Methodology Proposal for Soil Organic Carbon Estimation in Regenerative Cropping and Managed Grassland Ecosystems

REMOTE SENSING AND MACHINE LEARNING

Authors:

Hywel Evans, MBE PhD BA- *Chief Technical Officer*

David Wright, BSc – *Chief Executive Officer*

Version: 1.2.2

Document ID: ecometric-0044

Last updated: 11/07/2024

https://www.ecometric.co.uk

Contents

List of Tables

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
CO2 equivalents	CO2e
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 CO2 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, consisting of:

1. SOC stocks.
2. CO2 equivalents (CO2e).

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 the soil organic carbon 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 soil organic carbon 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 CO2 equivalent stocks.
6. Calculate the change in CO2e stocks between monitoring periods.

SOC stocks measured in the first sampling round (Baseline), will be compared to those calculated in subsequent sampling rounds to quantify changes in carbon stocks after project commencement. If historic SOC 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 based on using remote sensing data to train a neural network to estimate SOC stocks, within and beyond the project area. 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 one or more of the following data formats:

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

2.1.1. MASKING

Any areas outside the defined spatial boundaries will be masked.

2.2. 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. Sampling will be conducted at 12-month intervals to ensure temporal comparability and will follow laboratory guidelines on advised sampling time delays after the application of organic or inorganic fertilizer. The schema will be modified if an extreme climatic event or disaster is declared in or near the project area.

3. CALCULATING 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, Larsen and Yu, (2019), developed a SOC index (SOCI) using three bands of WorldView-2 imagery, with central wavelength (ρ).

$$SOCI = \frac{\rho_{478}}{\rho_{659} - \rho_{546}} \tag{1}$$

Bartholomeus et al., (2008) tested, in laboratory conditions, the performance of several word insert caption for equation right justified

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 (R2 = 0.80) and for the absorption features related to cellulose (around 2100 nm) (R2 = 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. 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. Their method resulted in a correlation between image reflectance and SOC% with R^2=0.67.

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. It is the intention to use additional sensors, if further research suggests indicates a benefit. Possible sensors might include Sentinel-1 radar imagery to address the problems of cloud cover.

3.3. NEURAL NETWORKS

This methodology will use Machine Learning (ML) in the form of an Artificial Neural Network (ANN). The choice of ANNs methods was made by research into their advantages and disadvantages and by comparison of results from other methods. ML has been used by other carbon sequestration methodologies, as reviewed by Odebiri, Odindi and Mutanga, (2021). A further review investigated remote sensing techniques for SOC estimation, highlighting the significant wavelengths and the use of ML [8]. Both shallow and deep learning networks have been reviewed with reference to using Sentinel-2 imagery by Odebiri, Odindi and Mutanga, (2021).

3.3.1. ADVANTAGES

The advantage of neural networks, over other ML methods, is that they can train directly on high dimensional data, such multispectral imagery. They have a high fault tolerance and can function on incomplete data.

3.3.2. DISADVANTAGES

Neural network training is stochastic, which produces slightly different predictions from each training session. Therefore, the network in this methodology will be run multiple times and the results will be the average of the multiple iterations. This approach reduces the potential for extremes to occur and smooths the data to provide more reliable and conservative estimates.

3.3.3. IMPLEMENTATION OF THE ANN

The neural network code will be compiled and packaged as a standalone App.

3.4. SOIL SAMPLING METHODS

Compact Geographic Stratification (CGS) will be the preferred soil sampling scheme used to collect the Year 0 baseline soil samples, but stratified random may also be used. The main reason for choosing the CGS method was its suitability to areas where there is no previous knowledge of variability of SOC or a proven suitable proxy to provide accurate stratification [9]. Also, in laboratory experiments targeted at SOC and remote sensing data CGS was used by Bartholomeus et al., (2008). CGS assumes that the sub-sample areas are smaller than the global variability [10]. Knowledge of the variability gained during baseline CGS sampling, will help optimize the number of samples in future sampling rounds and guide any change in the sampling methods in future sampling rounds. Soil samples will be taken to a depth of 0-30cm, with all cores gathered within each stratum composited into a single sample to avoid point inaccuracy and reduce cost. The DUMAS dry combustion laboratory test will be used exclusively for the direct measurement of SOC %, apart from geographies where the DUMAS test isn't available when Loss on Ignition (LOI) may be used.

3.5. PROCESSING WORKFLOW

The following workflow outlines the method used to estimate SOC stocks using Sentinel-2 imagery, ancillary data and ML. Other high spatial and spectral resolution imagery may be used in the future. 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.

3.5.1. SOIL SAMPLING SCHEME

The proposed CGS method will divide each field parcel into equal sized stratum with 10 or more composite core samples gathered per stratum. The location of the cores will be in an equally spaced grid pattern within each stratum. A custom GIS tool will be used to create the core location grid in each stratum and the geo-locations of the actual in-field core positions will be reported by the sampling team and recorded in supporting data to evidence that the sampling scheme was followed.

3.5.2. SAMPLING UNCERTAINTY

Maximum strata size will be determined by estimating Project Area SOC variability, to ensure that each stratum is as homogeneous as possible within the constraints of cost. Historical Project Area sample results or open-source regional SOC data will be used to estimate the range of SOC values within a project area and sampling uncertainty will be estimated for this range through a sampling simulation. This simulation calculates the mean of 100 random sample values arranged in a 2-dimensional grid, constrained between an upper and lower value that is representative of the estimated project area value range. Each simulation is repeated up to 500 times and a Mean Absolute Percentage Error (MAPE) is calculated from the mean of each individual calculation. The process is repeated at reducing sampling densities to graphically display the effect of reducing sample density (increasing CGS grid size) on the MAPE to identify the density that best balances accuracy with cost. The MAPE is reported as sampling uncertainty for each monitoring round, limited by a maximum acceptable uncertainty of 20%.

An example set of results using this method to simulate changes in core density are shown in Figure 1.

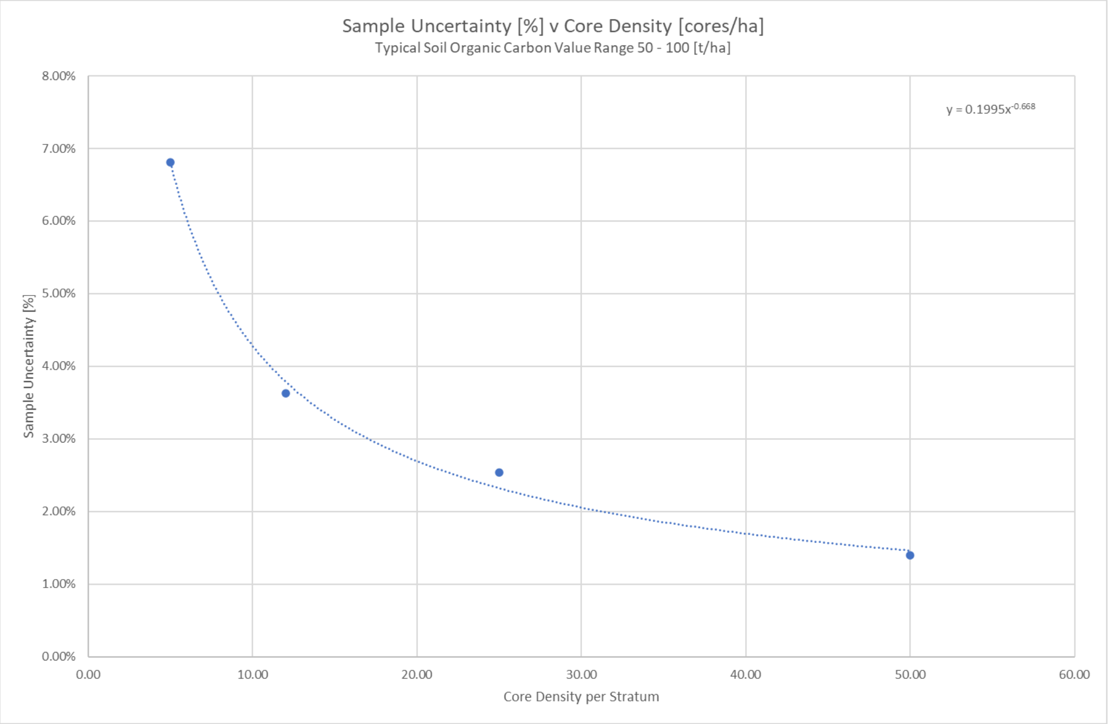

Figure 1 Sample Uncertainty with Core/Sample Density Change

Ecometric's methodology also relies on a minimum number of samples to train the network and the required minimum number of samples will also limit the minimum sample density.

3.5.3. 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.

3.5.4. SOIL CORE EXTRACTION

To maintain the integrity of the results the Regen Network soil sampling guide (G Booman *et al.*, 2021) will be the main reference. The following method will be used to collect soil samples:

1. This proposal will use a core depth of 30 cm.
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. The sampling depth will be consistent between all sampling rounds.
4. Each core will be georeferenced using a GNSS device with an accuracy of 4 metres or better.
5. Samples will be taken at least 10 meters away from any tree, structure, or body of water.
6. The date/time of each core will be recorded for each sampling round.
7. All sampling rounds will occur at least 6 months after the application of non-synthetic fertilizer.

3.5.5. SAMPLE ANALYSIS

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

$$SOC = Core\ depth * Bulk\ Density * SOC \tag{2}$$

Table 1 in Regen Network soil sampling guide (G 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 Loss on Ignition (LoI) only used in geographies where DUMAS is not available. Any future new methods of analysis that improve on the accuracy of current methods will be adopted. There shall be a requirement that the same analysis type is continued throughout the Project Crediting Period to ensure comparability, if possible, using the same laboratory.

3.5.6. IMAGERY SOURCE

Twelve bands of the Sentinel-2[1] multispectral image data (Table 1) will be used as the input for ML training; band 10 will not be used as it is only used for cirrus detection. The data will be downloaded from EOS Landviewer[2]. The download will include the respective metadata. Higher resolution data may be used in future projects.

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

3.5.7. IMAGE AND ANCILLIARY DATA PROCESSING

Imagery with a sensing date as close as possible to the sampling date but within +/- 4 months of the sampling date will be used and will be Cloud and defect free in the project area. Sentinel-2 image processing will include:

1. Processed to Level 2-A to provide BOA reflectance values.
2. Coordinate conversion as required (British National Grid for UK projects); Sentinel-2 tiles are ortho-images natively in UTM/WGS84 projection.
3. Resampling imagery and ancillary data to the same resolution (normally 10m).
4. Band stacking to create a 12-band multispectral image.

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

[2] https://eos.com/products/landviewer/

5. If multiple images are available the nearest, by date, cloud free image will be used.

3.5.8. NEURAL NETWORK TRAINING DATA

The QGIS[3] Point Sampling tool (or analogous tool) will be used to extract multispectral reflectance data at each soil sampling location. This data will then be paired with its respective SOC values to create a training dataset for the neural network.

3.5.9. NEURAL NETWORK TRAINING

The neural network will be trained using a set of design parameters that will remain fixed for each sampling project. The training may be run multiple times to determine the optimum parameters. Multiple training sessions will also be used to smooth the effects of the stochastic nature of neural networks.

3.5.10. ESTIMATE SOC USING A TRAINED NEURAL NETWORK

The trained network and Sentinel-2 imagery of the required project area will be used to estimate SOC at sampled and unsampled locations. The estimated SOC [t/ha] will be exported as a raster map(s).

3.6. SOIL SAMPLE ANALYSIS UNCERTAINTY

The declared uncertainty of the laboratory soil sample analysis will be propagated though any calculations made from this data.

3.7. NETWORK PREDICTION UNCERTAINTY

The network prediction uncertainty will be quantified using the mean absolute percentage error (MAPE), where number of values (n), soil sample value (A_t), network predicted value (F_t):

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \quad (3)$$

The MAPE will be calculated by overlaying the original soil sample positions (A_t) on the raster map (F_t) generated by the ANN prediction process. A proportion of the training dataset will be withheld for validation and test data. The standard fractions for validation and testing data are 15% [12]. Validation data may not be required for certain network architectures. The network inherits any uncertainty in the sample data during training as such the uncertainty in the sample data and the MAPE will define the final uncertainty in the predicted results.

3.7.1. CALCULATING SOC STOCKS

For areas where there are network SOC predictions the total SOC stocks [t] will be calculated by using the sum of the pixel values of the output image from the network.

3.7.2. CONVERTING SOC STOCKS TO CO_2 EQUIVALENTS

Converting soil organic carbon stocks to 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 [Da] of carbon dioxide (44) to that of carbon (12):

$$CO_2e[t] = SOC[t] \times \left(\frac{44}{12}\right) \quad (4)$$

[3] https://qgis.org/en/site/

3.7.3. WORKFLOW AUTOMATION

Automating the workflows within the methodology reduces the amount of time and work needed to complete a monitoring round. Automation also provides method integrity between monitoring dates and projects. The workflows that have been automated are image processing, SOC estimation using ML and the calculation of CO_2 equivalent. The automated workflows use QGIS modeler and a custom ANN App. The automation will also include detection of outliers in data to reduce their effect on the results of any process.

4. CALCULATING THE CREDITABLE CARBON CHANGE

4.1. BASELINE DEFINITION

The baseline SOC stocks or CO2e are defined here as the total carbon stocks calculated for the project's Initial Monitoring Date (IMD), or date of the first sampling round. The methodology will use a static baseline for each project, calculated as the total SOC stocks[t], from the IMD. All sampling rounds after the IMD will be compared to the baseline to calculate creditable carbon change. Parameters that can affect the SOC stock at baseline measurement will be taken into account in further sampling rounds.

4.2. CHANGES IN CO_2E BETWEEN REPORTING PERIODS

The change in SOC stocks between reporting periods is estimated as the difference between the total SOC stocks ($tSOC_{(t+1)}$) from the second monitoring period, minus total SOC stocks ($tSOC_{(t)}$) from the previous period:

$$SOC\ Stock = tSOC_{(t+1)} - tSOC_{(t)} \tag{5}$$

The same applies for estimating the change in the total SOC converted into CO_2 equivalents (CO_2e) between two sampling periods:

$$CO_2change = CO_2e_{(t+1)} - CO_2e_t \tag{6}$$

4.3. NET CO_2E REDUCTION

The net CO_2e reduction in the project area for a given reporting period is calculated as the difference between the changes in SOC, expressed as metric tons of CO2e, minus the total Greenhouse Gas Emissions (GHG) from all food and fibre production within the geographic boundaries of the Project Area emitted during the monitoring period, also in CO2e units, calculated using an IPCC higher tier (tier II or III) method of GHG reporting:

$$NET\ CO_2eReduction = CO_2e\ Change - E_{Farm\ Enterprise} \tag{7}$$

4.4. UNCERTAINTY AND DEDUCTIONS

Under this methodology framework, the total uncertainty for the project will be the sum of the uncertainties calculated throughout the methodology during a given monitoring period. Sources of uncertainties for creditable carbon stock calculations may include:

1. Laboratory percent soil organic carbon measurements.
2. Laboratory bulk density measurements.
3. Soil sample depth measurements.
4. ANN prediction uncertainty.

Acceptable methods for calculating uncertainty for percent soil organic carbon and bulk density will be in line with Regen Network Supplement S.2 [13]. If the uncertainty (U) for the reporting period is less than or equal to 20%, the net CO2 reduction value will be used without making any deductions to account for uncertainty (Uncertainty Deduction (UD) = 0). Table 2will be used to calculate the amount of uncertainty to deduct from the creditable carbon stocks. UD values are based on the Gold Standard LUF activity requirements Version 1.2.126 (Gold Standard, 2020).

Table 2 Ranges of uncertainties and the corresponding discounts

UNCERTAINTY (U)	Uncertainty Deduction (UD) (% of U)
U≤10%	-No Deduction-
10%<U≤20%	5-10% of U

The maximum uncertainty allowed for any measurement in a project is 20%. The Creditable Carbon Change after UD will then be estimated as:

$$Creditable\ Carbon\ Change = (Net\ CO_2e\ Reduction) \times (1 - UD) \qquad (8)$$

5. DATA REPORTING

5.1. 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 saved.

5.1.1. MONITORING REPORT

The monitoring report will include:

1. Sampling stratification design. This states the average strata size and core 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.
4. GNSS system and coordinate reference system used in the monitoring round.
5. Sampling date.
6. Selected Laboratory, laboratory accreditation, laboratory tests used.
7. AI training method summary.
8. Method used to quantify AI accuracy and the mean quantification accuracy achieved during the monitoring round.
9. The contact information for the Independent Contractor used to gather GHG Emissions data and calculate total Project Area GHG Emissions.
10. AI SOC Results, numerical and map form. Field level results to include:
    - Field area [ha].
    - Monitoring interval crop type.
    - Net mean SOCS [t/ha] after the deduction of AI error [MAPE].

- For all monitoring rounds after baseline, tabulated results will be listed for both the previous and current monitoring rounds to allow direct comparison.
- Field level results are summarised by crop type and totalled to complete the table.

11. Soil sample results, numerical and map form. Tabulated results reported by field to include:
    - Crop type.
    - Sampled field area.
    - SOCS total [t/ha].
    - Stone content % where applicable.
    - SOCS field total [t]. Field level results are summarised by crop type and totalled for the sampled area to complete the table.
12. SOCS change between monitoring rounds [tSOC] and [tCO_2e].
13. Baseline / Current year GHG Emissions report references to link with detailed GHG Emissions Reports.
14. GHG Emissions change between monitoring rounds [tCO_2e].
15. Yield related leakage reported against 5-year average crop yields. Any crop specific yield reduction of >10% from 5-year average to be justified against regional crop specific monitoring-year yield averages to differentiate between potential leakage and regional performance trends.
16. Carbon balance [tCO_2e] calculated by deducting monitoring period GHG Emissions [tCO_2e] from monitoring period SOCS gains [tCO_2e].
17. Credit statement. Positive carbon balance CO_2e tonnes are allocated Credits [1 Credit per tCO_2e]. The credit balance sub divides the allocated credit total into a buffer pool [20% of total] and credits pending issuance [80% of total].

5.1.2. PUBLICLY REPORTED DATA

The following supporting data will be publicly displayed on the Regen Registry:

Table 3 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

5.1.3. NON-PUBLIC DATA

The following 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 4 Non-Public Data

Folder	Sub Folder	Contents	File Type	Number of Files
AI Data	Image	Remotely Sensed Images	.tiff	As required
	Network Training Data	Training Data	.csv	As required
	Trained Network	Network Settings	.mat	1
	Network Results	Sampled Area Network Result	.csv	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	.docx	1 (may be omitted if duplicate of pdf)
		Monitoring Report	.pdf	1
Sampling Results		Combined Sampling Plan Results	.xlsx	1
		Individual Laboratory reports	.csv	Variable, by project size and laboratory reporting interval

6. 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.

7. 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

[1] U. Stockmann *et al.*, “The knowns, known unknowns and unknowns of sequestration of soil organic carbon,” *Agriculture, Ecosystems & Environment*, vol. 164, pp. 80–99, Jan. 2013, doi: 10.1016/j.agee.2012.10.001.

[2] E. A. Thaler, I. J. Larsen, and Q. Yu, “A New Index for Remote Sensing of Soil Organic Carbon Based Solely on Visible Wavelengths,” *Soil Science Society of America Journal*, vol. 83, no. 5, pp. 1443–1450, 2019, doi: 10.2136/sssaj2018.09.0318.

[3] H. M. Bartholomeus, M. E. Schaepman, L. Kooistra, A. Stevens, W. B. Hoogmoed, and O. S. P. Spaargaren, “Spectral reflectance based indices for soil organic carbon quantification,” *Geoderma*, vol. 145, no. 1–2, pp. 28–36, May 2008, doi: 10.1016/J.GEODERMA.2008.01.010.

[4] S. M. M. Rasel, T. A. Groen, Y. A. Hussin, and I. J. Diti, “Proxies for soil organic carbon derived from remote sensing,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 59, pp. 157–166, Jul. 2017, doi: 10.1016/J.JAG.2017.03.004.

[5] L. Gardin, M. Chiesi, L. Fibbi, and F. Maselli, “Mapping soil organic carbon in Tuscany through the statistical combination of ground observations with ancillary and remote sensing data,” *Geoderma*, vol. 404, p. 115386, Dec. 2021, doi: 10.1016/J.GEODERMA.2021.115386.

[6] L. Guo *et al.*, “Mapping soil organic carbon stock by hyperspectral and time-series multispectral remote sensing images in low-relief agricultural areas,” *Geoderma*, vol. 398, p. 115118, Sep. 2021, doi: 10.1016/J.GEODERMA.2021.115118.

[7] O. Odebiri, J. Odindi, and O. Mutanga, “Basic and deep learning models in remote sensing of soil organic carbon estimation: A brief review,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 102, no. June, p. 102389, 2021, doi: 10.1016/j.jag.2021.102389.

[8] T. Angelopoulou, N. Tziolas, A. Balafoutis, G. Zalidis, and D. Bochtis, “Remote sensing techniques for soil organic carbon estimation: A review,” *Remote Sensing*, vol. 11, no. 6, pp. 1–18, 2019, doi: 10.3390/rs11060676.

[9] J. J. De Gruijter, B. Minasny, and A. B. Mcbratney, “Optimizing stratification and allocation for design-based estimation of spatial means using predictions with error,” *Journal of Survey Statistics and Methodology*, vol. 3, no. 1, pp. 19–42, 2015, doi: 10.1093/jssam/smu024.

[10] J. J. de Gruijter, A. B. McBratney, B. Minasny, I. Wheeler, B. P. Malone, and U. Stockmann, “Farm-scale soil carbon auditing,” *Geoderma*, vol. 265, pp. 120–130, 2016, doi: 10.1016/j.geoderma.2015.11.010.

[11] G. Booman, A. Craelius, B. Deriemaeker, G. Landua, W. Szal, and B. Weinberg, “Regen Network Whitepaper,” 2021.

[12] M. A. Shahin, ; Holger, R. Maier, and M. B. Jaksa, “Data Division for Developing Neural Networks Applied to Geotechnical Engineering”, doi: 10.1061/ASCE0887-3801200418:2105.

[13] G. Booman, S. Bennetts, S. Leiker, and R. Steinherz, “Regen Methodology for GHG and Co-Benefits in Grazing Systems,” *Regen*, no. Version 0.91, pp. 1–42, 2021.

[14] Gold Standard, “Land Use & Forests Activity Requirements,” *Gold Standard*, vol. 1.2.1, no. October 2019, pp. 1–44, 2020.