Soil Organic Carbon Estimation

METHODOLOGY

ecometric.

Authors:
Hywel Evans, MBE PhD[2] BA- *Chief Technical Officer*
David Wright, BSc – *Chief Executive Officer*
Grace Milner, MSc BSc – *Geospatial Developer*

Version: 1.2.4
Document ID: ecometric-0045
Last updated: 05/12/2025
https://www.ecometric.co.uk

Contents

List of Tables

List of Figures

List of Equations

Units

Metric Tonnes	t
Hectare	Ha
Total Soil Volume	cm^3
Soil Bulk Density	g/cm^3
Dry Soil Weight	g
SOC Percentage	%
SOC	t/ha
SOC Stock	t
CO_2eq	t

Abbreviations

Soil Organic Carbon	SOC
CO_2 equivalents	CO_2e
Cation Exchange Capacity	CEC

Environmental Systems Research institute	ESRI
Keyhole Markup Language	KML
Keyhole Markup Language Zipped	KMZ
Geographic Java Script Object Notation	GeoJSON
Artificial Neural Network	ANN
Machine Learning	ML
Earth Observing System	EOS
Compact Geographic Stratification	CGS
Geographic Information System	GIS
Global Navigation Satellite System	GNSS

1. METHODOLOGY OVERVIEW

1.1. SCOPE

This methodology protocol uses remotely sensed multispectral imagery and soil sample results to train an Artificial Neural Network (ANN) to monitor changes in Soil Organic Carbon (SOC) stocks, within a project area, through time. The project area will be defined in the credit class document. The SOC change will be reported as CO_2 equivalent. SOC is crucial to soil health, fertility and ecosystem services including food production, making its preservation and restoration essential. This methodology will concentrate on the assessment of SOC sequestration as a major soil health characteristic, reported in:

1. SOC stocks (tSOC).
2. CO_2 equivalents (tCO_2e).

1.2. MOTIVATION

Soil contains approximately 2344 Gt of organic carbon globally and is the largest terrestrial pool of organic carbon (Stockmann et al., 2013). Small changes in SOC stock could result in significant impacts on the atmospheric carbon concentration. By monitoring the carbon levels in the soil, farmers and landowners will be able to measure the impact of their stewardship.

1.3. OUTLINE

The following steps will be followed to estimate change in SOC stocks within a project area:

1. Develop a soil sampling plan for the project area.
2. Sample collection and preparation.
3. Laboratory analysis of soil samples.
4. Estimation of SOC stocks using Machine Learning (ML) and remotely sensed multispectral imagery.
5. Convert SOC stocks to CO_2 equivalent stocks.
6. Calculate the change in CO_2e stocks between monitoring periods.

If historic SOC sampled data meets this methodology's sampling and laboratory analysis requirements, it may be used to calculate a historic baseline.

This methodology outlines an innovative approach using remote sensing data to train a neural network to estimate SOC stocks. This approach also allows for a significant reduction in the number of soil samples and in turn reduces costs.

2. PROJECT BOUNDARIES

2.1. SPATIAL BOUNDARIES

Spatial boundaries of the project area will be defined, including any parcels or stratification schemes, using an appropriate data format. Acceptable data formats include but are not limited to:

1. ESRI polygon shapefiles
2. KML/KMZ
3. GeoJSON.

2.2. MASKING

Any areas outside the defined spatial boundaries will be masked.

2.3. TEMPORAL BOUNDARIES

The project timeframe will be defined as the period during which SOC stocks will be monitored. This methodology will initially be based on annual sampling rounds and will follow existing guidelines on advised sampling time delays after the application of organic or inorganic fertiliser. Sampling will be conducted at 12-month intervals, or as close to 12-month intervals as possible depending on logistic requirements and required delays following fertiliser applications, to ensure temporal comparability. Modifications can be made if an extreme climatic event or disaster is declared in or near the project area, in which case clear justification will be provided.

3. ESTIMATING CARBON SEQUESTRATION USING REMOTELY SENSED MULTISPECTRAL IMAGERY AND NEURAL NETWORKS

3.1. BACKGROUND

Satellite imagery and other remote sensing data has been shown to provide a proxy for SOC; previous approaches were mainly based on spectral indices and some used machine learning. An example of the spectral index approach, (Thaler et al., 2019), developed a SOC index (SOCI) using three bands of WorldView-2 imagery, with central wavelength (ρ).

$$SOCI = \frac{\rho_{478}}{\rho_{659} - \rho_{546}} \#(\text{ SEQ Equation * ARABIC 1})$$

Bartholomeus et al., (2008) tested, in laboratory conditions, the performance of several spectral indices which had been developed to detect biochemical constituents (e.g., cellulose, lignin) for their ability to retrieve SOC. They found correlation for indices based on the visible part of the spectrum ($R^2 = 0.80$) and for the absorption features related to cellulose (around 2100 nm) ($R^2 = 0.81$). Rasel et al., (2017)) used remotely sensed variables such as elevation and forest type rather than image pixel values to estimate SOC. Gardin et al., (2021) used meteorological data, a land use map and MODIS Normalised Difference Vegetation Index (NDVI) imagery. This information was processed by advanced statistical methods to map SOC spatial distribution. Guo et al., (2021) estimated SOC and soil bulk density (SBD) through partial least square regression (PLSR) and extreme learning machine (ELM) neural networks, achieving a correlation between image reflectance and SOC% with R^2=0.67. They found that the combination of Sentinel 2 images and ELM obtained the best prediction results. ELMs are not as accurate as traditional backpropagation networks; they are generally used with problems that require real-time retraining of the network.

3.2. PROPOSED SOC PROXY

In this methodology Sentinel-2 multispectral data will be used as the proxy (inputs) and soil samples will be used to provide ground truth SOC (targets) data for the estimation of SOC stocks.

Additional data may be used if further research indicates a benefit. Alternative remote sensing imagery may be applied in place of Sentinel-2 data provided the spatial and spectral resolutions are similar or better.

3.3. NEURAL NETWORKS

This methodology will use Machine Learning (ML) in the form of an Artificial Neural Network (ANN). ML has been used by various other carbon sequestration methodologies, including both shallow and deep learning networks alongside Sentinel-2 imagery, as reviewed by (Odebiri et al., 2021). A further review investigated remote sensing techniques for SOC estimation, highlighting the most appropriate wavelengths and the use of ML (Angelopoulou et al., 2019).

The advantage of ANN, over other ML methods, is their ability to train directly on high dimensional data such as multispectral imagery. They have a high noise tolerance and can function on incomplete data. Neural network training is stochastic, producing slightly different predictions from each training session. The network may be trained multiple times and the results averaged to reduce the potential for extreme results, smoothing the data to provide more reliable and conservative estimates.

4. SOIL SAMPLING METHODS

For each project a suitable soil sampling scheme will be chosen based on field characteristics and practical considerations such as sampling capacity. Two primary stratification approaches may be applied: (i) area-based stratification using equal-sized blocks, and (ii) stratified random sampling using homogenous zones. In addition, two main sampling methods may be applied within strata: (i) distributed composite sampling, and (ii) clustered composite sampling.

Stratum boundaries, strata size, and sampling point locations may be adjusted in subsequent sampling rounds as knowledge on project SOC variability improves. Adjustments shall be documented and justified based on improved understanding of project-area conditions.

Projects may propose alternative sampling strategies where scientifically justified. Such approaches shall be supported by adequate project-specific rationale and submitted as a methodology deviation where required.

4.1. STRATIFICATION METHODS

Stratification refers to the technique of partitioning a population into smaller groups based on similar characteristics. It is the preferred sampling method as it can help to reduce the number of samples needed whilst increasing the accuracy of the results, (Booman et al., 2023).

4.1.1. AREA-BASED STRATIFICATION

Under area-based stratification, agricultural fields will be divided into blocks of approximately equal area. A single block may cover an entire field where appropriate. Each block will constitute a sampling stratum. This stratification approach is preferred in situations where there is limited understanding of existing project-specific SOC variation and/or limited data regarding variables known to impact SOC stocks.

4.1.2. ZONAL STRATIFICATION

Under zonal stratification, the project area will be divided into zones which represent relatively homogeneous conditions impacting SOC stocks. Relevant factors may include, but are not limited to, the following variables found to be good proxies to spatial variability of soil type, topography, Land Use/Land Cover (LULC), hydrology, satellite imagery, and climatic zone, (Lawrence et al., 2021). SOC estimations from machine learning algorithms validated within a comparable context to the project area (e.g. similar climate zone, soil type, management system) may also be used for stratification.

Strata may cover multiple fields, and a single field may contain multiple strata, dependant on project variability. This stratification approach is preferred in situations where there is existing knowledge on project-specific SOC variation and/or available data on variables known to impact SOC stocks.

4.2. SAMPLING METHODS

Within strata, two main sampling methods may be applied: (i) distributed composite sampling, and (ii) clustered composite sampling. For both sampling methods, multiple soil cores will be composited into a single sample for lab analysis. The exact number of soil cores within each composited sample will be dependent on field characteristics, sampling resources, and laboratory soil volume requirements. Typically, a composited sample will include 9-12 individual soil cores, but this number may vary within and between projects.

Single-core samples may also be used where determined to be more appropriate for the sampling objectives or analytical requirements.

In all cases, the locations of the soil sample positions will be reported by the sampling team.

4.2.1. DISTRIBUTED COMPOSITES

Under the distributed composites approach, cores will be arranged across the entire stratum in an evenly spaced grid design. All cores within a stratum will then be composited into a single soil sample. The density of cores within each stratum will primarily be determined by sampling logistics and laboratory soil volume requirements. This method aims to optimise the capture of spatial variation, particularly in situations where there is no prior knowledge of SOC patterns. This method is most appropriate where strata are relatively small in area to ensure spatial variation is adequately captured whilst retaining accuracy and avoiding averaging over excessive areas which could obscure real differences.

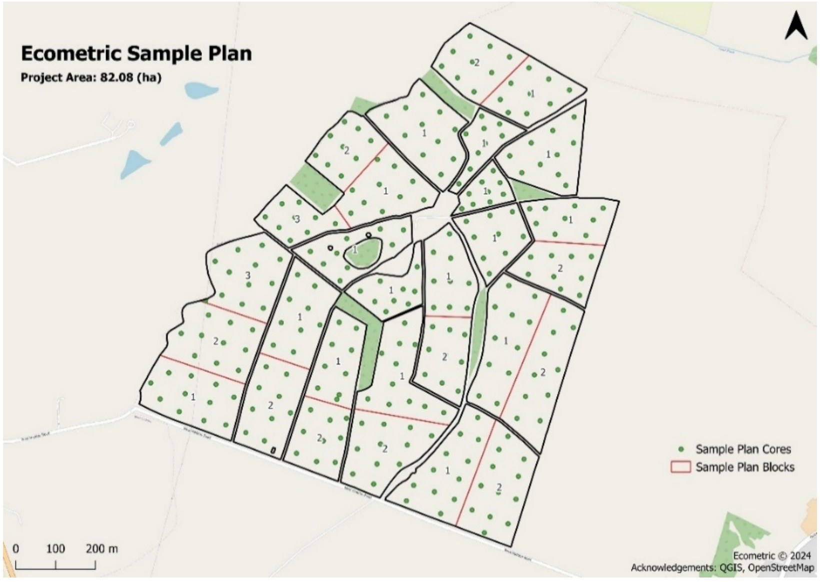

Figure 1 - Example of distributed composite sampling.

4.2.2. CLUSTERED COMPOSITES

For the clustered composites approach, the number and location of sampling points may be determined by stratum size and/or existing variability. Sampling points may be placed at random locations within the strata, in representative locations within the stratum where prior knowledge of SOC patterns is available, or in an evenly spaced grid pattern within the stratum. At each sampling point, several cores will be collected within a small radius (typically <20 m) and composited to produce a sample for that individual point. This differs from the distributed composites approach in which all soil cores collected across a stratum are combined into one composite. In the clustered approach, each point retains its own composite sample.

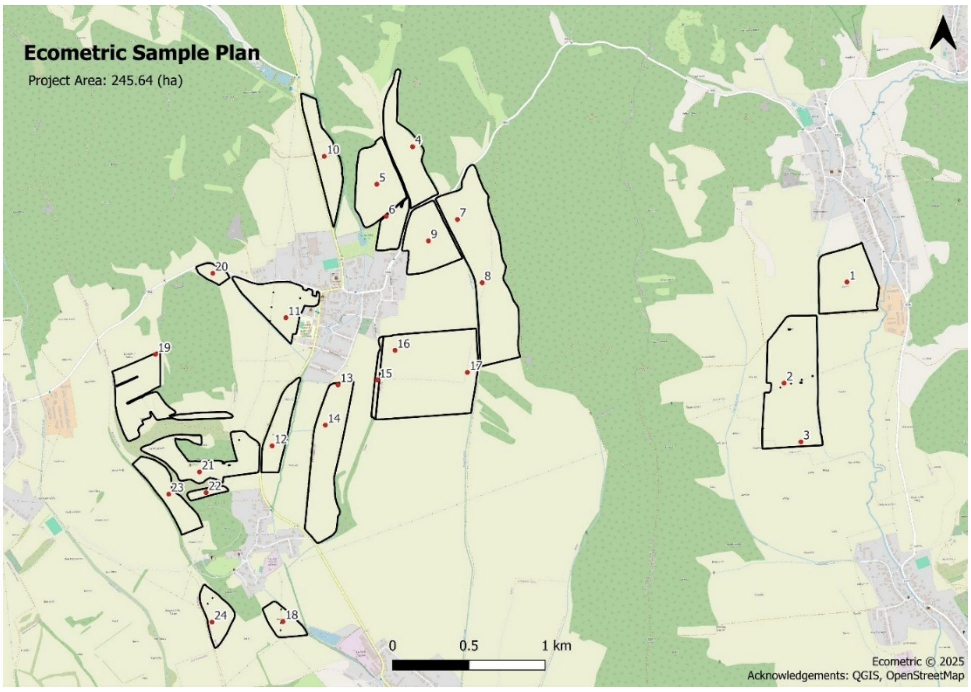

Figure 2 - Example of clustered composite sampling.

4.3. SOIL CORE EXTRACTION

To maintain the integrity of the results the Regen Network soil sampling guide (Booman et al., 2023) will be the main reference. The following method will be used to collect soil samples:

1. This proposal will use a consistent depth across monitoring rounds.
2. The sampling depth will be the same for all samples. The only exception to this will be where the nominated sampling depth cannot be reached due to bedrock or impenetrable layers. The sampling depth will be recorded where <30cm.
3. Each core will be georeferenced using a GNSS device with an accuracy of 10 metres or better.
4. Samples will be taken a sufficient distance from any tree, structure, or body of water so as not to be influenced.
5. The date/time of each core will be recorded for each sampling round.
6. Sampling rounds may be conducted earlier than 6 months after the application of organic amendments, provided that the contribution of the organic carbon (OC) content of the amendment can be estimated and deducted from the SOCS.

4.4. SAMPLING UNCERTAINTY

Soil represents a spatially continuous and highly heterogeneous medium, often treated as an infinite population in statistical terms due to the impracticality of sampling every individual unit (grain or location). Consequently, statistical inference is required to estimate population parameters (e.g. mean SOC concentration) from a finite sample set. These estimates inherently carry uncertainty, which can be estimated using confidence intervals, a method that reflects the variability of the estimate due to statistical sampling error (Montgomery & Runger, 2014). The precision of such estimates improves with increased sample size, as per the Central Limit Theorem, but collecting large sample sets is often constrained by cost and logistics (Webster & Oliver, 2007).

To address this, a methodology is employed that synthetically augments the sample density using a combination of observed sample data and multispectral imagery processed through a trained neural network. This approach generates synthetic data points that maintain spatial consistency with the resolution of the imagery, effectively increasing the sampling density without additional fieldwork. The reliability of inferences drawn from this synthetic dataset is contingent on the resolution of the imagery used during the baseline monitoring period. If resolution is reduced, the uncertainty in derived estimates increases proportionally.

The uncertainty in SOC estimation is inversely related to the sampling density and directly related to the spatial variability of SOC, which can be quantified using the standard deviation or coefficient of variation. This relationship can be simulated by generating random values with defined variability across grids of decreasing sample density. Such simulations illustrate the impact of sampling resolution on statistical confidence, as shown in Figure 3, and may be used to determine optimal sampling density if SOC variability is known. These principles align with geostatistical theory and spatial sampling design (Isaaks & Srivastava, 1991).

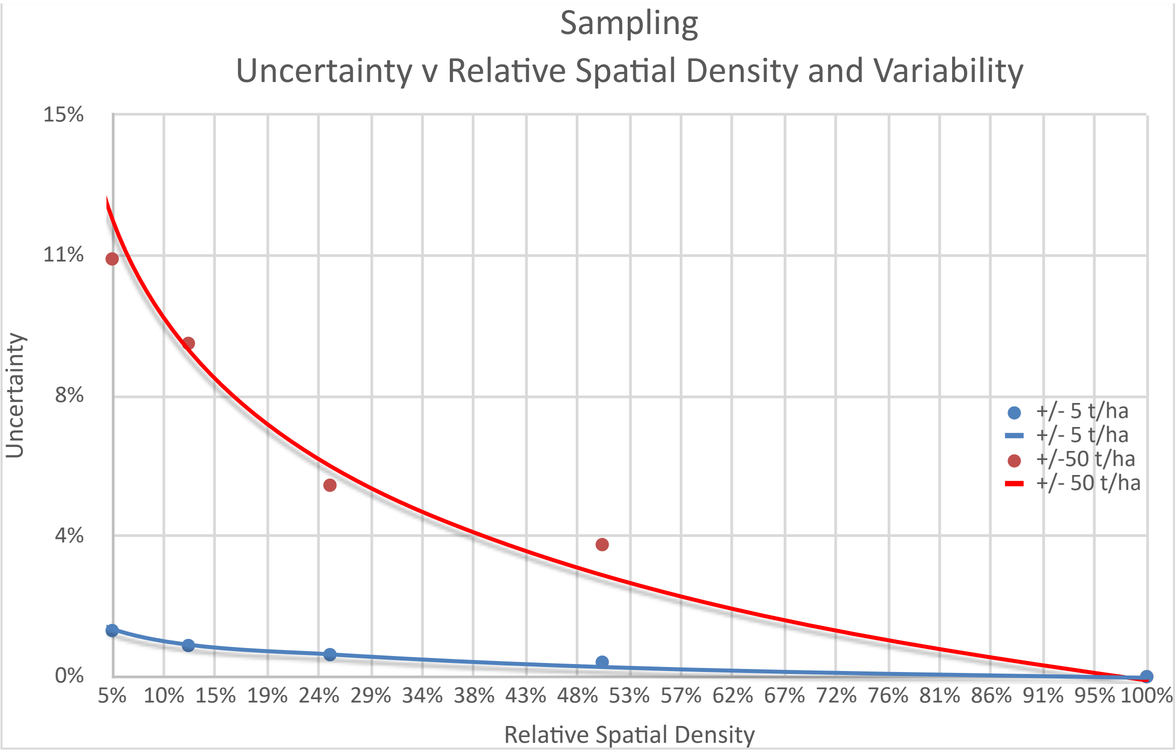

Figure 3 - Sample Uncertainty v Relative Sample Density. Colours denote the variability (Standard Deviation) within the sample area.

5. PROCESSING WORKFLOW

The following workflow outlines the method used to estimate SOC stocks using remotely sensed imagery, ancillary data and ML. All images and ancillary data included in the analysis will be specified in the project report. The workflow sequence will be:

1. Soil sampling.
2. Ancillary data, if used.
3. Sample analysis.
4. Image and ancillary data pre-processing.
5. Neural network training data.
6. Neural network training.
7. Estimating SOC with the previously trained neural network.
8. Project reporting.

5.1. ANCILLARY DATA

Ancillary data may be used to augment the ANN training dataset by adding additional predictors. The ancillary data may include:

1. Date, to allow for seasonal change.
2. Soil type.
3. Temperature.
4. Rainfall.
5. Soil moisture.
6. Nitrogen levels.
7. Slope.
8. Altitude.

The soil sample dates and the sample dates for the ancillary data, where relevant, will be chosen to be temporally close.

6. SAMPLE ANALYSIS

The DUMAS dry combustion laboratory test will be the preferred method for the laboratory measurement of SOC % due to its direct and complete measurement of total organic carbon with the highest reproducible accuracy.

The Walkley-Black wet combustion method is an acceptable moderate accuracy alternative for rapid field assessments, historical data continuity and low-resource settings where DUMAS capability is unavailable but is inferior to DUMAS due to incomplete oxidation leading to underestimation. It typically recovers only 60% to 80% of total SOC depending on soil type and conditions due to partial oxidation by potassium dichromate (Shamrikova et al., 2022). A conversion factor, commonly 1.3, is frequently used to adjust Walkley-Black results to better approximate SOC (Shamrikova et al., 2022), though this factor can vary with soil properties, introducing inconsistency and potential bias in analysis.

Loss on Ignition (LOI) may only be used when DUMAS or Walkley-Black laboratory tests are geographically unavailable as a low accuracy mass-loss measurement of organic matter, to maintain relative comparison with historic organic matter data records.

6.1. LABORATORY STANDARDS

Table 1 in Regen Network soil sampling guide (Booman et al., 2021) will be used to comply with Regen Network laboratory specific instructions, laboratory accreditation requirements and approved laboratories. The analysis must follow standard recommendations or standard procedures for SOC analysis. DUMAS will be the preferred laboratory method with Walkley-Black an acceptable secondary method alternative. Loss on Ignition (LoI) may only used in geographies where DUMAS and Walkley-Black are unavailable. Any future new methods of analysis that improve on the accuracy of current methods may be adopted.

If possible, the same analysis type using the same laboratory will be continued throughout the Project Crediting Period to ensure comparability. Exceptions are allowed where justified e.g. higher accuracy laboratory analysis methods become available, or the laboratory stops providing the service or increases cost to a degree that renders the project commercially unviable.

6.2. CARBON STOCK CALCULATION

The following equation will be used to quantify SOC stock (t/ha), using laboratory reported percent soil organic carbon (%), bulk density (g/cm^3) and sample depth (cm):

$$SOC\ Stock = Sample\ Depth * Bulk\ Density * SOC \#(\ SEQ\ Equation \backslash * \ ARABIC\ 2)$$

7. NEURAL NETWORK PROCESSING

The processing by the neural network requires the following steps:

- Collection of imagery.
- Collection of ancillary data, if applicable.
- Training.
- Estimation of SOC.

7.1. IMAGERY

Remotely sensed multispectral or hyperspectral image data will be used for ANN training. As an example, Sentinel-2[1] image bands are shown in Table 1.

Table 1 Sentinel-2 Multispectral Bands

Band Number	Central Wavelength (nm)	Bandwidth (nm)	Spatial Resolution (m)	Remarks
1	443	20	60	Aerosols
2	490	65	10	Blue
3	560	35	10	Green
4	665	30	10	Red
5	705	15	20	Red Edge 1
6	740	15	20	Red Edge 2
7	783	20	20	Red Edge 3
8	842	115	10	Near IR
8a	865	20	20	Red Edge 8
9	945	20	60	Water Vapour
10	1375	30	60	Cirrus Detection
11	1610	90	20	SWIR 1
12	2190	180	20	SWIR 2

7.2. IMAGE PROCESSING

Imagery with a sensing date as close as possible to the sampling date will be used. Image processing will include:

1. Processed to provide Bottom of Atmosphere (BOA) reflectance values.
2. Coordinate conversion as required (e.g. British National Grid for UK projects).
3. Resampling imagery and ancillary data to the same resolution (normally 10m).
4. Band stacking to create a multispectral image.

7.3. NEURAL NETWORK TRAINING DATA

A GIS sampling tool will be used to extract multispectral reflectance data at each soil sample location. This data will then be paired with its respective SOCS values to create a training dataset. A proportion of the training dataset will be withheld for validation and test data. The most common fractions for validation and testing data are 15% (Shahin et al., n.d.). Validation data may not be required for certain network architectures.

[1] https://eos.com/find-satellite/sentinel-2/

7.4. ESTIMATED SOIL ORGANIC CARBON STOCK

A trained network and remotely sensed imagery covering the project area will be used to estimate SOC stock at sampled and unsampled locations. The estimated SOC stock (t/ha) will be exported as a raster image. Post-processing will then be used to summarise the results as required e.g. mean and total stock values for strata, fields, and the total project area.

8. CALCULATING CREDITABLE CARBON CHANGE

8.1. BASELINE DEFINITION

The baseline SOC stock is defined here as the total carbon stock calculated for the project's initial monitoring date, or date of the first sampling round. The methodology will use a static baseline for each project, calculated as the total SOC stocks at the initial monitoring date.

Subsequent monitoring rounds will be compared to the baseline, or historical maximum SOC stock, whichever is greater, to calculate creditable carbon change.

8.2. GROSS SOIL ORGANIC CARBON STOCK CHANGE

The gross change in SOC stocks between reporting periods is estimated as the difference between the total SOC stock recorded by the current monitoring round ($tSOC_{(t)}$), minus the maximum historically recorded total SOC stock ($tSOC_{(Max)}$):

$$Gross\ SOC\ Stock\ Change = tSOC_{(t)} - tSOC_{(Max)} \#\left(\ SEQ\ Equation \backslash{*}\ ARABIC\ 3\right)$$

8.3. CONVERTING SOC STOCKS TO CO_2 EQUIVALENTS

Converting soil organic carbon stocks to CO_2 equivalent stocks (CO_2e) will be done by multiplying the SOC stocks (t) by a conversion factor of 3.67, the ratio of the molecular mass of carbon dioxide (44) to that of carbon (12):

$$CO_2e = SOC \times 3.67 \#\left(\ SEQ\ Equation \backslash{*}\ ARABIC\ 4\right)$$

The same applies for converting the change in the total SOC between two monitoring periods into CO_2 equivalents (CO_2e).

8.4. NET SOIL ORGANIC CARBON STOCK CHANGE

The gross SOC stock change must be adjusted for the error and uncertainty related to the SOC predictions from the ANN to calculate the net SOC stock change. In the context of this methodology, error refers to the difference between the ANN result and the lab analysis result from the soil sample. Uncertainty refers to the uncertainty in the ANN's SOC predictions in unsampled pixels.

8.4.1. ERROR

The method for calculating error in the ANN results will be to compare geographically corresponding network predictions to sample results data. The error will be quantified using the Absolute Percentage Error (APE) for each sample point or stratum, and the overall error for the project area will be calculated using the average of all APE values, referred to as the Mean Absolute Percentage Error (MAPE):

$$MAPE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right| \#\left(\ SEQ\ Equation \backslash{*}\ ARABIC\ 5\right)$$

with number of values (n), soil sample value (A_t), and network prediction (F_t).

8.4.2. UNCERTAINTY

The MAPE provides a scale-independent measure of the average prediction error of the ANN results in the sampled locations and can be used to estimate the uncertainty of the ANN results in the unsampled locations.

The reliability of the MAPE as a measure of uncertainty in unsampled locations depends on the density and spatial distribution of the validation soil samples used in the calculation. For this reason, it is preferable where possible to account for the stability of the MAPE calculation itself.

In the context of this methodology, the stability of the MAPE calculation will be determined through a Monte Carlo resampling procedure, which is commonly used in accuracy assessments within the remote sensing field (Hsiao & Cheng, 2016; Lyons et al., 2018). Using this approach, the MAPE will be repeatedly recalculated after omitting a small subset of validation points in each iteration. The resulting distribution of MAPE values reflects the stability of the original MAPE calculation.

From the distribution of MAPE values generated through the Monte Carlo resampling procedure, the 90th percentile ($MAPE_{90}$) will be calculated. This corresponds to the MAPE value that is only exceeded by 10% of the simulations, thereby representing a conservative, upper-bound estimate of model error.

A stability threshold of 1% is applied to determine whether the original MAPE ($MAPE_{orig}$) is sufficiently robust. If $MAPE_{90}$ exceeds $MAPE_{orig}$ by more than 1%, then $MAPE_{orig}$ is considered unstable and $MAPE_{90}$ will be used as the final MAPE ($MAPE_{final}$) in further calculations. If the difference is less than or equal to 1%, the stability is deemed acceptable and $MAPE_{orig}$ is retained as $MAPE_{final}$.

The final MAPE is therefore defined as:

$$MAPE_{final}\begin{cases} MAPE_{90} & \text{if } MAPE_{90} > MAPE_{orig} \times 1.01 \\ MAPE_{orig} & \text{if } MAPE_{90} \leq MAPE_{orig} \times 1.01 \end{cases}$$

$MAPE_{final}$ (%) and gross SOC stock change (tCO_2e) will be used to calculate the uncertainty adjustment (tCO_2e) for the entire project area. This represents a deduction as a proportion of total SOC stock change, accounting for ANN uncertainty.

$$Uncertainty\ Adjustment = MAPE_{final} \times Gross\ SOC\ Stock\ Change \#(\ SEQ\ Equation\ \backslash *\ ARABIC\ 6)$$

The exact procedure used to perform the Monte Carlo resampling approach of the MAPE may differ between projects and monitoring rounds, including the number of iterations performed, the proportion of the validation soil sample points omitted in each iteration, and the method for selecting a subset of validation points to omit in each iteration. These parameter choices must be clearly reported and justified.

Measuring the stability of the MAPE calculation provides an indirect but meaningful reflection of the uncertainty originating in soil sampling designs. When the validation soil sample dataset is relatively large and well distributed, the MAPE value typically remains stable across recalculation iterations which suggests a more reliable assessment of ANN error. In contrast, when sample density is low then each sample point has a proportionally greater impact on the MAPE calculation, typically leading to a wider spread of values hence a less stable MAPE.

Accounting for ANN uncertainty through both the MAPE value and the stability of the MAPE value itself allows the indirect inclusion of the impact of reduced sample density on ANN uncertainty, thereby providing a more robust and conservative estimation of the confidence on the ANN results.

Other suitable methods for error and/or uncertainty estimations may be chosen where scientifically justified.

8.4.3. NET CHANGE

The net change in SOC stocks is calculated by deducting the uncertainty adjustment from the gross change in SOC stocks.

$$Net\ SOC\ Stock\ Change = Gross\ SOC\ Stock\ Change - Uncertainty\ Adjustment \#(\ \text{SEQ Equation \textbackslash* ARABIC 7})$$

8.5. DEDUCTIONS

The deduction (D) for the project will be the sum of the following factors:

1. Yield-related Leakage Deduction (L)
2. Greenhouse Gas Emissions Deduction (E).

The total deduction (tCO_2e) is calculated as sum of the deductions:

$$D = L + E \#\left(\text{SEQ Equation * ARABIC 8} \right)$$

8.5.1. YIELD-RELATED LEAKAGE

Yield-related leakage refers to any significant reduction in yield within the project area which may lead to increased production elsewhere, hence increased emissions outside of the project area. In this methodology, a significant yield reduction refers to a project average percentage reduction in yield of 10% or more in comparison to the 5-year average or, in the event of climatic conditions that reduce regional yield, a reduction in yield of 10% or more below the regional average for that crop type.

The yield leakage exceeding the 10% threshold will be calculated as shown for each relevant crop type.

$$Yield\ Leakage\ = \left(5\ Year\ Average\ Yield \times 0.9 \right) - Project\ Average\ Yield \#\left(\text{SEQ Equation * ARABIC 9} \right)$$

This is then multiplied by the relevant crop area and crop-specific average emissions reported by the project during the current monitoring period to calculate the leakage deduction.

$$Leakage\ Deduction\ = Yield\ Leakage \times Crop\ Area \times Crop\ Average\ Emissions \#\left(\text{SEQ Equation * ARABIC 10} \right)$$

The leakage deductions for each relevant crop type in the project are summed to give a total project leakage deduction.

8.5.2. GREENHOUSE GAS EMISSIONS

The total Greenhouse Gas Emissions (GHG) deduction refers to emissions from all food and fibre production within the geographic boundaries of the Project Area emitted during the monitoring period, in tCO_2e, calculated using an IPCC higher tier (tier II or III) method of GHG reporting or other approved credible Registry GHG Calculation method that Regen Registry recognises and accepts, in accordance with the Credit Class.

8.6. CREDITABLE CARBON CHANGE

The total creditable carbon change (tCO_2e) in the project area for a given reporting period is calculated as the net change in SOC stock between the current monitoring period and the previous monitoring period with the highest total SOC stock, minus total deductions.

One EcoCredit will be issued per tCO2e positive creditable carbon change. Of the total creditable carbon change, 20% will be placed in a buffer pool to whilst the remaining 80% is available for trade.

9. DATA REPORTING

9.1. REPORT

After each monitoring round, a report will be submitted to the Regen Registry including a description of the methods used for soil sampling, analysis of samples, as well as the equations and references that were used. The reported results for each section of this Methodology will be accompanied by the supporting data. In the case of GIS or remote sensing data the SOC maps will be included as images within the report for illustrative purposes. The original vector and raster files will be kept by ecometric ltd. Any documentation containing calculations and statistical analysis will also be retained.

9.2. MONITORING REPORT

The monitoring report will include:

1. Sampling stratification design. This states the average strata size and sample numbers per stratum employed in the monitoring round.
2. Method of assigning strata boundaries and core locations including, GIS file format used, sample labelling system and minimum GNSS absolute accuracy of the sampling team georeferencing equipment.
3. Soil sampling contractor and sampling equipment type used in the monitoring round including GNSS accuracy.
4. Coordinate reference system used for geoprocessing.
5. Sampling date.
6. Selected laboratory, laboratory accreditation, laboratory tests used.
7. AI training method summary.
8. Method used to quantify AI error and the mean associated error for the monitoring period.
9. Contact information of the independent contractor used to gather GHG emissions data and calculate total project area GHG emissions.
10. AI SOC results. Field level results to include, where possible:
    - Field area (ha).
    - Monitoring interval crop type.
    - Gross mean SOCS (t/ha).

    For all monitoring rounds after baseline, tabulated results will list previous monitoring rounds to allow comparison.
11. Gross and net SOCS change between monitoring rounds (tSOC) and (tCO_2e).
12. GHG emissions summaries and full report references for relevant monitoring rounds.
13. Yield related leakage reported against 5-year average crop yields.
14. Credit statement, with total creditable credits separated into buffer credits (20%) and tradeable credits (80%).

Where appropriate, the relevant information may be presented in alternative formats e.g. a monitoring report and a supporting credit statement, either in separate files or combined into a single file.

9.3. PUBLICLY REPORTED DATA

The supporting data that will be publicly displayed on the Regen Registry is shown in Table 2:

Table 2 Public Data

Folder	Sub Folder	Contents	File Type	Number of Files
Public data		Project Plan	.docx or .pdf	1
		Emissions Report	.pdf	1
		Emissions Report Methodology	.pdf	1
		Monitoring Report	.docx or .pdf	1

9.4. NON-PUBLIC DATA

The additional commercially sensitive data will not be displayed publicly but will be made available to auditors, verifiers, diligence agencies and ratings agencies on application.

Table 3 Non-Public Data

Folder	Sub Folder	Contents	File Type	Number of Files
AI Data	Image	Remotely Sensed Images	.tiff/.tif	As required
	Network Training Data	Training Data	.csv	As required
	Trained Network	Network Settings	.mat	1
		Project Area Network Result	.csv	1
Emissions	Harvest Report Raw Data	Emissions Calculator Input Data	.docx .pdf .xlsx	Variable, by diversity of farming system and operations
Historic Yields and Cropping.		Historic 5-Year Management Plan.	.xlsx	1
		Monitoring Season Cropping Plan	.xlsx	1. May be omitted if included in 5-year
Land Registry		Land Registry Titles	.pdf	Variable, by farm size
Report		Monitoring Report	.pdf	1
Sampling Results		Sample Results	.xlsx	1
		Individual Laboratory reports	.csv	Variable, by project size and laboratory reporting interval

10. DATA STORAGE

All data used during the analysis will be stored for 5 years after the completion of the project. This data includes:

1. All raster and vector data used in geospatial analysis to generate results for any section of the methodology.
2. A copy of all laboratory reports.

3. All the relevant field data from the soil sample collection process (dates, tools, procedures, sample locations).
4. Documentation outlining calculations and results of statistical analysis.

11. DATA VERIFICATION

All data and information collected during the monitoring reporting as stated in para 5.1.1, including all publicly reported data as outlined in para 5.1.2 and non-public data as outlined in para 5.1.3 will be made available to the verifier.

References

Angelopoulou, T., Tziolas, N., Balafoutis, A., Zalidis, G., & Bochtis, D. (2019). Remote sensing techniques for soil organic carbon estimation: A review. *Remote Sensing*, *11*(6), 1–18. https://doi.org/10.3390/rs11060676

Bartholomeus, H. M., Schaepman, M. E., Kooistra, L., Stevens, A., Hoogmoed, W. B., & Spaargaren, O. S. P. (2008). Spectral reflectance based indices for soil organic carbon quantification. *Geoderma*, *145*(1–2), 28–36. https://doi.org/10.1016/J.GEODERMA.2008.01.010

Booman, G., Craelius, A., Deriemaeker, B., Landua, G., Szal, W., & Weinberg, B. (2021). *Regen Network Whitepaper*.

Booman, G., Horning, N., Bennetts, S., Leiker, S., & Steinherz, R. (2023). *Methodology for GHG and Co-Benefits in Grazing Systems v1.0 Regen Network Development, Inc*. https://www.regen.network

Gardin, L., Chiesi, M., Fibbi, L., & Maselli, F. (2021). Mapping soil organic carbon in Tuscany through the statistical combination of ground observations with ancillary and remote sensing data. *Geoderma*, *404*, 115386. https://doi.org/10.1016/J.GEODERMA.2021.115386

Guo, L., Sun, X., Fu, P., Shi, T., Dang, L., Chen, Y., Linderman, M., Zhang, G., Zhang, Y., Jiang, Q., Zhang, H., & Zeng, C. (2021). Mapping soil organic carbon stock by hyperspectral and time-series multispectral remote sensing images in low-relief agricultural areas. *Geoderma*, *398*, 115118. https://doi.org/10.1016/J.GEODERMA.2021.115118

Hsiao, L. H., & Cheng, K. S. (2016). Assessing uncertainty in LULC classification accuracy by using bootstrap resampling. *Remote Sensing*, *8*(9). https://doi.org/10.3390/rs8090705

Isaaks, E., & Srivastava, M. (1991). An Introduction to Applied Geostatistics. *Computer & Geosciences*, *17*(3), 471–473.

Lawrence, P., Boom, G., Thornton, A., & Kane, D. (2021). *WG Project 1: Modeling and Stratification Technology Assessment: Phase 1a Ecosystem Services Market Consortium/ Ecosystem Services*.

Lyons, M. B., Keith, D. A., Phinn, S. R., Mason, T. J., & Elith, J. (2018). A comparison of resampling methods for remote sensing classification and accuracy assessment. *Remote Sensing of Environment*, *208*, 145–153. https://doi.org/10.1016/J.RSE.2018.02.026

Montgomery, D., & Runger, G. (2014). *Applied statistics and probability for engineers*. Wiley.

Odebiri, O., Odindi, J., & Mutanga, O. (2021). Basic and deep learning models in remote sensing of soil organic carbon estimation: A brief review. *International Journal of Applied Earth Observation and Geoinformation*, *102*(June), 102389. https://doi.org/10.1016/j.jag.2021.102389

Rasel, S. M. M., Groen, T. A., Hussin, Y. A., & Diti, I. J. (2017). Proxies for soil organic carbon derived from remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, *59*, 157–166. https://doi.org/10.1016/J.JAG.2017.03.004

Shahin, M. A., Holger, ;, Maier, R., & Jaksa, M. B. (n.d.). *Data Division for Developing Neural Networks Applied to Geotechnical Engineering*. https://doi.org/10.1061/ASCE0887-3801200418:2105

Shamrikova, E. V., Kondratenok, B. M., Tumanova, E. A., Vanchikova, E. V., Lapteva, E. M., Zonova, T. V., Lu-Lyan-Min, E. I., Davydova, A. P., Libohova, Z., & Suvannang, N. (2022). Transferability between soil organic matter measurement methods for database harmonization. *Geoderma*, *412*. https://doi.org/10.1016/j.geoderma.2021.115547

Stockmann, U., Adams, M. A., Crawford, J. W., Field, D. J., Henakaarchchi, N., Jenkins, M., Minasny, B., McBratney, A. B., Courcelles, V. de R. de, Singh, K., Wheeler, I., Abbott, L., Angers, D. A., Baldock, J., Bird, M., Brookes, P. C., Chenu, C., Jastrow, J. D., Lal, R., … Zimmermann, M. (2013). The knowns, known unknowns and unknowns of sequestration of soil organic carbon. *Agriculture, Ecosystems & Environment*, *164*, 80–99. https://doi.org/10.1016/j.agee.2012.10.001

Thaler, E. A., Larsen, I. J., & Yu, Q. (2019). A New Index for Remote Sensing of Soil Organic Carbon Based Solely on Visible Wavelengths. *Soil Science Society of America Journal*, *83*(5), 1443–1450. https://doi.org/10.2136/sssaj2018.09.0318

Webster, R., & Oliver, M. (2007). *Geostatistics for Environmental Scientists*. Wiley.