change results from two monitoring round spanning the 2024 cropping season. Both projects experienced +250% of the 20-year average rainfall during this cropping year, dramatically increasing the risk of erosional SOC loss from bare overwinter soils common in the conventional production system practiced by Farm 1 and highlighting the vital role of regenerative practices including minimized cultivation and cover cropping in mitigating this risk as practiced by Farm 2:

- Farm 1: Loss -28 (t/ha)
  - No regenerative practices
  - Clay Loam soil
  - No cover cropping
  - Very little catch cropping
  - No applications of Organic Nitrogen
- Farm 2: Gain +7 (t/ha)
  - Regenerative practices
  - Clay Loam soil
  - Cover cropping
  - Applications of ON

2.10 Summary

Soil health is governed by a complex interplay of biological, chemical, and physical factors. For practical management and monitoring, a single indicator metric is essential as it provides a clear, cost-effective, and reliable way to track changes over time. SOC is the most appropriate choice for this role because it reflects the core functions of the soil. It directly impacts soil structure, nutrient cycling, water holding, and biological activity, and it is measurably influenced by land management and environmental factors.

3 In-Situ Soil Health Monitoring

3.1 Soil Sampling Practicalities

Soil sampling is the basis for assessing SOC as well as other soil health metrics such as SOM and nutrient availability. It involves the systematic collection of representative soil material from designated field locations for laboratory analysis. Accurate sampling is essential to ensure that measured SOC values are representative of the field conditions and robust enough to inform management decisions.

However, sampling is resource intensive and subject to logistical, financial, and methodological constraints. Key challenges include cost, accessibility of sampling sites, consistency of methodology, and the inherent uncertainty associated with subsampling a highly heterogeneous medium such as soil. Other real-world challenges include different spatial and depth sampling requirements for different metrics, for example nutrient sampling is conducted to a depth of 15cm while SOC and SOM sampling is conducted to a minimum depth of 30cm and up to 90cm, often denying the potential economies of scale from sampling for multiple metrics at the same time.

3.1.1 Sampling Cost

Soil sampling teams require specialized equipment, including coring tools, transport containers, appropriate vehicles to move between sampling positions, and GNSS devices with sufficient positional accuracy (typically within 10 meters) to record sampling locations and follow pre-determined plans in GIS software. This requirement for specialized equipment to ensure samples are accurate and appropriately tracked has the disadvantage of higher capital and operating costs.

Mechanized sampling systems tend to be readily available in developed agricultural regions due to their use in routine soil nutrient analysis, but in developing agricultural regions hand sampling using augers or pot-corers often remains the only option. Sampling by hand is lower cost as an important inclusivity protection but necessitates training and quality control measures to ensure consistency.

Sampling teams typically price by the total area covered rather than the number of samples collected. As a result, reducing the number of samples does not necessarily reduce costs. Reducing the sampling density does however increase the uncertainty of the measurement of SOC stock. Sampling design must therefore consider trade-offs between cost and spatial coverage, considering the increased uncertainty of reduced sampling density against the cost saving it represents.

Sampling costs vary considerably by country and region due to the availability of equipment and suitably trained personnel, and the operator business models. Across Ecometric projects as a current and real-world example, sampling costs in 2025 ranged between <$1/ha in South America to >$15/ha in the UK.

3.1.2 Reducing Cost

Soil sampling often represents the largest element of the total SOC monitoring cost, priced by area covered or time taken, depending on the number of samples and the distance between the samples (density). Sampling variable costs include labor, fuel, travel and subsistence, and sample shipment logistics, with savings in some or all elements needed to achieve a reduction in cost.

Reducing the number of samples may not reduce the cost if the geographic locations are as widely dispersed as the original high density sample plan as the total distance travelled remains the same. Clustering the samples in a smaller area may reduce the sampling cost but may also fail to capture the entire SOC value range in the project area, reducing the representativeness and accuracy of the results. Cost reductions must hence be carefully balanced against accuracy requirements to avoid false economy.

3.1.3 Sample Volume

In general, there are two main approaches to soil sampling: the analysis of a single intact core, or the combination of multiple cores across a sampling area into a conflated sample. Intact core analysis can offer more precise spatial attribution and maintain the soil’s vertical structure which allows the analysis of soil horizons or root penetration, however they can be biased by local heterogeneity and so are limited when representing the average condition of an area. Collecting larger diameter intact cores also requires more specialized, heavier, hydraulic vehicle mounted corers that are less readily available and more costly adding a financial barrier to use.

In comparison, conflated samples involve the physical averaging of multiple samples across the sampled area. This process smooths out small-scale spatial variability, resulting in a more representative estimate of the value for the total area. Conflated cores can also reduce the impact of potential outliers by not relying on a single core. While it is not possible to analyze the soil’s vertical structure using this sampling technique, it is appropriate for the measurement of soil nutrients such as SOC and SOM. Conflated sampling is also generally more practical than other methods when considering transportation and storage requirements, especially in comparison to intact cores. As such, conflated samples are preferable for SOC monitoring.

The volume of soil collected for each sample depends on the dimensions of the soil probe used to take the soil cores, the depth to which the cores are taken, and the number of cores that are combined to create a conflated sample. Laboratories typically have a minimum and maximum requirement for soil volume per sample to ensure practical sample storage and accurate analysis, which must be considered when developing the sample design. For example, most soil analyzer equipment typically requires sub-sample volumes of no larger than 3cm3 so to ensure sub-sample representativeness of the whole sample, laboratories often recommend a maximum submitted weight of 750g per sample.

3.1.4 Depth

Soil depth is a critical factor in designing sampling plans because carbon levels, and other soil nutrients, vary significantly with depth. SOM usually only occupies the topmost layer of the soil and may not represent the deeper soil health. In comparison, SOC can be found much deeper in soil profiles. For accurate results, it's important to sample in consistent depth layers such as 0-10cm, 0-30cm, or even deeper if long-term carbon storage is being assessed (Franzen et al., 1998). An intermediate depth such as 0-30cm optimizes the balance between practicality and ease of sampling with representativeness, capturing both surface conditions most influenced by inputs and management practices as well as deeper soil processes. This is also a widely standardized depth in many agronomic studies, allowing easier comparability.

3.1.5 Bulk Density

Bulk density is used in conjunction with SOC% and sample depth to calculate the SOC stock (t/ha) within a field or region. Errors in measurement of SOC or bulk density are amplified by the core depth, e.g., a 0.1% error in SOC measurement, with a core depth of 30cm, will create a 3% error in Soil Organic Carbon Stock (SOCS). Changes to bulk density as a result of soil management can bias carbon estimates if only fixed depths are used. Overall, careful attention to sampling depth helps avoid misleading conclusions about SOC changes over time.

3.1.6 Timing of Sampling

The timing of sampling must align with field operations to minimize disruption. For example, sampling equipment can damage crops at late growth stages so, ideally, the sampling will take place after harvest or at early crop growth stages. Also, manure or other organic input application may inflate the SOC measurements by the additional organic carbon (OC) content of the amendment itself. With farmyard manure applied at up to 2tOC/ha and green waste compost at up to 4tOC/ha, the potential to inflate SOC measurements is significant. Each type of amendment will decompose at a different rate over time and only a fraction of the applied carbon will stabilize in the soil, however decay curves vary significantly over time by type of amendment, soil conditions, ambient and soil temperature, and moisture levels, making residual OC prediction very difficult. The ideal solution is to sample before the manure has been applied or if that is not logistically possible, to delay sampling by minimum of six months after manure application and use peer-reviewed OC decay curves to estimate and deduct residual tOC.

3.1.7 Accessibility

Practical access to fields/sample area can also limit sampling. It is important to consider locked gates, difficult terrain, and any potential obstacles to corer placement which may reduce the feasibility of intended sample plans. Personnel safety and equipment security must also be considered at the design stage of large-scale projects in austere environments.

3.2 Soil Sampling Design

In addition to the practical considerations of soil sampling outlined above, soil sampling design must also consider potential sources of uncertainty in SOC measurements.

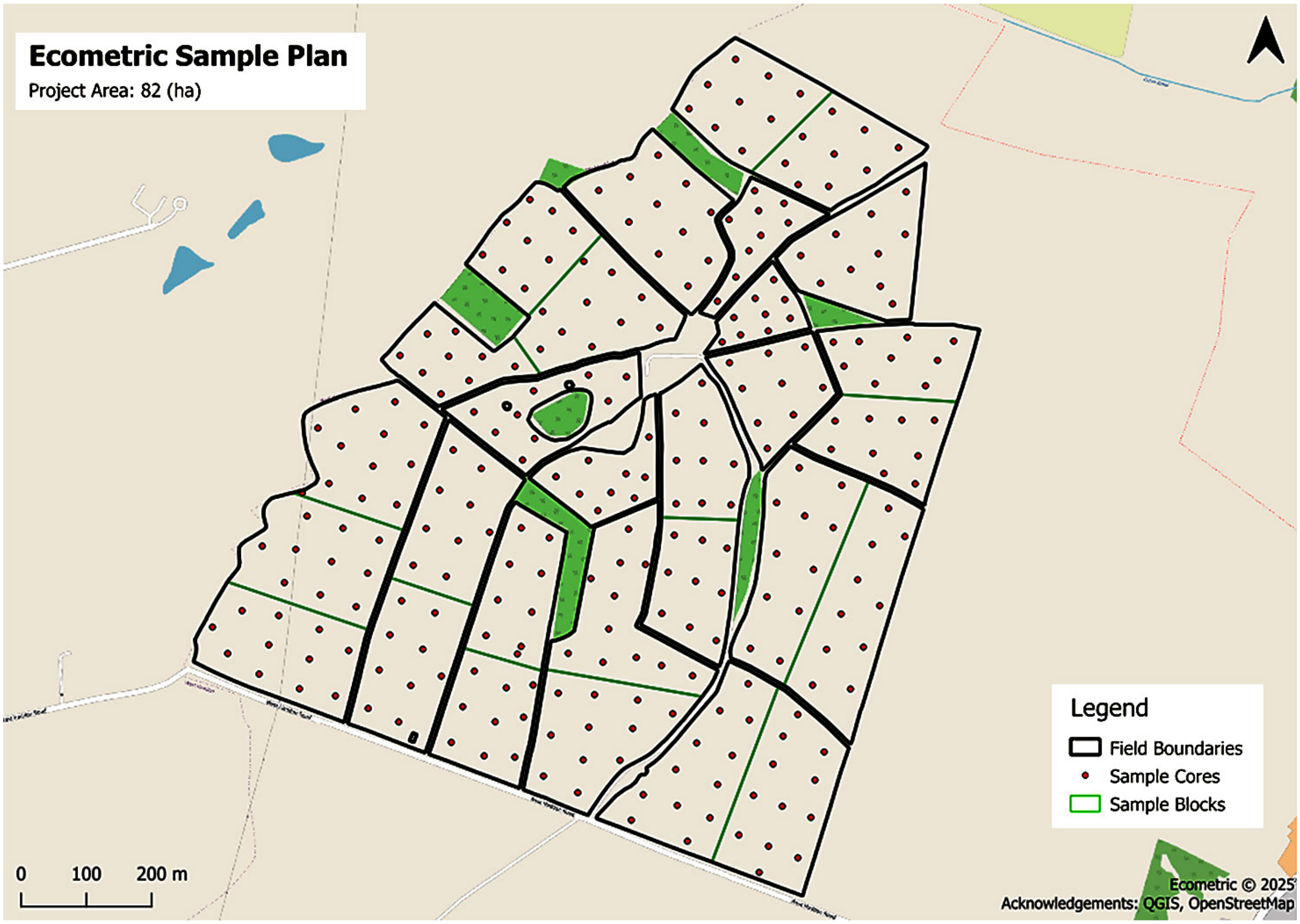
3.2.1 Uncertainty

Soil sampling statistics deal with infinite populations therefore uncertainty is used instead of accuracy. Uncertainty in SOC monitoring reflects the degree of confidence in the measurements. Sampling design plays a major role in uncertainty. In areas with complex topography, low sampling density can fail to capture key variations in the soil, increasing the uncertainty in SOC estimates. Differences in coring technique, equipment, or field conditions can also introduce variability. Understanding and quantifying these sources of uncertainty is essential for making more informed decisions about soil health and SOC (Glavič-Cindro et al., 2023).

3.2.2 Sample Density

Sample density defines the minimum number of samples required per unit area. The preferred approach to avoid bias is to divide the project area into equally sized blocks and collect multiple cores within each block, which are then conflated into a single sample, as visualized in Figure 3 (next page). The number of cores per block is typically constrained by laboratory requirements for sample weight and volume (see Section 3.1.2).

Figure 3: Example Sample Plan at High Density (Ecometric, 2025)

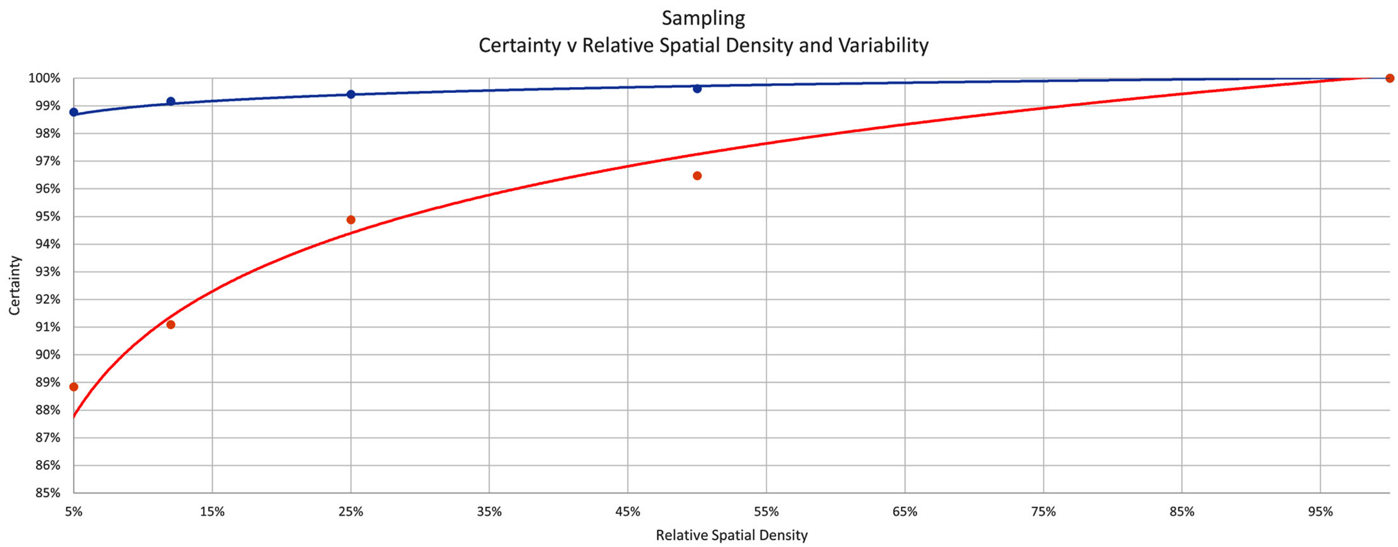


Smaller block sizes and a larger number of samples across the project area, meaning higher sample density, increases costs but improves confidence in results. In contrast, larger blocks and a smaller number of samples across the project area reduces costs but increases uncertainty. An adaptive approach to sample design is necessary to balance these trade-offs and ensure sampling meets project objectives within budgetary constraints.

The level of spatial autocorrelation in SOC values can also impact the accuracy of spatial interpolation between sample points and the confidence in results. Spatial autocorrelation refers to the correlation of a variable, in this case SOC, with itself through space, meaning the degree to which similar SOC values occur near to each other spatially. In some cases, sparse sampling can still yield accurate results in situations where spatial autocorrelation is strong, meaning areas where SOC variability through space is minimal (Long et al., 2018). This means that in some contexts, sampling density can be reduced in subsequent monitoring periods following a baseline monitoring round which demonstrates strong spatial autocorrelation in SOC values.

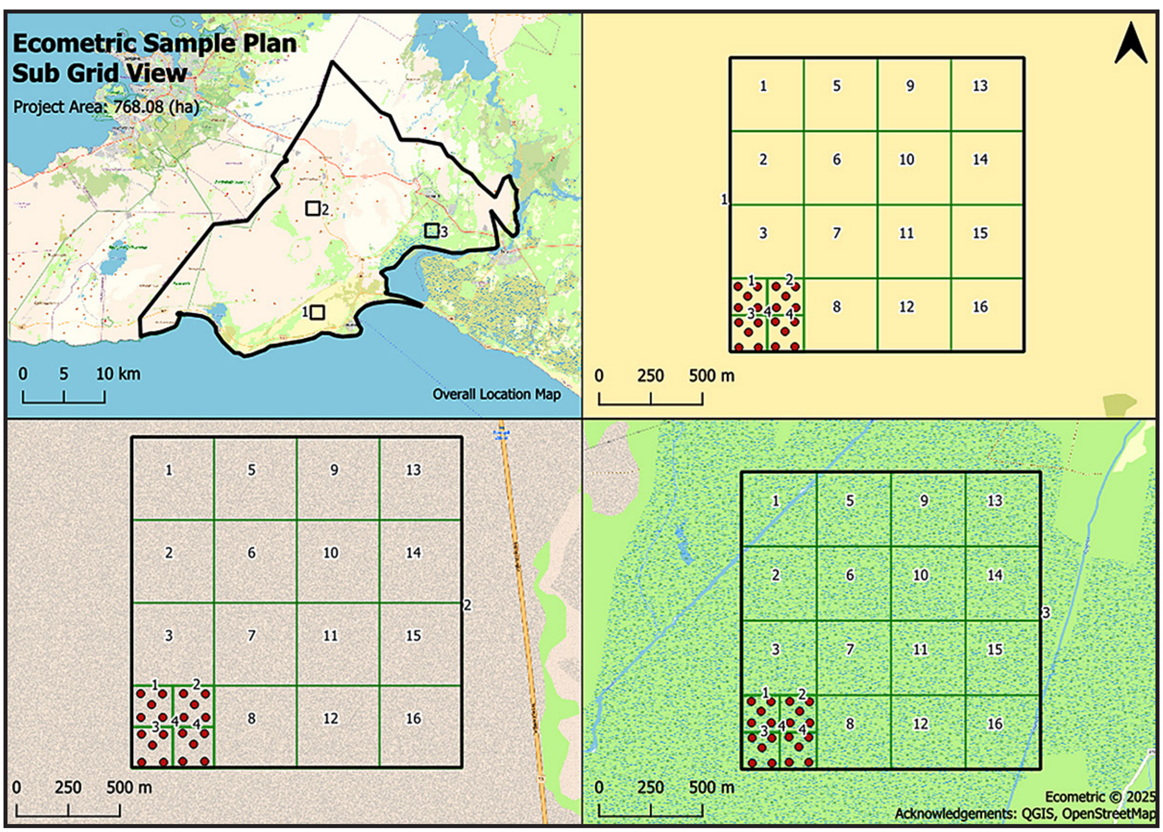
Ecometric has developed a sampling certainty simulator to calculate the rate of sampling uncertainty increase associated with reduced sampling densities, for a given site-specific SOC value range established by pilot soil sampling and laboratory testing. Certainty is used as the measure instead of accuracy because the sample population (the soil) is infinite. By simulating reduced sampling densities, the site-specific certainty curve can be graphically displayed to establish the sampling density that best balances accuracy and cost. An example curve for a low and high variability site (standard deviation of +/-5 t/ha & +/-50 t/ha) is shown in Figure 4 (next page) as certainty.

Figure 4: Low (Blue +/- 5t/ha) & High (Red +/- 50 t/ha) variability (Ecometric, 2025)



In situations where the project area is too large to practically support full sampling coverage, an alternative approach involves sampling in representative clusters or zones. Smaller areas deemed representative of the variability across the total project area are identified, for example by assessing land cover or soil type or other appropriate variables. These smaller representative areas are sampled according to a chosen sampling density, and machine learning models are trained on this representative sampled data, often incorporating remote sensing covariates, to predict into the unsampled areas. This approach enables cost-efficient SOC monitoring across large scales but requires additional considerations for potential uncertainty in the results introduced by interpolation, shown in Figure 5 (below).

Figure 5: Example Sample Plan of Representative Area at Low Density (Ecometric, 2025)



3.2.3 Digital Soil Mapping

Digital soil mapping provides a more advanced framework for optimizing sample design at large scales. This method uses environmental covariates that are known to influence soil properties such as elevation, slope, land cover, and a variety of remote sensing indices in combination with soil sample results. These covariates are integrated into quantitative models which generate predictive maps of SOC, highlighting patterns of soil variability (Zeraatpisheh et al., 2022).

Stratification can then be performed on the predictive maps, for example using a segmentation algorithm, to divide the total project area into distinct zones. Sampling can then be targeted towards zones with the highest predicted levels of SOC heterogeneity, improving efficiency while ensuring the full SOC variability is captured. Furthermore, uncertainty estimates from the digital soil mapping predictions can be produced, allowing sampling density to be increased in areas where the predictions are less reliable and reduced where confidence is high to make the most efficient use of available resources.

3.3 Laboratory Analysis

Laboratory analysis involves meeting standards, financial cost, requires quality assurance, a defined accuracy, and the complexity of transportation and storage.

3.3.1 Methods of Measuring SOC and SOM

Laboratory measurement methods have developed over time and become more accurate. The main methods that are in use for SOC and SOM analysis are:

- DUMAS (SOC)
- Walkley Black (SOC)
- Loss on Ignition (SOM)

3.3.2 Laboratory Standards

Many laboratories must hold specific accreditations or certifications to ensure that their methods meet recognized standards. Within the UK, the United Kingdom Accreditation Service (UKAS) is a common accreditation. The Agriculture and Horticulture Development Board (AHDB) lists members of the Professional Agricultural Analysis Group (PAAG). Others include United States Department of Agriculture (USDA), the National Soil Survey Lab, and Kellogg Soil Survey Laboratories (KSSL) in the United States. Adherence to these requirements is essential for ensuring the consistency and comparability of results, particularly in projects where outputs must undergo independent validation or verification. Using accredited laboratories also strengthens confidence among stakeholders that reported SOC estimates are reliable.

3.3.3 Laboratory Cost

Laboratories typically charge per sample analyzed, which creates a strong cost incentive to minimize the total number of samples submitted. This discourages practices such as repeated analysis of the same sample, despite replicate testing being preferable for the identification of outliers. Similarly, analyzing multiple sub-samples from a single field sample would provide a clearer picture of within-sample variability, but is rarely undertaken due to additional expense.

Logistical costs of transporting samples to the laboratory must also be taken into account as such costs may be considerable over large project areas, especially if the areas are remote from laboratories or if climate control is required to preserve sample quality. In regions lacking DUMAS laboratory capability, logistics costs alone may pragmatically dictate the monitoring of SOM using LOI instead of SOC, to avoid the prohibitive cost of shipping bulky soil samples long distances to reach DUMAS equipped facilities.

3.3.4 Quality Assurance

Robust quality assurance procedures are essential to ensure the accuracy and reliability of laboratory results. This includes identifying outliers within the dataset, for example where results deviate significantly from values expected within a given field or across the project area. These can be identified using methods such as the Z-Score test or other statistical approaches. Outliers may indicate issues in sampling, handling, or laboratory processing and should be reviewed carefully. Where results are considered low-confidence, retesting or additional sampling may be necessary. Consistent application of quality assurance protocols helps prevent erroneous data from biasing SOC estimates.

3.3.5 Accuracy

Accuracy in laboratory measurements is critical because even very small inaccuracies can result in amplified errors in SOCS. SOCS is calculated by multiplying SOC concentration (%) with bulk density (kg/l) and sample depth (cm). While modern laboratory equipment provides highly accurate SOC concentration measurements, bulk density remains a more manual measurement with wider error margins. As such, it is important to carefully control the bulk density measurement approach to minimize inaccuracies as much as possible. Otherwise, errors in bulk density measurement can propagate through the SOCS calculation, particularly when scaled across sampled depth and project area, leading to significant uncertainty in final SOCS estimates.

By analyzing the standard deviation and related Z-Scores of laboratory results, outliers can be identified as potentially resulting from laboratory error and sent for retesting. As a second check for errors, SOC estimations generated using machine learning techniques (see Section 6) can be compared with the laboratory results. The difference between the two can be assessed through statistical calculations such as the Absolute Percentage Error (APE), which measures the accuracy of the predicted value. APE values falling outside of the expected accuracy range can indicate potentially erroneous laboratory results, which once identified can be sent for retesting.

3.3.6 Sample Storage and Shipping

Logistics surrounding sample storage and shipment can introduce additional challenges. This is particularly evident in tropical regions, where high ambient temperatures and humidity may alter sample properties if not carefully managed. In such cases, samples must either be shipped rapidly to the laboratory or stored in climate-controlled conditions, both of which add cost and complexity, as mentioned in Section 3.3.3. As such, it is important to consider storage and transport at the planning stage.

3.4 Soil Sampling Advantages and Limitations

Estimating SOC using soil sampling alone can be advantageous in certain contexts. It is a relatively simple approach which can provide highly accurate point results if proper sampling and laboratory analysis techniques are followed. It is also an accurate method for measuring bulk density and stone fraction in soils, both of which are important to consider when monitoring SOC.

However, there are several limitations to consider. Soil sampling is expensive when considering the costs of the sampling team, equipment requirements, soil storage and transportation, and laboratory costs and can be as high as $30/ha where high SOC spatial variability demands a high sampling density. As such, the possible density of soil samples is often limited. In addition, some regions may have limited sample team or laboratory availability. Furthermore, GNSS accuracy when recording the location of sample points may limit the ability of repeat samples in the same location. Finally, human error during laboratory analysis or sampling may limit accuracy of results.

4 Proximal Sensing

Proximal soil sensing (PSS) involves using field-based instruments placed in contact with, or very close to, the soil surface (within about 2 meters). PSS applies to the measurement of SOC and SOM. These tools, ranging from optical and electromagnetic sensors to ion-selective electrodes, can rapidly collect data on soil physical, chemical, and biological properties under real field conditions. While individual readings may not match laboratory precision, the dense, on-the go measurements provide rich, practically valuable information that supports better soil mapping and management decisions.

PSS technologies can capture key soil health indicators, such as SOC, SOM, moisture, nutrient levels, pH, and texture. Both direct measurements (e.g., electrical conductivity) and indirect estimates via calibrated proxies (e.g., spectral reflectance for organic matter) can be used. These methods can be faster, cheaper, and less labor intensive than traditional sampling in some circumstances, as they avoid the time delay and cost of removing and shipping samples to a laboratory for analysis, so offer the future potential of enabling high-resolution spatial and temporal characterization of soil health.

However, from Ecometric's direct trial of two of these systems (Stenon[3] and PES Technologies[4]), they have delivered inconsistent levels of SOC quantification accuracy when directly tested against laboratory analyzed physical samples. While these capabilities are particularly valuable for precision agriculture, environmental monitoring, and soil landscape studies (Egmond et al., 2010; Gholizadeh et al., 2018; Ji et al., 2019) they have not yet demonstrated comparable accuracy to direct laboratory measurement for SOC estimation. The in-field use of these sensor types is also more time consuming per core than physical sampling, so although they offer potential cost savings by replacing laboratory testing, they are likely to increase the sampling cost element. An in-field device with comparable accuracy to the DUMAS lab test would however potentially offer cost savings of 2x - 10x despite this relative increase in sampling time and cost.

4.1 Ground Penetrating Radar

Ground penetrating radar (GPR) is an experimental sensing technique that emits high frequency electromagnetic waves into the ground and detects reflections from subsurface that have different dielectric properties. The primarily soil attributes that impact the reflected energy include moisture, bulk density, and organic matter, all key aspects of soil health. In peatland, GPR combined with electrical resistivity imaging has been shown to map peat thickness and estimate carbon (Comas et al., 2017). However, GPR cannot differentiate between organic and inorganic carbon.

4.2 Spectroscopy

Spectroscopy techniques, such as Inelastic Neutron Scattering (INS) and Visible–Near Infrared (VisNIR) reflectance, offer fast, non-destructive ways to measure SOC. INS detects gamma rays from neutron interactions in the soil, while VisNIR analyzes how soil reflects light. These methods make it possible to monitor soil components directly in the field, without needing lab analysis. Currently these in-field tools are not as accurate as soil testing in a laboratory and can also require long time periods to capture the data for each soil sample. As with GPR these methods cannot measure bulk density.

4.3 Summary

Current proximal sensing techniques are still experimental, and most are reliant on physical soil sampling to calibrate the instruments at regular intervals. Also, bulk density must be measured separately to calculate carbon stock, which requires continued laboratory analysis, negating the convenience and cost saving potential of in-field measurement.

5 Remote Sensing for Soil Health

5.1 Limitations of Using Soil Sampling Alone

To be able to accurately estimate SOCS across an entire project area, some form of interpolation is required to provide the SOCS values between the sampled points. There are several traditional methods for spatial interpolation such as Inverse Distance Weighting (IDW) and kriging. These methods rely on assumptions about how SOCS changes across space, typically expressed through mathematical functions. In practice, SOCS values between known sample points vary in a much more complex way, influenced by dynamic environmental and management factors. The Ecometric ML approach bases its interpolation and extrapolation on measured proxy data rather than mathematical functions increasing the accuracy of SOCS estimations in this dynamic natural environment. Once trained, the ML system approach allows for SOC prediction without sample data, with subsequent calibration sampling to validate the results and quantify uncertainty. This capability can be used to reduce physical soil sampling requirements and costs, while maintaining measurable accuracy and uncertainty.

Soil samples alone cannot fully capture these dynamic patterns across an entire project area. ML addresses this limitation by associating SOCS with proxy variables that can be observed at much higher density across space. This allows interpolation to reflect underlying drivers of SOC variation, rather than relying solely on distance or statistical assumptions.

[3] https://www.stenon.io/en
[4] https://www.pestechnologies.com/

Remote sensing in combination with machine learning techniques provides a practical source of such proxy data. This refers to data collected from a distance, without physical contact, including imagery from satellites or Unmanned Aerial Vehicles (UAVs).

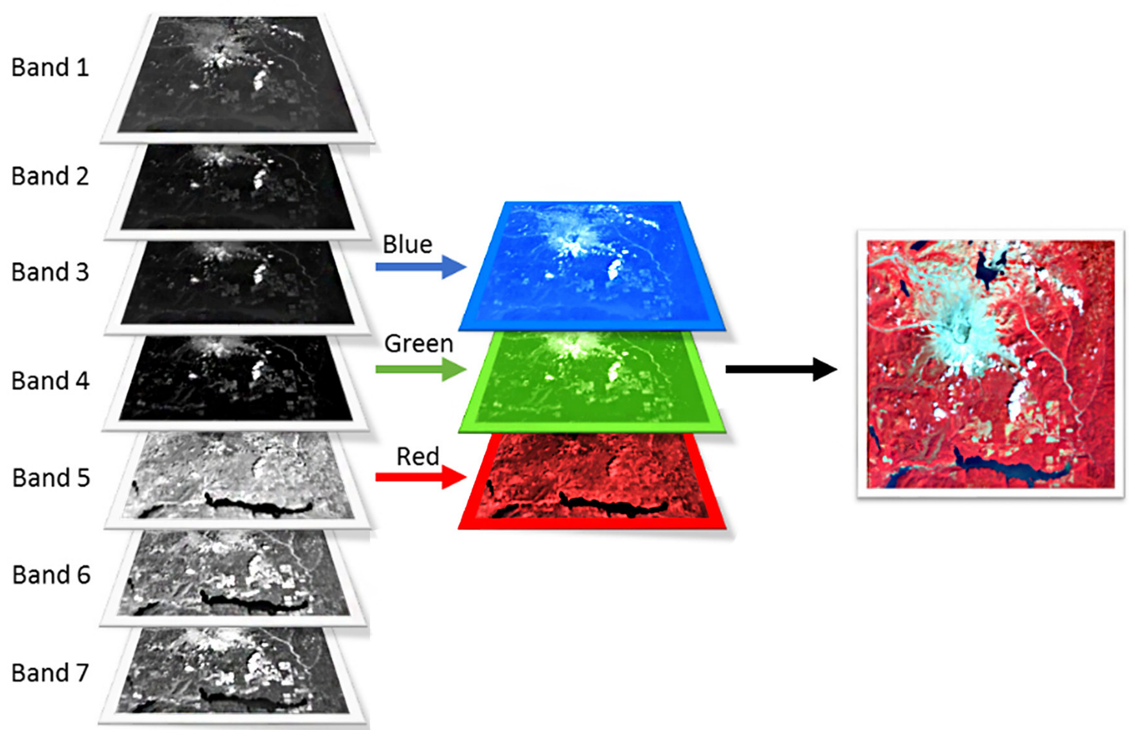
5.2 Machine Learning

Machine learning offers a flexible approach to integrating soil sampling with proxy data from remote sensing. By combining field-based SOC measurements with spatially dense satellite data, machine learning models such as artificial neural networks (ANNs) can be trained to recognise patterns and predict SOC across unsampled areas. The higher the resolution of this remote sensing data, the closer the estimation of the infinite population can be achieved (Mondal et al., 2017).

5.3 Multispectral Imagery

Multispectral imagery is a form of remote sensing data, visualized in Figure 6 (below). This imagery makes use of the patterns of reflectance, absorption, and emittance of electromagnetic radiation, which is the energy the Earth receives from the sun. Different materials such as soil, vegetation, and water reflect and absorb radiation differently across wavelengths in the electromagnetic spectrum. Distinct sections of the electromagnetic spectrum are referred to as spectral bands, and multispectral imagery captures data across a number of these bands including visible, near infra-red (NIR), and longwave infra-red (LWIR).

Figure 6: Multispectral Imagery (Humboldt State Geospatial Online, 2014)

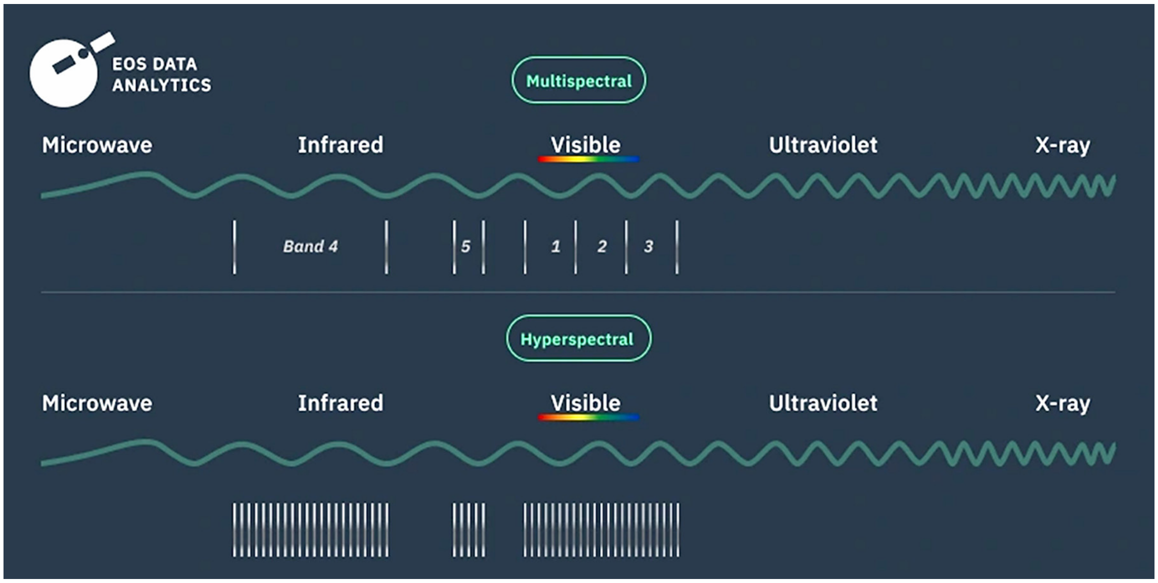


This imagery can be used to estimate key soil health indicators like vegetation cover, soil moisture, organic matter, and even some nutrient levels. For example, indices such as Normalized Difference Vegetation Index (NDVI) and Soil-Adjusted Vegetation Index (SAVI), derived from red and NIR spectral bands, can reflect underlying soil conditions by indicating vegetation density, which is often linked to soil fertility. Studies have shown that multispectral imagery combined with machine learning models can predict soil properties like organic carbon, nitrogen, and pH with useful accuracy, especially when integrated with environmental or topographic data (Piccoli et al., 2023; W. Zhao et al., 2024; Žížala et al., 2019).

Many multispectral datasets are open source. For example, the Landsat and Sentinel-2 satellite missions from the European Space Agency. These datasets offer global coverage with minimal or no cost, making multispectral imagery a practical tool for soil health monitoring over large areas.

An alternative to multispectral imagery is hyperspectral imagery which captures data in hundreds of narrow spectral bands across the electromagnetic spectrum, as shown in Figure Figure 7 (below). This imagery provides much higher spectral resolution, which is useful for detecting very subtle patterns, especially in complex environments. However, the greater spectral resolution is often a trade-off resulting in limited spatial or temporal resolution.

Figure 7: Multispectral vs Hyperspectral Imaging (EOS Data Analytics, 2025)

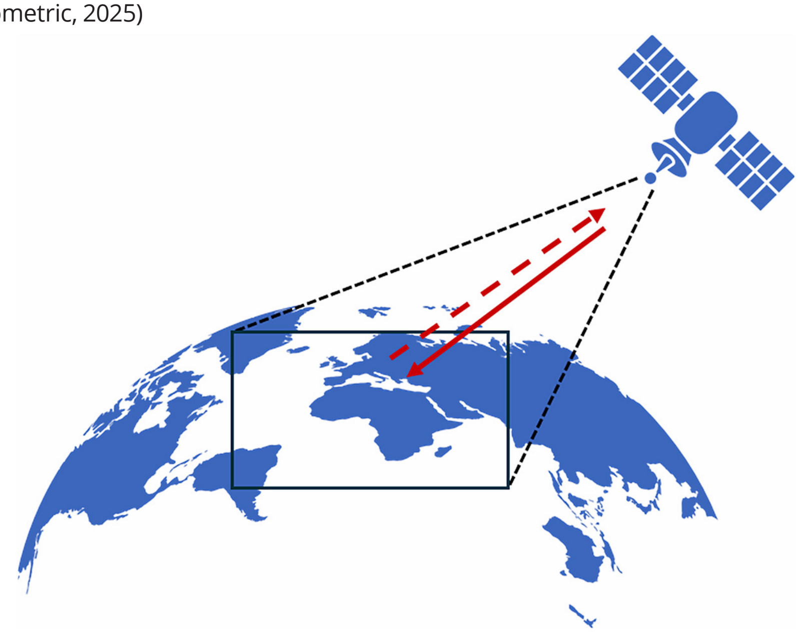


When paired with robust calibration and ML, multispectral imagery is a useful tool for assessing soil health. Its open availability (e.g., Sentinel-2, Landsat), relatively low computational costs, and integration with AI make it especially valuable for agricultural planning, carbon monitoring, and environmental conservation (Ratke et al., 2024; Žížala et al., 2019). Whilst the spectral resolution is more limited in comparison to hyperspectral imagery, multispectral imagery still provides sufficient spectral detail for a range of applications including soil health monitoring.

5.4 Active Sensors

Active sensors provide their own illumination rather than passively measuring the sun's illumination of the Earth's surface, Figure 8. Space-based sensors typically use RADAR and airborne sensors use LASER as the active transmitter.

Figure 8: Active Sensor (Ecometric, 2025)



Active remote sensing technologies such as, P-band radar, Synthetic Aperture Radar (SAR), Interferometric SAR (InSAR), and even cellular radio mast signals are valuable tools for assessing soil health, especially by detecting soil moisture, which is a key indicator of biological activity, structure, and nutrient cycling.

P-band radar penetrates vegetation and upper soil layers, making it ideal for estimating root-zone moisture and biomass. L- and C-band SAR are effective for measuring surface moisture in croplands, while InSAR provides precise changes in moisture content by tracking surface movement (LIU et al., 2019). These systems work in all weather, day or night, and cover large areas with high spatial and temporal resolution. As soil moisture links closely with SOC, these sensors could act as proxies for monitoring soil health.

The combination of data from both active sensors and multispectral sensors, in conjunction with machine learning techniques, provides a wide range of valuable information for soil health monitoring.

5.5 Temporal Frequency

Traditional soil sampling provides highly accurate measurements, but the process is limited by the time and resources required to visit each sample site. On a small scale, such as individual farms in the UK, sites could feasibly be revisited daily. However, at continental scales, the logistics and costs involved mean that re-visiting the same locations for repeated sampling may only be possible after several months or years. In contrast, remote sensing provides more efficient means of monitoring soil conditions over large areas; however, this approach is still constrained by the temporal resolution of the satellite's sensors used. The frequency with which a satellite revisits the same location varies depending on the mission, and this can limit the ability to capture short-term changes in soil properties and its environment. The temporal frequency of some example remote sensors is shown in Table 1.

Table 1: Temporal Frequency of Sensors

Sensor	Revisit Time
Sentinel 2	5 days
PlanetScope	1 day
Landsat 9	18 days

5.6 Advantages and Limitations of Remote Sensing

5.6.1 Advantages

One of the key advantages of remote sensing is the wide availability of open-source datasets from multiple satellite missions, offering a broad range of spatial, spectral, and temporal resolutions to suit diverse monitoring needs. This accessibility allows users to select data that best suit their objectives, whether for high frequency monitoring or detailed spatial analysis. Additionally, users of these open-source datasets often benefit from well-documented resources and standardized methodologies, which support accurate processing and ensure consistency in the use and interpretation of satellite imagery.

5.6.2 Limitations

Despite its many advantages, remote sensing also faces several limitations that can impact data quality and availability. Cloud cover is a significant issue in many regions, often obscuring the land surface and limiting the usability of cloud-sensitive imagery, specifically optical imagery which relies on the reflectance of solar radiation. In addition, both cloud cover and vegetation shadows can distort spectral reflectance values, reducing the accuracy of analyses derived from the imagery.

Another limitation is the revisit time of some satellite systems, which may be too infrequent to capture rapid changes or to provide consistent monitoring in dynamic environments. For example, the revisit time for Sentinel-2 imagery, as shown in Table 1, is roughly five days. This means that the satellite with the required sensor will pass over a certain location once every five days, hence there is the opportunity for new imagery at a certain location every five days. In certain situations, consistent cloud cover may correspond with this small-time window for new imagery, essentially blocking the sensors from capturing useable data. In some contexts, this can mean that there are weeks or even months between useable imagery, greatly limiting the potential to perform detailed temporal analysis. Data from satellites with a shorter revisit time, such as PlanetScope's daily imagery, offers more opportunities to collect useable, cloud-free data however this data is not open-source and costs can be significant when working over large areas.

Cloud cover limitations can be overcome in some cases by using active sensors such as SAR which is not susceptible to cloud cover interference. However, the ability to use this imagery type is dependent on the specific analysis requirements of the project because SAR data provides different insights to optical imagery. These factors must be taken into consideration when selecting remote sensing data for soil and land surface assessment.

6 Machine Learning for Soil Health Monitoring

Artificial Intelligence (AI) refers to a broad field centered around the development of machines capable of mimicking human cognitive functions. In mainstream discourse, AI is often used in relation to Large Language Models (LLMs) or similar prompt-response systems.

Machine learning is a subset of Artificial Intelligence technology which focusses on building machines that can learn from data, identify patterns, and make predictions based on these insights while improving their performance over time or with additional data. This approach can be used to interpolate SOC values at high resolution between discrete sample points using appropriate proxy data sources.

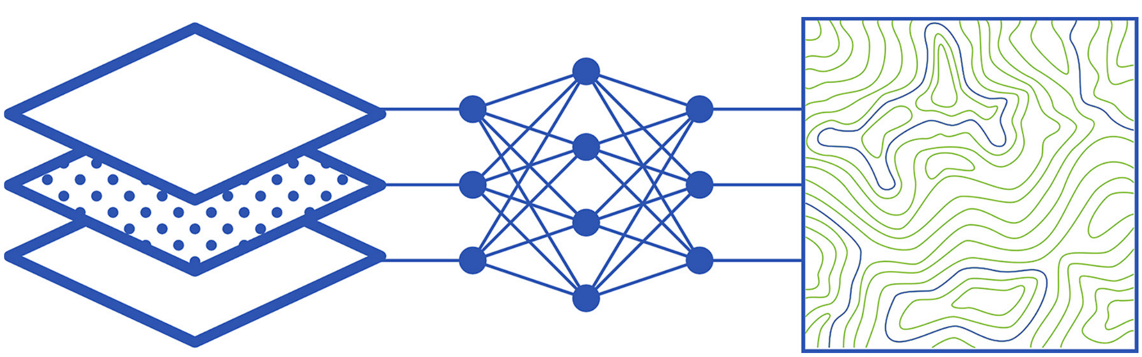
6.1 Technology Overview

Artificial Neural Networks (ANN) evolved out of a study of brain function and memory recall. ANNs have been created both in hardware and software, but generally now are implemented in software. Hardware implementations were developed to overcome the speed and memory limitations of early computers (I. Aleksander and T.J. Stoneham, 1979; Lehmann & Bruun, 1994). Neural networks associate one set of data (predictors) with elements of a second set (targets). Tasks include land use classification, facial recognition, and navigation for autonomous vehicles. The use of remotely sensed imagery in such contexts started as early as 1995 (Paola & Schowengerdt R. A., 1995).

6.2 Existing Methodologies

In the specific case of SOC monitoring using machine learning techniques, multispectral imagery is often used as a proxy. The network learns to associate multispectral data with SOC data and then predicts SOC values between sampled points, as visualised in Figure 9. This can be achieved through segmentation or regression tasks.

Figure 9: Multispectral Imagery and Machine Learning to Estimate SOC (Ecometric, 2025)



Specifically, geolocated soil sample SOCS results are combined with multispectral imagery data in the same location. This data is then used to train an ANN, which refers to the process in which the ANN learns the existing patterns and relationships between the multi-dimensional data by repeated exposure to the data and fine-tunes predictions until the error between the true and the predicted value no longer decreases, signalling that the optimal results. Error statistics can be calculated using the training and testing datasets to help quantify uncertainty in the results.

The training stage can take seconds, minutes, hours, or even days depending on the complexity of the ANN architecture, size of the training data set, and available computing power. A huge range of ANN architectures exist and have been developed for a multitude of specific use cases, for example shallow vs deep networks, supervised vs unsupervised, multiple approaches to input and output handling, number of epochs, neuron connectivity types, etc. Each type is tailored to different types of data, tasks, and computational availability.

Once the ANN is trained, it can be deployed to perform SOC predictions using only multispectral imagery. By using the relationships and patterns it learnt from the training dataset, the ANN takes in multispectral imagery data and predicts SOC values on a pixel-by-pixel basis. This gives a much higher spatial density of SOC predictions than using soil core positions alone.

In addition to multispectral imagery, other datasets representing important covariates which impact SOC can be utilized to improve the strength of the trained ANN. This can include soil type maps, climactic variables, radar data, and more. Inclusion of these covariates can strengthen the ANN's ability to generalize into unseen areas and reduce uncertainty in the predictions, with the assumption that the covariate data is of high quality and reliable.

The development of Ecometric's ANN to accurately predict SOC between sampled points and across unsampled areas required multiple training and testing cycles to optimize both the ANN architecture and training data structure to achieve the best possible performance and minimize uncertainty. The ANN had to be developed taking into consideration the specific characteristics and format of the input ground-truth data and multispectral imagery, as well as rigorously tested to ensure all potential sources of error were identified and minimized. The network continues to be developed with more input data, new data structures, more advanced machine learning approaches, and more sophisticated uncertainty calculation techniques to improve the quality of the output SOC predictions.

6.3 Data Requirements

To train an ANN and then accurately estimate SOC requires a sufficient quantity of soil samples which meet minimum accuracy standards and high-quality proxy imagery. Predicting SOC into the future is possible using a time series network design and several years of historical soil and imagery data.

6.3.1 Soil Sample Data

Historic soil sample data can be used to train an ANN; the quality of the results is dependent on multiple factors such as sampling method and density, accuracy of georeferencing, defined laboratory methods and associated accuracy, and availability of supporting variable test results such as bulk density alongside SOC which is important to quantify total SOCS. The uncertainty in historic SOCS sample results tends to be high due to outdated laboratory analysis methods and the absence of bulk density measurements alongside SOC values.

Some SOC measurement approaches rely entirely on historical data sources to train ANN and generate SOC predictions, without the use of site-specific soil sample results to either train or validate results. However, site-specific soil samples are incredibly important to ensure that the ANN can accurately predict over the area of interest. Soil characteristics are hugely variable even within areas of the same soil type or land use type.

6.3.2 Earth Observation Data

Current Earth observation data includes multispectral and hyperspectral imagery obtained from passive sensors, Synthetic Aperture Radar (SAR), Interferometric Synthetic Aperture Radar (InSAR) and P-band radar (currently in the commissions phase) from active sensors, and Light Detection and Ranging (LiDAR) data which uses laser pulses to measure distance to a very high accuracy and is useful for recording topography. This Earth observation data can be utilized in various forms for ANN training.

The limitations of earth observation data in the context of machine learning approaches are similar to the general limitations of remote sensing data, including the absence of clouds or waterlogging in multispectral imagery, potentially long revisit times, and high computational requirements of certain data types.

6.4 Advantages and Limitations of Machine Learning Approach

The use of machine learning with remotely sensed imagery offers several advantages for SOC monitoring. This approach lowers the overall cost of monitoring by reducing reliance on expensive physical soil sampling and laboratory analysis. It also provides a sophisticated method for interpolation between sampling points and across unsampled areas, generating far more detailed insights into soil health than sampling alone and mapping variability which would not be detected through mathematical interpolation methods. Furthermore, machine learning models provide a framework for quantifying uncertainties in predictions which is important when integrating the results into decision-making. Finally, in some cases SOC predictions from machine learning methods can accurately identify laboratory errors by cross-checking sample results against predicted values.

However, the use of machine learning for SOC monitoring does have important limitations. Predictions are strongly influenced by the quality and completeness of the input training datasets and the appropriateness of the model architecture for the specific use case. Input data quality is limited by soil sampling design and laboratory data quality and integrity at the time the sampling was carried out. Historic open and closed source SOC data has generally proved to be of insufficient quality to train the Ecometric ANN system, due to a lack of sample georeferencing, sampling depth inconsistencies between data sets, sample preparation inconsistencies between laboratories and the use of multiple laboratory analysis techniques that would by today's standards be considered legacy. The lack of high-quality historic data has necessitated the gathering of new baseline data to train the Ecometric ANN system, that met the latest sampling and laboratory equipment standards and through automation of the sampling and laboratory processes minimized human error. As an example, Ecometric has introduced bar coding of sample bags, which removes the potential for transcription errors by the sampling or laboratory technician teams as a simple process step that has dramatically improved data integrity.

Training can also be computationally intensive, particularly when working with large or complex datasets. Depending on the design and type of machine learning system training can take from seconds to days, while the prediction time from a trained system typically only takes seconds or minutes. These times are highly dependent on the machine learning system design and computing resource availability. This challenge is increasingly mitigated by cloud-based processing and storage availability. Such services can be relatively expensive, but costs are expected to decrease through time.

6.5 Recent Developments

Recent developments in the soil health monitoring space highlight the growing adoption of digital soil mapping techniques, with certification bodies such as Verra and Gold Standard incorporating these approaches into their frameworks. As mentioned in Section 3.2.4, these methods rely on combining input soil sample data with a variety of proxy datasets, including temperature, precipitation, crop growth dynamics, and soil moisture, to extend SOC predictions across unsampled areas. The increasing availability of space-based sensors capable of capturing these environmental variables has been instrumental in advancing these approaches and signals that the sector is moving toward greater scalability.

Nonetheless, the accuracy and robustness of such predictions remain constrained by the limited availability of project-specific ground truth data. Without sufficient local sampling, model outputs cannot be properly calibrated or validated, and uncertainties in predictions cannot be fully quantified. Our practical experience demonstrates that maintaining a strong link to project-level soil sampling is essential. Careful sample design must balance uncertainty, local conditions, project-specific SOC variability, and implementation costs to ensure that predictions achieve measurable accuracy and remain credible for both scientific and decision-making purposes.

7 Conclusions

7.1 Soil Health Indicators

Soil Organic Carbon (SOC) is often chosen as a key indicator for soil health because it is stable over time, more precisely measurable and less affected by short-term biological activity. Soil Organic Matter (SOM) remains a valuable metric for understanding nutrient cycling, microbial activity, and soil fertility alongside SOC. Higher SOC levels are linked to better water retention and infiltration. Also, high SOC levels improve soil structure by binding soil particles together into stable aggregates, reducing the risk of erosion and compaction. Carbon-rich soils also support diverse communities of bacteria, fungi, and other organisms which drive nutrient cycling and availability. These connections make SOC a useful and practical proxy for many other soil health measures.

Samples can be analyzed consistently in a laboratory with sample plans designed to ensure affordable measurement of the site-specific SOC value range. SOC also changes more slowly over time than other indicators, making it ideal for long-term soil monitoring and management decisions. Due to its wide relevance and ease of measurement, SOC can be the most appropriate single indicator for assessing soil health.

7.2 Monitoring Soil Health

Measuring soil health involves a range of methods, each with its own strengths and limitations. Physical soil sampling can be costly and time consuming but provides highly accurate point data. The use of remote sensing imagery allows for wider coverage than soil sampling alone and can be more time and cost efficient, however it often lacks the precision of direct sampling and can be interrupted by external factors such as cloud coverage. Remote sensing can be combined with soil sampling by ML techniques, which can reveal important insights but often require significant computing power to achieve well-developed models. Taking all of the specific strengths and weaknesses into account is vital to selecting the most suitable method for monitoring specific soil health measures.

7.2.1 Sampling

Physical soil sampling provides highly accurate data specific to the area of interest and can include parameters such as chemical composition, nutrient content, and soil texture, which are essential for detailed monitoring. However, it can be labor-intensive, time-consuming and expensive when covering large or remote areas. These limitations can be alleviated by planning the soil sampling in relation to the area size and choosing smaller representative regions across the area of interest. The choice of laboratory analysis method is also essential, as there may be differences in analysis techniques, methodology, and equipment between laboratories which can introduce uncertainties in result accuracy.

7.2.2 Remote Sensing and Machine Learning

Remote sensing allows extensive spatial coverage and can provide detailed information into aspects such as soil moisture, texture, and vegetation indices over time. While this can be more cost effective for large scale or repeated monitoring than soil sampling alone, it is subject to limitations such as trade-offs between temporal, spatial, and spectral resolution and cloud cover interference.

Machine learning combined with soil sampling and remote sensing imagery provides a cost effective and detailed approach to monitoring SOC. It reduces the need for extensive physical soil sampling, enhances spatial coverage and can reveal patterns that traditional interpolation methods might miss. These models also help quantify uncertainty and, in some cases, can even identify errors in laboratory data. Prediction uncertainties can then guide targeted higher density physical sampling to fill knowledge gaps and improve results. The accuracy of predictions from machine learning methods ultimately depends on the quality of the training data and the suitability of the model used. Additionally, training complex models can be computationally demanding, although cloud-based solutions are addressing these challenges.

Although soil sampling and remote sensing have limitations when used independently, their combined application enables a more comprehensive and scalable approach to soil analysis by leveraging the precision of physical sampling and the broader efficiency of remote sensing imagery with machine learning techniques.

8 Recommendations for Global Soil Health Monitoring

Designing an effective global soil health monitoring system requires balancing spatial resolution, temporal frequency, and computational feasibility. Monitoring at very high spatial and temporal resolutions across the entire globe is not currently possible due to data availability and computational limitations. Instead, trade-offs must be made to ensure that resources are directed towards the areas where monitoring will have the greatest impact. As the sampling density decreases the measurement accuracy decreases, moderated by the variability of SOC within the study area. As an example of significant regional differences in variability, Ecometric's annual SOC monitoring of 75 projects in the UK over a four-year period has demonstrated high spatial variability with a mean SOCS of 80 (t/ha) and a standard deviation of 25 (t/ha), necessitating a higher sampling density than projects in South America where mean SOCS values are as low as 35 (t/ha) and standard deviation of 10 (t/ha).

A first step is to identify regions where soil health is most at risk. These priority areas can be detected using global or continental-scale remote sensing datasets, which provide valuable indicators of soil degradation despite limitations in spatial resolution. Examples include datasets tracking deforestation, land-use change, vegetation stress, and water availability. Using such indicators at broad scales allows monitoring efforts to focus on regions where poor soil health is most likely, without the prohibitive costs of high-resolution data processing across the entire globe.

Once priority areas have been identified, more detailed monitoring should be undertaken at regional or project scales. This process begins with selecting small-scale areas that are considered representative of the region, which are then sampled. The sample results are then used in conjunction with proxy data such as multispectral imagery to provide training data for ANNs. These trained models can then predict SOC across much larger unsampled areas. Importantly, these predictions must include explicit quantification of uncertainty to ensure confidence in results and guide subsequent sampling efforts.

Areas where model uncertainty is high can then be targeted for additional soil sampling. This adaptive approach ensures that fieldwork is focused where it is most needed, capturing the full variability of SOC in the target area and reducing uncertainty in the model predictions. By combining targeted sampling with remote sensing and machine learning, global monitoring can be made both more efficient and more accurate.

For areas selected for ongoing monitoring, temporal frequency should be as high as resources allow, ideally with annual monitoring rounds. Establishing a baseline with high-density soil sampling at the outset is also critical. While the initial cost may be significant, subsequent monitoring rounds can reduce sample density as the ANN becomes increasingly effective at predicting SOC trends in specific areas. Over time, advances in AI, improvements in computational efficiency, and the expansion of open-access earth observation datasets are likely to reduce costs further, making global soil health monitoring increasingly feasible.

References

Balenzano, A., Mattia, F., Satalino, G., Lovergine, F. P., Palmisano, D., & Davidson, M. W. J. (2021). Dataset of Sentinel-1 surface soil moisture time series at 1 km resolution over Southern Italy. *Data in Brief, 38*, 107345. https://doi.org/10.1016/J.DIB.2021.107345

Balenzano, A., Mattia, F., Satalino, G., Lovergine, F. P., Palmisano, D., Peng, J., Marzahn, P., Wegmüller, U., Cartus, O., Dąbrowska-Zielińska, K., Musial, J. P., Davidson, M. W. J., Pauwels, V. R. N., Cosh, M. H., McNairn, H., Johnson, J. T., Walker, J. P., Yueh, S. H., Entekhabi, D., ... Jackson, T. J. (2021). Sentinel-1 soil moisture at 1 km resolution: a validation study. *Remote Sensing of Environment, 263.* https://doi.org/10.1016/j.rse.2021.112554

Bastida, F., Eldridge, D. J., García, C., Kenny Png, G., Bardgett, R. D., & Delgado-Baquerizo, M. (2021). Soil microbial diversity–biomass relationships are driven by soil carbon content across global biomes. *ISME Journal, 15*(7), 2081–2091. https://doi.org/10.1038/s41396-021-00906-0

Brito, I., & Goss, M. J. (n.d.). *Soil Organic Matter as a Global Metric for Soil Heath: Integrating Soil Biology and Agro-Ecosystem Management.*

Chhabra, V., Sreethu, S., Kaur, G., Singh, A., Kaur, M., Siddiqui, M. H., & Gupta, R. K. (2025). Response of wheat crop to water-logged conditions under different land configurations and nutrient management. *Scientific Reports, 15*(1). https://doi.org/10.1038/s41598-024-83752-2

Comas, X., Terry, N., Hribljan, J. A., Lilleskov, E. A., Suarez, E., Chimner, R. A., & Kolka, R. K. (2017). Estimating belowground carbon stocks in peatlands of the Ecuadorian páramo using ground-penetrating radar (GPR). *Journal of Geophysical Research: Biogeosciences, 122*(2), 370–386. https://doi.org/10.1002/2016JG003550

Cruz-Martínez, K., Rosling, A., Zhang, Y., Song, M., Andersen, G. L., & Banfield, J. F. (2012). Effect of rainfall-induced soil geochemistry dynamics on grassland soil microbial communities. *Applied and Environmental Microbiology, 78*(21), 7587–7595. https://doi.org/10.1128/AEM.00203-12

Davidson, E. A., & Janssens, I. A. (2006). Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. *Nature, 440*(7081), 165–173. https://doi.org/10.1038/NATURE04514

De Baets, S., Poesen, J., Meersmans, J., & Serlet, L. (2011). Cover crops and their erosion-reducing effects during concentrated flow erosion. *CATENA, 85*(3), 237–244. https://doi.org/10.1016/J.CATENA.2011.01.009

Egmond, F. Van, Loonstra, E., & Rossel, R. A. V. (2010). Proximal Soil Sensing. *Proximal Soil Sensing, October 2015*. https://doi.org/10.1007/978-90-481-8859-8

Franzen, D. W., Specialist, S., Cihacek, L. J., Tools, S., Handling, S. S., & Collecting, S. S. (1998). *Importance of Soil Sampling Depth of Sampling. 990*(3).

Gholizadeh, A., Žižala, D., Saberioon, M., & Borůvka, L. (2018). Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging. *Remote Sensing of Environment, 218*(September 2017), 89–103. https://doi.org/10.1016/j.rse.2018.09.015

Glavič-Cindro, D., Bruggeman, M., Črnič, B., Nečemer, M., Petrovič, T., Prem, P., Vodenik, B., & Zorko, B. (2023). Comparison of different approaches of soil sampling uncertainty determination. *Applied Radiation and Isotopes, 194*. https://doi.org/10.1016/j.apradiso.2023.110676

I. Aleksander and T.J.Stoneham. (1979). *Guide to pattern recognition using random-access memories.*

Ji, W., Adamchuk, V. I., Chen, S., Mat Su, A. S., Ismail, A., Gan, Q., Shi, Z., & Biswas, A. (2019). Simultaneous measurement of multiple soil properties through proximal sensor data fusion: A case study. *Geoderma, 341*(January), 111–128. https://doi.org/10.1016/j.geoderma.2019.01.006

Jones, C. (2018). *Light Farming: Restoring carbon, organic nitrogen and biodiversity to agricultural soils.* www.amazingcarbon.com

Kibblewhite, M. G., Ritz, K., & Swift, M. J. (2008). Soil health in agricultural systems. *In Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 363, Issue 1492, pp. 685–701). Royal Society. https://doi.org/10.1098/rstb.2007.2178

Kirschbaum, M. U. F. (1995). The temperature dependence of soil organic matter decomposition, and the effect of global warming on soil organic C storage. *Soil Biology and Biochemistry, 27*(6), 753–760. https://doi.org/10.1016/0038-0717(94)00242-S

Laganière, J., Augusto, L., Hatten, J. A., & Spielvogel, S. (2022). Editorial: Vegetation Effects on Soil Organic Matter in Forested Ecosystems. In *Frontiers in Forests and Global Change* (Vol. 4). Frontiers Media S.A. https://doi.org/10.3389/ffgc.2021.828701

Lehmann, T., & Bruun, E. (1994). Hardware Learning in Analogue VLSI Neural Networks. In *Downloaded from orbit.dtu.dk on.*

Li, B., Wang, L., Kaseke, K. F., Li, L., & Seely, M. K. (2016). The impact of rainfall on soil moisture dynamics in a foggy desert. *PLoS ONE, 11*(10). https://doi.org/10.1371/journal.pone.0164982

Li, H., Van den Bulcke, J., Wang, X., Gebremikael, M. T., Hagan, J., De Neve, S., & Sleutel, S. (2020). Soil texture strongly controls exogenous organic matter mineralization indirectly via moisture upon progressive drying — Evidence from incubation experiments. *Soil Biology and Biochemistry, 151,* 108051. https://doi.org/10.1016/J.SOILBIO.2020.108051

Li, J., Wu, L., Chen, L., Zhang, J., Shi, Z., Ling, H., Cheng, C., Wu, H., Butler, A. D., Zhang, Q., Arslan, Z., Pierce, E. M., Su, Y., & Han, F. X. (2024). Effects of slopes, rainfall intensity and grass cover on runoff loss of mercury from floodplain soil in Oak Ridge TN: A laboratory pilot study. *Geoderma, 441,* 116750. https://doi.org/10.1016/J.GEODERMA.2023.116750

Liu, C., Chen, Z., Shao, Y., Chen, J., Hasi, T., & Pan, H.. (2019). Research advances of SAR remote sensing for agriculture applications: A review. In *Journal of Integrative Agriculture* (Vol. 18, Issue 3, pp. 506–525). Chinese Academy of Agricultural Sciences. https://doi.org/10.1016/S2095-3119(18)62016-7

Long, J., Liu, Y., Xing, S., Qiu, L., Huang, Q., Zhou, B., Shen, J., & Zhang, L. (2018). Effects of sampling density on interpolation accuracy for farmland soil organic matter concentration in a large region of complex topography. *Ecological Indicators, 93,* 562–571. https://doi.org/10.1016/j.ecolind.2018.05.044

Marchant, B. P., Dailey, A. G., & Lark, R. M. (2012). *Cost-effective sampling strategies for soil management.*

Masood, T. K., & Ali, N. S. (2023). Effect of Different Soil Organic Carbon Content in Different Soils on Water Holding Capacity and Soil Health. *IOP Conference Series: Earth and Environmental Science, 1158*(2). https://doi.org/10.1088/1755-1315/1158/2/022035

Mirsoleimani, H. R., Sahebi, M. R., Baghdadi, N., & El Hajj, M. (2019). Bare soil surface moisture retrieval from sentinel-1 SAR data based on the calibrated IEM and dubois models using neural networks. *Sensors (Switzerland), 19*(14), 1–12. https://doi.org/10.3390/s19143209

Mondal, A., Khare, D., Kundu, S., Mondal, S., Mukherjee, S., & Mukhopadhyay, A. (2017). Spatial soil organic carbon (SOC) prediction by regression kriging using remote sensing data. *Egyptian Journal of Remote Sensing and Space Science, 20*(1), 61–70. https://doi.org/10.1016/j.ejrs.2016.06.004

Nozari, S., & Borůvka, L. (2023). The effects of slope and altitude on soil organic carbon and clay content in different land-uses: A case study in the Czech Republic. *Soil and Water Research, 18*(3), 204–218. https://doi.org/10.17221/105/2022-SWR

Onwuka, B. (2018). Effects of Soil Temperature on Some Soil Properties and Plant Growth. *Advances in Plants & Agriculture Research, 8*(1). https://www.researchgate.net/profile/Brown-Onwuka/publication/323685894

Paola, J. D., & Schowengerdt R. A. (1995). A review and analysis of backpropagation neural networks for classification of remotely-sensed multi-spectral imagery. *International Journal of Remote Sensing, 16*(16), 3033–3058. https://doi.org/10.1080/01431169508954607

Piccoli, F., Barbato, M. P., Peracchi, M., & Napoletano, P. (2023). Estimation of Soil Characteristics from Multispectral Sentinel-3 Imagery and DEM Derivatives Using Machine Learning. *Sensors, 23*(18). https://doi.org/10.3390/s23187876

Prairie, A. M., King, A. E., & Francesca Cotrufo, M. (2023). Restoring particulate and mineral-associated organic carbon through regenerative agriculture. *Proceedings of the National Academy of Sciences of the United States of America, 120*(21). https://doi.org/10.1073/pnas.2217481120

Rao, D. L. N. (2019). *Microbiology and Biochemistry of Soil Organic Matter, Carbon Sequestration and Soil Health.* https://www.researchgate.net/publication/332038028

Ratke, R. F., Viana, P. R. N., Teodoro, L. P. R., Baio, F. H. R., Teodoro, P. E., Santana, D. C., Santos, C. E. da S., Zuffo, A. M., & Aguilera, J. G. (2024). Multispectral Sensors and Machine Learning as Modern Tools for Nutrient Content Prediction in Soil. *AgriEngineering, 6*(4), 4384–4394. https://doi.org/10.3390/agriengineering6040248

Salim, B. A., Mrema, G. C., Mlengera, *, Salim, B. A., & Mrema, G. C. (2017). Conservation tillage methods for soil erosion control on steep slopes: Case study of southern Uluguru Mountains, Tanzania. *International Journal of Agriculture, Environment and Bioresearch, 2*(04). www.ijaeb.org

Trivedi, P., Delgado-Baquerizo, M., Anderson, I. C., & Singh, B. K. (2016). Response of soil properties and microbial communities to agriculture: Implications for primary productivity and soil health indicators. *Frontiers in Plant Science, 7* (JULY2016). https://doi.org/10.3389/fpls.2016.00990

Vannoppen, W., Vanmaercke, M., De Baets, S., & Poesen, J. (2015). A review of the mechanical effects of plant roots on concentrated flow erosion rates. *Earth-Science Reviews, 150,* 666–678. https://doi.org/10.1016/j.earscirev.2015.08.011

Viscarra Rossel, R. A., & Brus, D. J. (2018). The cost-efficiency and reliability of two methods for soil organic C accounting. *Land Degradation and Development, 29*(3), 506–520. https://doi.org/10.1002/ldr.2887

Wuest, S. B., & Durfee, N. (2024). Temporal variability is a major source of uncertainty in soil carbon measurements. *Soil Science Society of America Journal, 88*(3), 830–845. https://doi.org/10.1002/saj2.20660

Yao, Y., Dai, Q., Gao, R., Gan, Y., & Yi, X. (2021). Effects of rainfall intensity on runoff and nutrient loss of gently sloping farmland in a karst area of SW China. *PLoS ONE, 16*(3 March). https://doi.org/10.1371/journal.pone.0246505

Young, M. A., Macreadie, P. I., Duncan, C., Carnell, P. E., Nicholson, E., Serrano, O., Duarte, C. M., Shiell, G., Baldock, J., & Ierodiaconou, D. (2018). Optimal soil carbon sampling designs to achieve cost-effectiveness: A case study in blue carbon ecosystems. *Biology Letters, 14*(9). https://doi.org/10.1098/rsbl.2018.0416

Zeraatpisheh, M., Garosi, Y., Reza Owliaie, H., Ayoubi, S., Taghizadeh-Mehrjardi, R., Scholten, T., & Xu, M. (2022). Improving the spatial prediction of soil organic carbon using environmental covariates selection: A comparison of a group of environmental covariates. *Catena, 208*(September 2021), 105723. https://doi.org/10.1016/j.catena.2021.105723

Zhao, W., Chuluunbat, G., Unagaev, A., & Efremova, N. (2024). *Soil nitrogen forecasting from environmental variables provided by multisensor remote sensing images.* http://arxiv.org/abs/2406.09812

Zhao, X., Yang, Y., Shen, H., Geng, X., & Fang, J. (2019). Global soil-climate-biome diagram: Linking surface soil properties to climate and biota. *Biogeosciences, 16*(14), 2857–2871. https://doi.org/10.5194/bg-16-2857-2019

Zhou, J., Sun, T., Shi, L., Kurganova, I., Lopes de Gerenyu, V., Kalinina, O., Giani, L., & Kuzyakov, Y. (2023). Organic carbon accumulation and microbial activities in arable soils after abandonment: A chronosequence study. *Geoderma, 435.* https://doi.org/10.1016/j.geoderma.2023.116496

Zhou, M., Liu, C., Wang, J., Meng, Q., Yuan, Y., Ma, X., Liu, X., Zhu, Y., Ding, G., Zhang, J., Zeng, X., & Du, W. (2020). Soil aggregates stability and storage of soil organic carbon respond to cropping systems on Black Soils of Northeast China. *Scientific Reports, 10*(1). https://doi.org/10.1038/s41598-019-57193-1

Žížala, D., Minarík, R., & Zádorová, T. (2019). Soil organic carbon mapping using multispectral remote sensing data: Prediction ability of data with different spatial and spectral resolutions. *Remote Sensing, 11*(24). https://doi.org/10.3390/rs11242947