Development Carbon Monitoring Protocol

A testcase for three farms in the Netherlands

Y. Fujita
S.E. Verweij
G.H. Ros

Nutrient Management Institute BV
e-mail: nmi@nmi-agro.nl
website: www.nmi-agro.nl

Reference

Fujita Y, Verweij SE & GH Ros (2021). *Development Carbon Monitoring Protocol. A testcase for three farms in the Netherlands*, Nutrient Management Institute BV, Wageningen, Rapport 1831.N.21, 34 pp.

Summary

There is growing interest in sustainable soil management practices increasing organic carbon content and stocks in arable soils in order to mitigate climate change, to enhance soil fertility and improve the agronomic resilience to climate change. Given the need for credible and reliable measurement, monitoring, reporting and verification we developed a robust sampling protocol (cheap, affordable, reliable) for evaluation of C stocks on farm level. The protocol is tested and evaluated for three farms in the Netherlands with varying site properties and land management history. We showed that the proposed protocol has the potential to be applied across the Netherlands (and even outside the Netherlands given the use of open source available satellite data and soil databases). Soil C stocks on farm level can be quantified accurately with a maximum deviation less than 5%.

Distribution

ESTI	digital

Table of contents

Summary and Conclusions

There is growing interest in sustainable soil management practices increasing organic carbon content and stocks in arable soils in order to mitigate climate change, to enhance soil fertility and improve the agronomic resilience to climate change. Given the need for credible and reliable measurement, monitoring, reporting and verification we developed a robust sampling protocol (cheap, affordable, reliable) for evaluation of C stocks on farm level. The protocol is tested and evaluated for three farms in the Netherlands with varying site properties and land management history.

We showed that the proposed protocol has the potential to be applied across the Netherlands (and even outside the Netherlands given the use of open source available satellite data and soil databases). Soil C stocks on farm level can be quantified accurately with a maximum deviation less than 5% and a confidence interval of 95% by a sampling density of 1 sample per 2 ha. This means that, if one repeats the same sampling procedure 100 times, the estimated farm C stock falls between -5% to +5% of mean for more than 95 times. By doubling the sampling density, the deviation decreases till 3%.

From this analysis we can conclude that:

- required sampling densities decline with the number of farms sampled. This improvement might be even bigger when smart machine learning algorithms can be applied. Adding weather related covariables or machine sensing data derived from machinery of the farmer might be another option.
- required sampling densities also decline when the accuracy of the Handheld scanner increases. For the current study we used a relative mean error of about 30% given a prior calibration on a global database (and this is likely a worst-case scenario since NIR sensors are known for their high reproducibility when used on the same location).
- the use of a priori knowledge of fields (derived from either earlier scans, results from agricultural laboratories or spatial models derived from these lab analyses) enhance the accuracy of the C stock estimates and reduces the number of additional samples needed.
- a priori knowledge of fields derived from results from agricultural laboratories alone can be used to estimate the C stock with a deviation between 10 á 20% but only when all the fields are sampled (and field size do not increase the 5 hectare; the common approach of Eurofins Agro includes a mixed sample of 40 subsamples taken from the field). Historical analyses cannot be used as a surrogate for spatial variation inside fields.
- the satellite derived maps of BioScope show a positive relationship with soil organic C levels within a field, but their functionality overlaps with the method used in this study and their applicability is currently limited due to the fact that grasslands and field boundaries are excluded and the underlying algorithm is unsupervised.

1 Introduction

There is growing interest in sustainable soil management practices increasing organic carbon content and stocks in arable soils in order to mitigate climate change, to enhance soil fertility and improve the agronomic resilience to climate change (Smith et al., 2019). Since the soil organic carbon (SOC) content cannot be easily measured with high precision, there is the need for credible and reliable measurement, monitoring, reporting and verification, both for national reporting and for emissions trading.

Direct measurements of C stocks on farm level rely on appropriate study designs and sampling protocols dealing with high spatial variability in SOC levels across fields. A large number of soil samples is often required. Sufficient sampling depth is a crucial factor for properly evaluating changes in SOC (IPCC recommends a depth of 30 cm) since SOC impacts may occur over the complete rooting zone along the soil profile. Since costs associated with collecting, processing and storing soil samples and SOC measurements, there is a need to evaluate these costs against the (accuracy) of estimated carbon gain. A combination of direct measurements (at the plot scale) and modelling (at larger spatial scales) can greatly help defining the efficacy of land management practices in enhancing SOC.

In behalf of and in collaboration with *Ecosystem Services Trade Initiative* we aim to develop a cheap, fast and robust sampling methodology for determination of C stock baseline using smart combination of historical soil data, satellite data and sensor derived soil analyses. With proper *a priori* knowledge of the spatial variation within and among fields it is possible to design a sampling protocol to reach the desired accuracy: to estimate the C stock on farm level with a maximum deviation less than 5% and a confidence interval of 95%. This means that, if one repeats the same sampling procedure 100 times, the estimated farm C stock falls between -5% to +5% of mean for more than 95 times.

2 Materials and Method

2.1 Sample locations

The development and evaluation of the protocol for C monitoring on farm level has been tested on three farms located in the Netherlands, located in Texel, Friesland and Brouwershaven (Figure 2.1). The arable farm in Brouwershaven is located on soils varying from loamy sand / sandy loam to light clay texture. Most of the fields are minimum cultivated and the crop rotation plan includes crops like cereals, potatoes, onion and chicory. The farm in Texel is located on sandy soils and are managed with a strong focus on soil organic matter build-up in the last decade due to high decomposition rates. Soils are cultivated for potatoes, sugar beets, grass seed, corn and cereals. The farm in Friesland is managed by a dairy farmer, is mainly used for grassland and is characterized by clayey and loamy soils, sometimes with peat being present in the top soil profile.

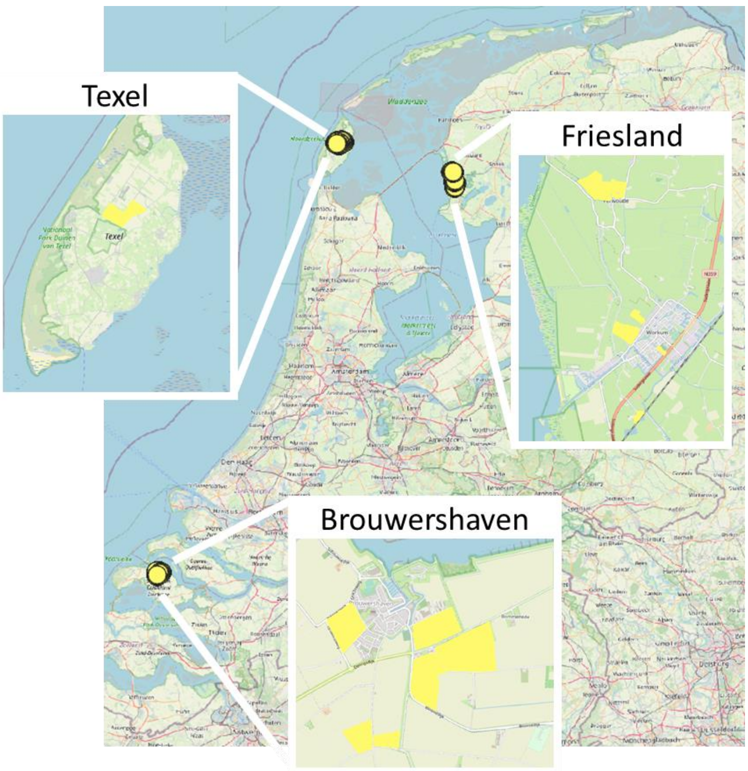


Figure 2-1. Location of the three farms used for development and testing of a C monitoring protocol.

2.2 Sampling scheme design

2.2.1 Covariables for (potential) sampling points

For the sampling scheme for each farm, the fields are divided in a grid of 10 meter resolution. Each of those units is a potential sampling location. See Appendix I for example maps of these potential sampling points. For all these potential locations covariates have been collected that might be related to the spatial variation in soil C levels, either from private organisations or from open source available data sources.

For each potential sampling point, a series of covariables were collected. These covariables were selected given our earlier evidence from developed machine learning models reflecting spatial variation in soil organic matter levels across the Netherlands. Covariables are stored in a form of either vector (v) or raster (r) data. See full list in Table 1 of Appendix II.

The private and open source (soil) data for the Netherlands include:

- Eurofins' soil measurement data on parcel level [v]
- NMI database including predictions for all kind of agronomic relevant soil properties, including mineralogy, organic matter, pH and soil nutrients (all derived from geostatistical models built on existing databases of soil analyses and land use, being input for the Open Soil Index) [v]
- Soil map 1:50,000 (Soil type, groundwater table) [v]
- Fraction of Absorbed Photosynthetically Active Radiation (FAPAR) [r]
- BRP parcel crop rotation data of 2009-2019 [v]
- AHN [r]
- RIVM earthworm database [r]
- Sentinel Mosaic data (band 2, 3, 4, 5, 6, 7, 8A, 11, 12) [r]

All satellite derived information is globally available. Similar to the NMI-database one might make use of ISRIC soil grids for other countries outside the Netherlands to obtain prior knowledge on soil properties (although the accuracy of the soil properties varies among countries).

Prior to the analysis for each farm, covariables were excluded when:

- the variable is available for less than 80% of the potential sampling points,
- the variable is strongly correlated with others (i.e. rank correlation coefficient rho > 0.95),
- the value of the variable is identical for more than 95% of the potential sampling points.

Subsequently, missing values were imputed with the median values of the variable. These missing values occur in only in a few cases and imputation is needed to avoid the removal of valuable covariates due to single missing data points. It has little impact on the stratification algorithm used.

2.2.2 Stratified sampling method cLHS

For selection of sampling points for analysis, we used the conditioned Latin hypercube sampling (cLHS) algorithm. The stratified sampling algorithm with cLHS works efficiently when the covariables can capture the spatial variations of the target variable (in this case soil organic carbon, SOC). Therefore, we made a preselection of the covariables that are related to variation in soil carbon. In other words, the preselection of the covariables is meant to avoid that the selected sampling points reflect a gradient of covariables which has little impacts on SOC.

We selected the top 20 numerical covariables which have a high correlation coefficient (Spearman's rank correlation) with the organic matter content data of Eurofins, measured on field level. In addition we also added two general categorical variables (soil texture and groundwater level) as well as organic matter content on field level. In total, 23 covariables were selected.

Sampling points were subsequently selected with a stratified sampling technique cLHS. With cLHS a subset of strata are selected using a stratified random procedure based on the multivariate distribution of the covariables (Minasny & McBratney, 2006). This will ensure that strata that are closely related to each other will not be measured both, making the sampling efficient to estimate a spatial trend. For each farm, 100 points were selected with cLHS (see maps in Appendix I). The selected sampling points by cLHS successfully mimic the range of covariables of the entire potential sampling points (Figure 2.2). This is a requirement for future extrapolations.

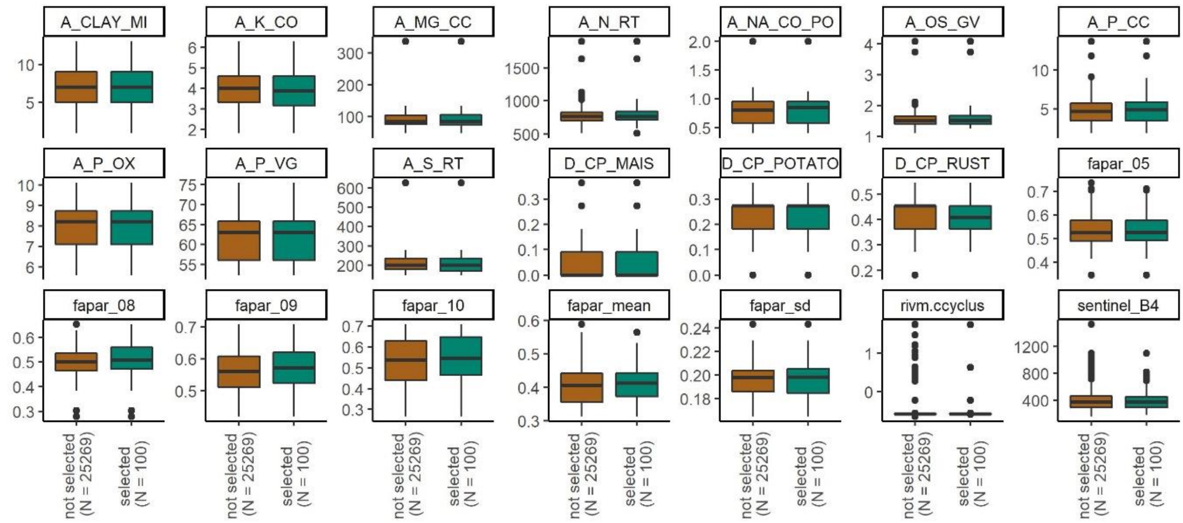


Figure 2-2. Difference in distribution of 21 numeric covariables (including organic matter on field level) between selected (n = 100) and not-selected points (n = 25269) in the farm of Texel. The box represents the 1st and 3rd quantile of the covariables.

The selected points also cover a wide range of multidimensional covariable space, as visualized on the two major axes of PCA (Principal Component Analysis, Figure 2-3, left). PCA extracts the major axes of variation of multiple variables, and therewith helps to reduce dimension of multiple variables. For a comparison purpose, we also made another subset of 100 points with random sampling.

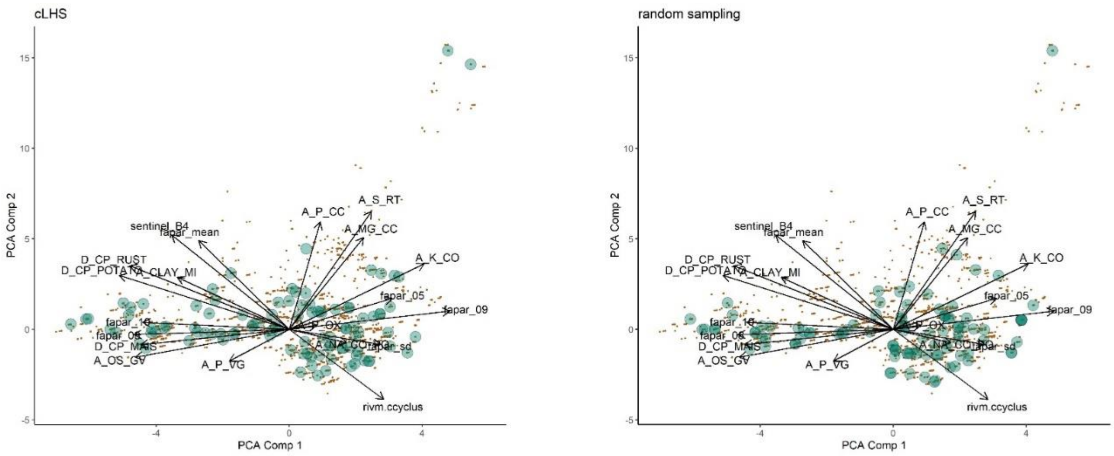


Figure 2-3. Distribution of 100 selected samples (green points) on PCA biplot, compared to the non-selected potential sampling points (brown points) of the farm of Texel. The 100 points were selected with cLHS (left) and random sampling (right). Arrows depict the loadings of covariables.

The randomly selected 100 samples cover the PCA plane less effectively compared to the samples selected by cLHS: it misses some ranges, and it overlaps more often in the ranges where potential sampling points are densely located (Figure 2-3, right). This indicates that our stratified sampling with cLHS could effectively select the sampling points which represent the variation of covariables related to soil carbon.

2.3 Field sampling protocol

For each farm, soil samples were collected at 100 locations and scanned with the AgroCares handheld scanner. Depth of the sample was between 0 and 30 cm. Among these 100 locations, 25 locations were randomly selected for more detailed lab analysis. At these locations, the scanned soil was collected and analysed for wet chemistry variables at the AgroCares Golden Standard Laboratory (GSL).

In the Friesland farm, 10 locations were randomly selected from the 100 locations to test the difference in soil C at two depths. These 10 locations do not overlap with the GSL sampling locations. At each location, the soil was first scanned for 0-10 cm and 10-30 cm depth separately. Subsequently, the soil was mixed well and scanned again for the entire depth (i.e. 0-30 cm).

During field sampling, three locations were not suited for a proper sampling. Therefore, remaining 297 samples of handheld scanner are used for the following analysis.

The sampling protocol is further described in Appendix III.

2.4 Sampling scheme optimization

The aim of this study is to develop a cheap and robust sampling strategy. To that aim, it was tested if the sampling effort can be reduced without large expense of accuracy. The sampling scheme was evaluated by randomly selecting subsets of the soil C measurement (selected with cLHS), and calculating the farm-level carbon stock with each of the subsets. By repeating the random sub-setting for many times, representative estimates of farm-level C stock for different degree of reduced sampling can be obtained. For each iteration of the random sub-setting, soil C content of all grids was estimated with the algorithm trained for the subset, and with random measurement error of handheld scanner. The error of Handheld scanner is set to be 30% given a prior calibration on a global database (and this is likely a worst-case scenario since NIR sensors are known for their high precision when used on the same location). In this way, for each number of sampling points, mean and standard deviation of farm-level C can be computed.

In addition, we experimentally tested two different ways to reduce sampling number. One is to include measurements of previous field campaigns of other farms (assuming that these farms have been analysed prior the visit to the selected farm), and another is to use priori knowledge on soil properties from either Eurofins Agro, NMI or any other organisation. Here also all relevant input variables for the Open Soil index are included (see 2.2.1). Impacts of both methods on reducing the number of sampling points were evaluated with a full-factorial design, and variation of estimated farm-level carbon stock were calculated for all combinations.

2.5 Carbon stock calculation

Soil carbon concentration of all 10x10m grids within the farms was predicted based on the measurements of 297 locations, using a Partial Least Square (PLS) regression model. This approach is chosen since it can handle multi-collinearity well and it doesn't require huge sample numbers like Locally Weighted Regression or machine learning algorithms such as XGBoost. Locally weighted regression is a way of estimating a regression surface through a multivariate smoothing procedure, fitting a function of the independent variables locally and in a moving fashion analogous to how a moving average is computed for a time series. XGBoost is a popular and efficient open-source implementation of the gradient boosted trees algorithm. Gradient boosting is a supervised learning algorithm, which attempts to accurately predict a target variable by combining the estimates of a set of simpler, weaker models. In contrast to PLS, the latter two approaches will be valuable as soon as more than 1000 locations have been analysed.

In total, 73 covariables were used as explanatory variables in the PLS model used for upscaling using the selection criteria mentioned in section 2.2.1 (see Table 1 of Appendix II). The number of PLS components used for the prediction was minimized given the RMSE of the cross-validation.

Subsequently, soil carbon concentration of top 30 cm (g C kg^{-1}) was converted to a carbon stock (g C $100m^{-2}$) by multiplying the soil C content (g C kg^{-1}) by the bulk density (kg m^{-3}), the depth of the soil (m) and the area of the grid cell (100 m^2).

The bulk density was estimated from soil organic matter content and clay content based on a well-known empirical relationship built on a large dataset of Dutch agricultural soils, as being used in agronomic research in the Netherlands (CBAV, 2019) as well as the Open Soil Index. Soil organic matter content was estimated from soil C content, using the conversion factor of 0.58, being the most popular conversion algorithm in the Netherlands. Recently there was some scientific debate whether the value should be rather 0.5 rather than 0.58. Since this is only used for the calculation of soil bulk densities, the impact of this choice is not further evaluated[1].

The formula of bulk density estimation is as follows:

$$BD = cf \cdot BD_{clay} + (1 - cf) \cdot BD_{sand}$$

$$cf = \min\left(1, \frac{Clay}{25}\right)$$

$$BD_{clay} = 1000 \cdot (a_1 \cdot OS^4 + a_2 \cdot OS^3 + a_3 \cdot OS^2 + a_4 \cdot OS + a_5)$$

$$BD_{sand} = 1000 \cdot \frac{1}{b_1 \cdot OS + b_2}$$

$$OS = \frac{1}{0.58} \cdot soilC \cdot 10^{-3} \cdot 10^2$$

where *BD* is the bulk density (kg m^{-3}), *OS* is the soil organic matter (%), *soilC* is the soil C content (g C kg^{-1}), a_i and b_i are the empirically derived coefficient values, derived from fitting a polynomial equation to the data presented in the Fertilizer Recommendation Manual (a_1 = 0.00000067, a_2 = - 0.00007792, a_3 = 0.00314712, a_4 = - 0.06039523, a_5 = 1.33932206, b_1 = 0.02525, b_2 = 0.6541).

Finally, farm C stock was calculated as sum of soil C of all grids located within the fields of the farm.

[1] Since the current analysis strongly focuses on the required accuracy of the farm C stock, the spatial variation in soil organic matter levels receives less attention. Since the actual certification is done on changes in C levels, the conversion from SOC to SOM is only supporting for the communication to farmers. For future use of this protocol it is recommended to add estimated SOM levels besides the C stocks calculated. We also recommend to use the more generic conversion factor of 2 when converting from SOC to SOM.

2.6 Handheld correction with GSL

When farmers or soil testing laboratories take a soil sample from a field they mostly aim at one sample representing the average status of that field. They take several soil cores (40 per 2-5 ha) and mix this to one sample, which is then used for soil testing at the laboratory. The laboratory reports back the soil test value of this mixed sample. The current approach in this study follows the use of the Handheld scanner of AgroCares, a handheld device using a NIR sensor to determine (changes in) soil organic C levels in soil. Diffuse reflectance near infrared (NIR) spectroscopy has extensively been applied for analyses in very diverse fields including, agriculture, geology and soil science. In soil science, numerous studies have demonstrated that the NIR spectral range combined with a multivariate calibration method (or any machine or deep learning calibration) could be used as a non-destructive rapid analytical technique to simultaneously estimate several soil properties (including soil organic carbon) in a very short time. Using these NIR sensors it is possible to take a lot more soil samples within a field for low costs. When several soil samples are taken and measured, this may lead to different soil testing values because of spatial heterogeneity within the field. In addition, while wet chemistry is known for its analytical error (high precision and low reproducibility), the opposite is true for NIR sensors: assuming a well-designed calibration and validation procedure, the reproducibility of NIR measurements is substantially higher although the precision is often lower. This means that much more samples have to be analysed with NIR to achieve the same precision, but at the same time is it much easier to detect changes over time with NIR due to its high reproducibility.

For the quantification of the actual C stock, this study strongly depends on the soil organic carbon levels determined by the Handheld scanner. To correct for possible bias, one out every four samples is randomly assigned to be also measured for wet chemistry analysis of soil organic carbon. Using these wet chemistry data as a check of the “true C content”, a linear regression model was fitted between both methods to be used for normalization of the Handheld scanner thereby combining the power of NIR sensors (high reproducibility) with wet chemistry (high precision).

2.7 Evaluation BioScope maps

For two farms satellite derived stratification maps were prepared by Bioscope and evaluated for their potential as covariable in future sampling schemes. Since the underlying algorithm of these maps is unsupervised, each field is separated into five subsections that show spatial variation in bare soil reflectance during winter. The method as such is not (yet) applicable for fields with crops growing, reducing its potential use for grassland and fields cultivated with catch crops during harvest and winter. Nevertheless, for the two arable farms included, stratification maps were evaluated given the variation in estimated soil organic carbon levels. When the within strata variation is less than the variation across strata, then the Bioscope maps have potential value as covariable optimizing the sampling protocol.

3 Results

3.1 Maps of carbon content

The spatial variation in soil C levels across the farm in Texel are visualised in Figure 3.1. The soil C levels vary from 6 to 14 g C kg^{-1} and is on average 9.3 g C kg^{-1}. There is no clear distribution of soil C levels over the fields though fields in the southeast might be slightly higher than the surrounding fields. A limited number of the fields has values below 8 g C kg^{-1}.

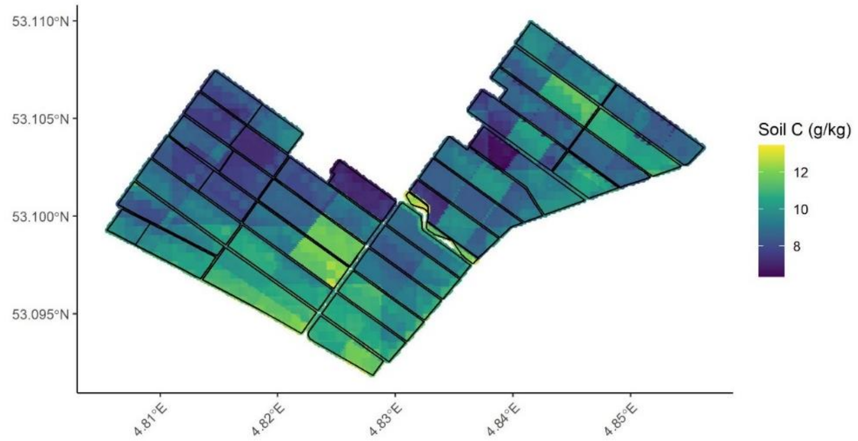


Figure 3-1. SOC of the farm in Texel. Black lines are the boundary of fields.

For Brouwershaven, SOC levels vary from 8 to 19 g C kg^{-1} and is on average 13.5 g C kg^{-1}. There is a clear west east pattern visible with higher levels in the western part of the farm.

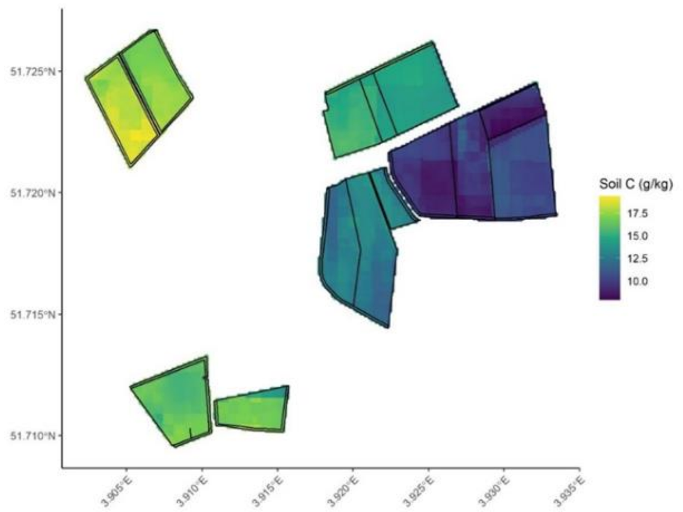


Figure 3-2. Soil C levels of the farm in Brouwershaven. Black lines are the boundary of fields.

For the farm in Friesland, the SOC levels are much higher than those in the other two farms. They vary from 22 to 32 g C kg^{-1} and the mean SOC level is about 28 g C kg^{-1}.

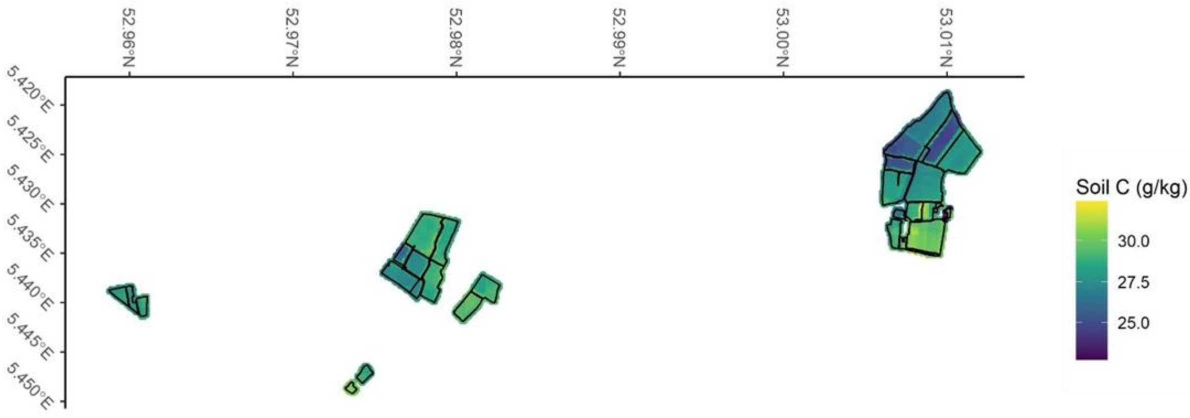


Figure 3-3. Soil C levels of the farm in Friesland. Black lines are the boundary of fields.

The summary statistics of the C levels within and among the fields of the three farms are summarized in table 3-1. Largest variation is observed for the farm at Brouwershaven, whereas the spatial variation is quite similar among the two other farms although the absolute levels strongly differs.

Table 3-1. Summary statistics of predicted soil C (g kg^{-1}) for the 3 farms. Mean, standard deviation, max, minimum, 95th percentile, 5th percentile, and number of potential sampling location (10x10m grids) are shown.

farm	mean	sd	max	min	q95	q05	n
Brouwershaven	13.5	3.0	19.4	7.9	18.0	8.8	10312
Friesland	27.9	1.5	32.4	22.7	30.1	25.0	6200
Texel	9.3	1.2	13.5	6.3	11.6	7.3	25370

3.2 Farm-level carbon stock

Given the estimated C content on each potential sample location on the farm (each being 10x10m), and the estimated bulk density, we estimated the C stock on farm level. The total C stock for each farm is shown in table 3.2. For the three farms evaluated, the C stock varies from about 5 to 10 kton carbon in the topsoil of 30 cm depth. When expressed per area agricultural land, the total stock varies from 39 to 100 ton C ha^{-1}. As expected, the highest C stock was found in the farm in Friesland due to the presence of peat and grasslands, whereas the lowest C stock was found in the sandy soils in Texel.

Table 3-2. Total C levels for the three farms evaluated.

Farm	Area (ha)	Carbon stock (ton C)	(ton C ha^{-1})
Brouwershaven	103	5434	53
Friesland	61	6172	100
Texel	254	9811	39

3.3 Optimizing sampling protocol

To explore the minimum number of soil sampling needed for robust and reliable estimates of the C stock, the sampling scheme was evaluated. For the evaluation, we compare all combinations of two sources of uncertainty:

- A situation where there is no a priori knowledge on soil properties (either from Eurofins analysis or from NMI-databases). This means that only publicly-available information, such as soil map and satellite data, are used as covariables.
- A situation where the PLS model includes a priori knowledge of the relation between covariables and SOC levels, derived from other farms.

As expected, the variation in farm-level carbon stock estimates decreases as the number of samples increases for all scenarios (Figure 3.4). When both a priori knowledge is used (Figure 3.4 top left), the carbon stock estimate has the highest accuracy, compared to when one of the a priori knowledge is lacking. Biggest gain can be obtained from using the data from previous field campaigns (comparing left with right, Figure 3.4) indicating that the number of samples required to reach the desired accuracy will decline as the sampling campaign expands with more farms.

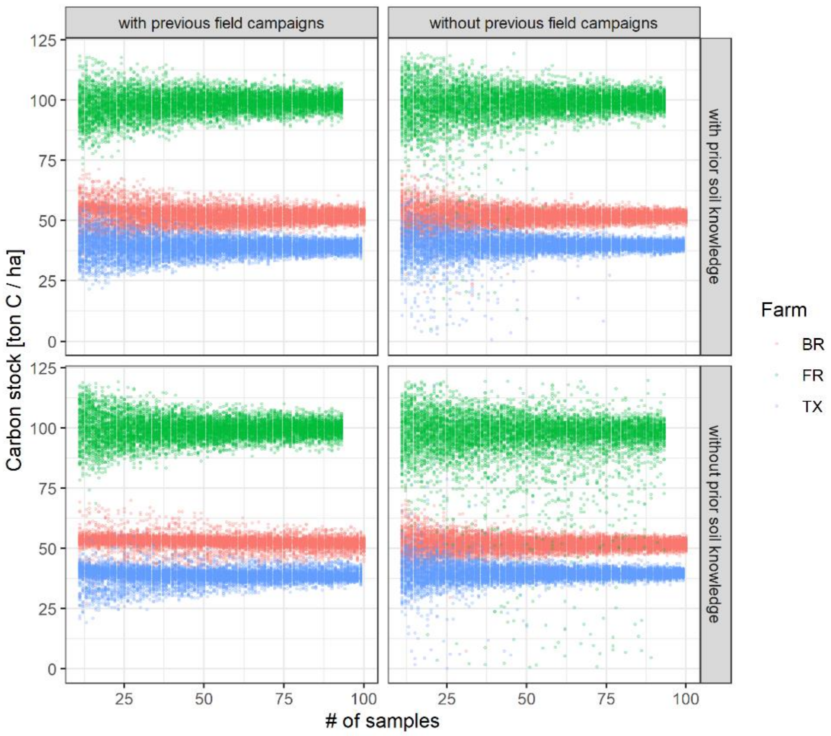


Figure 3-4. The calculated carbon stock for the three farms with (top) and without (bottom) prior soil knowledge, and with (left) and without (right) data from previous field campaigns. As the number of samples increases the variation in carbon stock decreases, with the lowest variation when the prior soil knowledge is used and the data from previous field campaigns.

Figure 3-5 presents the same results as Figure 3.4, but with variation of farm-level C stock on the y-axis and actual sampling density as x-axis. Here coefficient of variation (i.e. standard deviation divided by mean) is used to quantify the variation. It is clearly shown that the highest accuracy (and subsequent the lowest sampling density given a desired accuracy) can be found when both existing data from previous field campaigns as well as prior soil knowledge is used (Figure 3-5, top left). With this condition, the observed coefficient of variation declines from about 17% when only 1 sample per farm is analyzed down to about 3% variation when 1 sample per ha is taken. Using existing farm data substantially improved the accuracy (visible by comparing left and right figures from Figure 3-5).

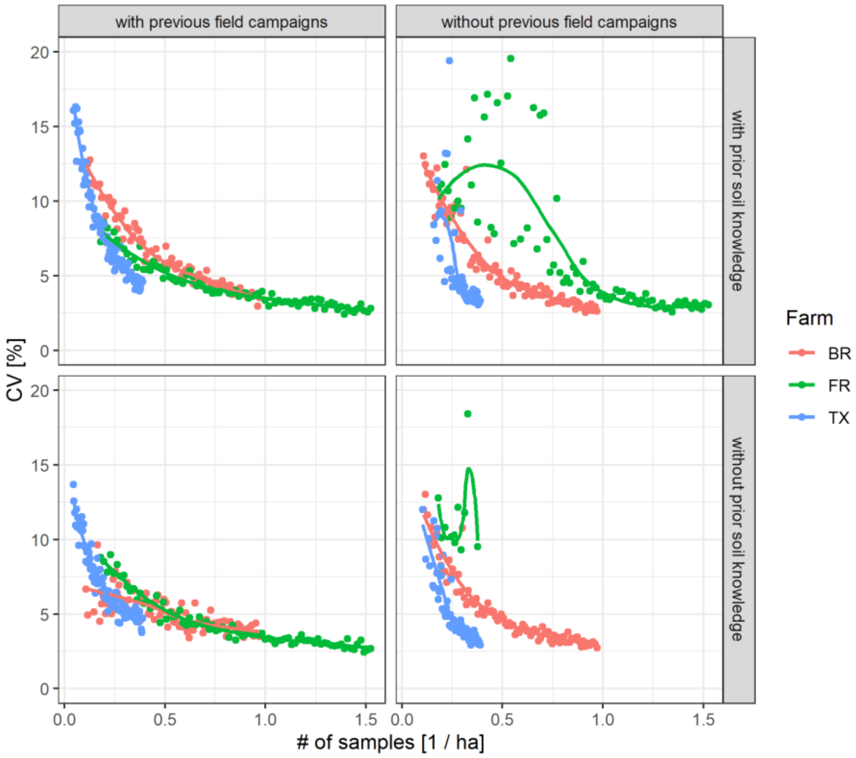


Figure 3-5. The Coefficient of Variation (CV, %) of the carbon stock for the different farms. The CV of the carbon stock that uses prior soil knowledge and data from the previous field campaigns reaches the threshold of 5% with the lowest density of samples. Part of the fitted green line as well as the green datapoints in the figure below-right are not visible since the CV exceeds the maximum value of the y-axis for most of the scenario's evaluated.

The variation in calculated carbon stock depends heavily on the numbers of samples and so on the sampling density. The sampling density required to reach a threshold of maximal 5% deviation can be lowered with the use of prior soil knowledge and data from previous field campaigns. With presence of both prior knowledge, an accuracy of 5% can be met for all farms with only one samples per two hectares.

3.4 Comparison with Eurofins Agro

The alternative approach by using Eurofins soil analyses (or NMI-databases when no data was available for a given field) to calculate the carbon stock of the farm shows that it is not possible to get a robust estimate of the C stock within the expected deviation less than 5% given a 95% confidence interval. About 30% of the actual C stock predictions in Brouwershaven, 22% of the predictions in Friesland and 13% of the predictions in Texel results in a C stock that deviates more than 5% of the "true mean".

This is visualised in Figure 3.6 where the variation on C stocks for the three farms is visualised given the variability on the NIR measurement in the laboratory, excluding all inaccuracies due to sampling (CV of 20%, being 10% less than the actual CV of the Handheld Scanner). The solid lines show the prescribed and acceptable confidence interval for the farm to estimate the C stock with a desired maximum deviation of 5%. As soon as the solid lines are in between the dashed lines, then the actual uncertainty (represented by the dashed lines) is bigger than the desired one.

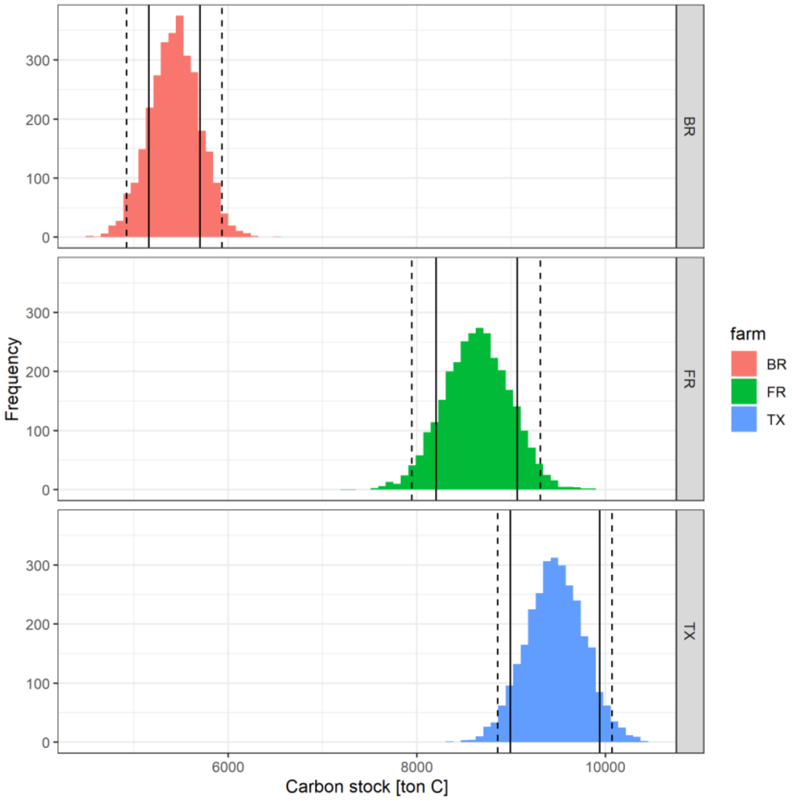


Figure 3-6. Carbon stock calculated when using Eurofins Agro soil analysis. The solid line shows the prescribed 95% confidence interval with deviation of 5%. The dashed line is actual 95% confidence interval for the farms.

Lowering the desired accuracy down to 10% allows the use of a prior soil analyses for quantification of C stocks. In other words, no additional samples are needed above the field averaged ones. Be aware, this is only valid as long as all fields of the farm have been analysed and the field size do not exceed the 5 hectares (since an identical error for fields with a bigger area has a much higher impact on the C stock on farm levels). In addition, given the actual carbon stock on the three farms studied, a desired accuracy of 10% equals to an acceptable deviation in calculated C stocks ranging from 500 to 1000 ton, being equal to 4 to 10 tonnes C per ha. These high numbers questions whether the use of a desired accuracy of 10% is sufficient to valorise sustainable soil management on the short and long term.

To illustrate this, we illustrate this for the ambition of the 4-promille initiative. Assuming an organic carbon content of about 1.2% for a clay or sandy soil results in a C stock of about 44 to 49 ton C per hectare[2]. An annual change of 4-promille (in absolute terms) results in an expected increase of about 2.6 to 3.0 tonnes C per ha over a period of 20 years. This increase corresponds to a relative change of about 6% for a sandy soil with an initial SOC level of 1.2%. The relative change in C stocks rapidly decline to about 2.5% when the initial SOC level increases up to about 2.6% SOC (corresponding to 5.2% SOM). At higher SOM levels one need an accuracy below the 2.5% to detect even the desired 4-promille change

[2] A sandy soil with 1.2% SOC has a density of 1362 kg m^{-3}. For a soil layer of 30 cm this results in a C stock of 1362 kg m^{-3} x 3000 m^3 ha^{-1} x 1.2% x 0.001 ton kg^{-1} = 49 ton C ha^{-1}. For a clay soil with the same amount of SOC the density is 1211 kg m^{-3} resulting in a C stock of 44 ton C ha^{-1}. When 20 years of sustainable soil management practices leads to an increase of 20 x 0.004 this results in a new C stock in the sandy soil of 1.28%. The soil density slightly decreases down to 1354 kg m^{-3} whereas the C stock increased up to 52.0 ton C ha^{-1}, and increase of 6.06%.

in SOC. However, the 4-promille promotes not an absolute annual increase of 4‰ but a relative increase. For a sandy soil this leads to a minimum detectable change of 4.8 to 7.7% (as percentage deviation of the original content) given a SOM content ranging from 20 to 0.8%. For a clay soil this leads to a minimum detectable change varying from 5.8 to 7.7%. Hence, a protocol for monitoring of changes in soil Carbon should have a desired accuracy far less than 10% on farm level. The methodology presented in this study aims therefore on a desired accuracy less than 5%. Given the relatively high soil organic matter levels in the Netherlands one should focus for the long term on a desired accuracy of about 2.5% (to be able to detect changes in SOC for arable soils with a soil organic matter level up to 6.4%).

To show the implications of the aforementioned high uncertainty of 10%, we estimated that more than 63 ton green compost per hectare need to be applied[3] (including more than 300 kg P ha^{-1}, being far more than allowed by legislation) to exceed the allowed uncertainty of 10% in the farm stock as determined by the field averaged soil analyses. Although this can be management on field level, this challenges the soil and nutrient management to bring sufficient carbon to the soil while reducing the phosphorus inputs for environmental aspects.

Since the soil analyses done with the Handheld sensor show that much higher accuracies are possible when spatial variation within the field is accounted for, we recommend to increase the desired accuracy up to 5% at minimum. With this accuracy the expected change in organic matter given the 4-promille initiative can be quantified over a period of 10 to 20 years for most arable soils. In addition, the desired accuracy might theoretically increase to almost 1% as shown for a fake simulation where the measurement error of the Handheld sensor is reduced to zero (Figure 3.7.). As a consequence, the required sampling density goes down to maximally one sample per four hectares.

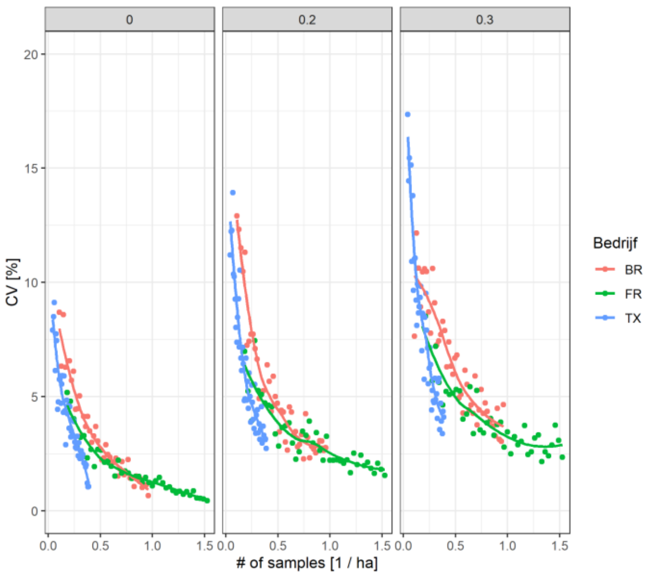


Figure 3-7. Change in predicted Coefficient of Variation (CV, %) of the carbon stock for the different farms in relation to the sampling density and the assumed measurement error of the handheld (30%, 20% and 0%).

[3] With a farm C stock of 100 ton per hectare, and an allowed deviation of 10% (with 90% confidence intervals), the uncertainty in C stock varies between 90 and 110 ton per hectare. So, the deviation varies from minus 10 to plus 10 ton C per ha. This variation equals to ±6.2 ton green compost (containing 218 kg Effective Organic Matter per ton compost).

When the mean C stock is derived from the 100 samples analysed at each farm as well with the Eurofins field analysis (Table 3.3), we found that there is huge agreement for the farm in Brouwershaven, a negative difference for the C stock of the farm in Texel and a much higher C stock for the farm in Friesland. The reason for these observed differences is unknown and might be attributed to missing spatial variability by Eurofins, differences in wet chemistry protocols, or uncertainties in the soil database of NMI which was used when Eurofin's field soil data was missing.

Table 3-3. Comparison of carbon stock [ton C] between Eurofins Agro and AgroCares Handheld for the different farms.

Farm	Eurofins Agro	Handheld	Difference
Brouwershaven	5432	5434	-2
Friesland	8636	6172	2464
Texel	9467	9811	-344

As an illustration, we compared the spatial variability of the wet chemistry samples of GSL with the field averaged samples analysed by Eurofins laboratory (Figure 3.8). If both methods represent the actual variation of soil organic matter in the fields, one should expect that these samples deviate along the dotted line (reflecting the ratio between SOC and SOM). This is not confirmed for the three farms analysed, in particular for the farm in Friesland. First, we see that substantial variation might exist within fields that is overlooked when taking a field averaged soil sample. Secondly, we see that soils with a higher soil organic matter level (Friesland in this case) has relatively lower SOM levels using the NIR estimate of Eurofins in comparison with the C measurement of the GSL, suggesting that the NIR model is less accurate at high SOM levels. At low soil organic matter levels the differences between both laboratories decline.

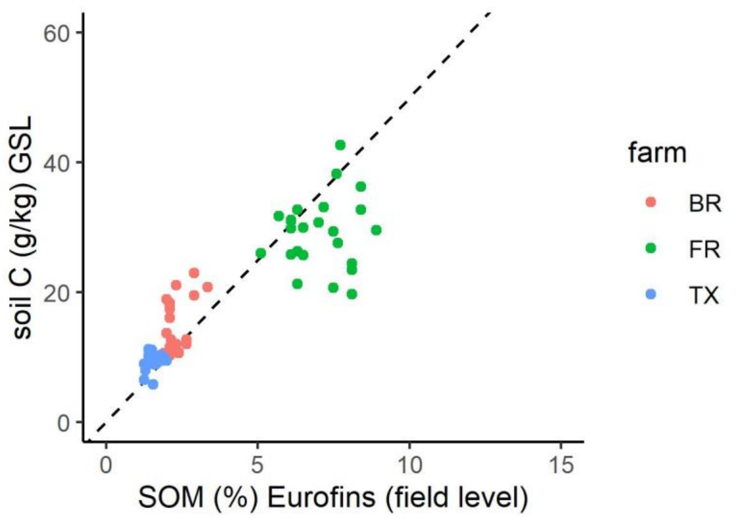


Figure 3-8. Comparison of wet chemistry SOC levels (measured in GSL) with corresponding SOM levels measured via NIRS on field level (measured by Eurofins). The dotted line is the SOM-SOC ratio of a factor 2.

3.5 Potential of BioScope zonation

Soil C is significantly different between zones for 35 (out of total 36) fields in Texel and for 13 (out of 14) fields in Brouwershaven ($p<0.05$ with 1-way ANOVA). This indicates that the zonation / stratification of Bioscope, which were identified based on satellite data, has a good potential of explaining the variation in soil C. Comparison per field are shown in Appendix IV. This observation coincides with the results of our PLS model in which satellite data (both Sentinel 2 and FAPAR) played an important role as predictor. The currently available zonation maps are not readily applicable as a covariable to predict soil C. The map does not cover grasslands as well as the fringes of parcels, and it only clusters areas within each parcel. If it is able to make zones for all parcels of a farm simultaneously, then the added value of using the zonation map becomes promising.

3.6 Impact of sampling depth

Soil C concentration is on average higher for top soil (0-10cm) than subsoil (10-30cm) (Figure 3.8). On individual location, however, subsoil C concentration sometimes exceeds top soil C concentrations. This might have been caused by recent top-soil removal in the grassland, or effect of oxidation in top soil of peat. In most arable soils one would expect a gradual decline over depth, in particular below the ploughing zone. In grassland soils, highest organic matter build-up occurs in and slightly below the rooting zone. Grassland renewal however can have strong disturbing impacts on the re-allocation of carbon over the first 30 cm, not necessarily declining the total C stock on farm level. Given the illustrative nature of this exercise, the exact reason for the observed patterns in the farm in Friesland have not been investigated and explored.

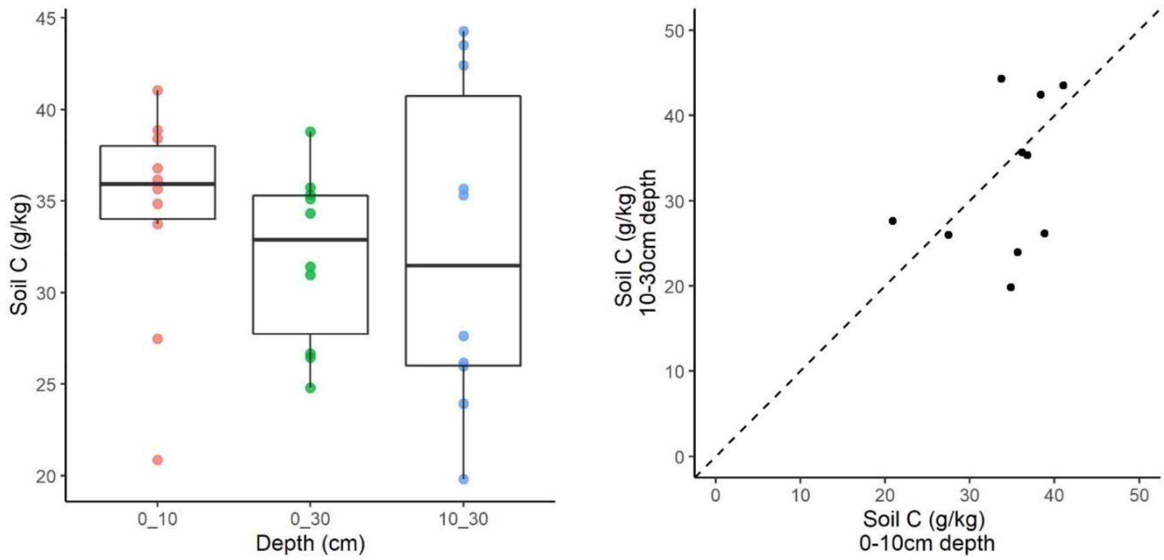


Figure 3-9. Soil organic C concentration of different depth of soil samples, measured at 10 locations in the farm of Friesland (left). Boxes represent 25th, 50th and 75th percentiles. Same data represented as correlation plot (right).

3.7 Sample costs

Sampling costs of the proposed methodology is mainly controlled by the manual labour needed to collect soil samples and scan these samples after mixing the soil to a homogenous sample. The costs for using the Handheld scanner is currently 1.600 euros per year (with unlimited number of scans for soil organic matter, pH and nitrogen) and the hardware costs are 3.000 euros per scanner. We recommend to use a certified sample taker in order to ensure high quality measurements that are needed for calculation of farm C stocks as well for the further extension of the database that is used to link satellite data to soil C levels.

Assuming a first start with 100 farms a year by using one Handheld scanner and where each farm is analysed by 30 sampling points (1 sample per 2 hectares), the analysis costs are estimated on about 1.5 euro per analysis. Increasing sampling density or number of farms linearly decreases sampling costs. Over time, the analysis costs are therefore negligible in comparison with the field worker needed to take and analyse the sample. It might be an option to request the farmer to take the soil samples on his/her own farm, reducing these costs. If an experienced field worker is used then one might expect that an accurate estimation of farm C stock takes 1 to 2 days of sampling per farm.

An option to increase sampling speed might be to analyse the C content of the top soil layer without digging and mixing a soil sample over a depth of 30 cm. In most arable systems the top soil layer is quite well mixed, and it is likely that the C content of the first cm can be used to estimate the C content deeper in the soil. This approach has not been tested yet but might have potential for upscaling the protocol to a large number of farms.

Assuming that the sampling protocol is optimized up to a sampling density of about 1 sample per 4 hectare, we estimate that the costs for C monitoring varies around 39 euro per hectare (Table 3.4). The cost estimation is indicative and is done for farms with an average area of 100 ha, for a period of 20 years and a measurement frequency of five years (n = 4 per farm). The available webservices include the estimation of soil properties, the collection of satellite covariates, the design of the sampling plan and an automated report generation. Sampling is done by a certified sample taker collecting 20 samples a day. Handheld scanners are used given an annual licence of 1300 euro.

Table 3-4. Fictive cost calculation (euro ha^{-1}) for future sampling and C monitoring for a given number of farms.

Costs	n = 500	n = 1000	n = 5000	n = 10000
4 webservices	0.26	0.19	0.14	0.14
Sample taking	38	38	38	38
SOC sensing	1.87	1.56	1.35	1.32
Total price per hectare	40.0	39.6	39.3	39.3
Number of scanners	9	15	65	127

4 Conclusion & outlook

4.1 Applicability of Carbon Protocol

Given the current price per unit Carbon sequestered there is an urgent need for cheap, affordable and robust soil organic carbon measurements taken at strategic sample locations reflecting the actual spatial variability within and among fields. The current protocol is designed and evaluated for application on three Dutch farms with varying crop rotation plan and soil types. This analysis shows that there is certainly potential for accurate quantification of changes in C stocks over time and space given the variability in management across the three farms. From this analysis we can conclude that:

- required sampling densities are around 1 sample per 2 hectares to achieve an accuracy of 5% of the C stock on farm level (with a 95% confidence interval). Increasing sampling density up to 1 sample per hectare decreases the observed deviation from the C stock to about 3%.
- required sampling densities decline with the number of farms sampled. This improvement might be even bigger when smart machine learning algorithms can be applied as soon as about 1000 samples are available (given our experience in developing predictive models for soil properties across the Netherlands). Adding weather related covariables or other site specific data available on farms might be another option to improve the performance of the PLS model used for upscaling and sampling design.
- required sampling densities also decline when the accuracy of the Handheld scanner increases. For the current study we used a relative mean error of about 30% given a prior calibration on a global database (and this is likely a worst case scenario since NIR sensors are known for their high reproducibility when used on the same location).
- the use of a priori knowledge of fields (derived from either earlier scans, results from agricultural laboratories or spatial models derived from these lab analyses) enhance the accuracy of the C stocks and reduces the number of additional samples needed.
- a priori knowledge of fields derived from results from agricultural laboratories alone can be used to estimate the C stock with a deviation between 10 á 20% but only when all the fields are sampled (and field size do not increase the 5 hectare; the common approach of Eurofins Agro includes a mixed sample of 40 subsamples taken from the field).
- the satellite derived maps of BioScope show a positive relationship with soil organic C levels within a field, but their applicability is currently limited due to the fact that grasslands are excluded and the underlying algorithm is unsupervised while excluding field boundaries.

4.2 Comparison with other sampling protocol

Different sampling approaches have been proposed in last years to estimate the farm C stock with high precision and accuracy. A commonly taken approach is stratified sampling or clustering. This method divides areas into discrete strata which have homogeneous environmental conditions based on prior knowledge. Sampling is designed to ensure that target metrics for each stratum is efficiently estimated. Several advanced methods were developed recently to make the stratified sampling more efficient and robust, including integrative hierarchical stepwise sampling approach (Yang et al., 2016) and optimized stratified random sampling (de Gruijter et al., 2016).

Our current approach, cLHS, selects samples that spread across the feature space of many environmental variables (Minasny & McBratney, 2006). cLHS can be considered as a stratified random procedure, as it stratifies the marginal distribution of the covariables into equally probable non-overlapping strata and draws a random sample from each stratum. A notable advantage of cLHS is that it guarantees to cover the full range of a multivariate feature space. Accordingly, an increasing number of studies use this method for soil mapping (e.g. Zhang et al., 2021, Adhikari & Hartemink, 2017, Taghizadeh-Mehrjardi et al., 2016). Furthermore, cLHS can be expanded to include additional conditions to evaluate sampling costs and accessibility to sampling points, so that sampling design can be optimized with explicit consideration on more practical constraints. Since cLHS requires high-quality data of environmental variables, clustering method has been proposed as a more feasible alternative in particular when sampling resources and legacy soil maps are limited (Yang et al., 2020). Although clustering and cLHS have their own pros and cons, success of both methods highly depends on how the selected environmental covariables can capture the spatial variation of the local soil (Yang et al., 2016).

The method of De Gruijter et al. (2019) optimizes sampling size and locations for farm scale carbon monitoring. However, the major drawback of this method is that it needs a carbon prediction map as input to calculate a sample design. For almost all fields and farms, such a map is not available with enough spatial detail. The method heavenly depends on the quality of the map, so an independent and extensive field campaign is needed a priori to create a map for an efficient sampling design. Besides this, the method of De Gruijter et al. (2019) optimizes the sample design on a number of assumptions. One of these assumptions is that costs of measurements linearly increases with the number of samples. This is true when using classical chemical soil analysis, but with the advent of modern soil sensors, like the AgroCares Handheld sensors, this is not the case anymore as the costs are determined by a fixed price for a license with unlimited measurements.

4.3 Sample Collection and Measurement

Monitoring the variation in SOC levels measured with the Handheld scanner within the different farms show unexpectedly higher variability within the Friesland farm compared to the two other farms. After discussion with the field workers we observed that extreme weather conditions during sampling was responsible for this; high wind disturbance complicated stable soil sensing and rainy conditions caused the soil to become wet (reaching the applicability range of the Handheld Scanner). We recommend to update the current sampling protocol with regard to these circumstances.

In addition to this, some general observations are made to uniform soil collection in the field, potentially reducing sampling error. These include:

- Fixed depth by using optimized augers that avoids sample collection from soil layer below the 0-30cm depth sampled.
- Predefined mixing procedures that vary with soil textural classes to avoid inhomogeneous soil samples to be scanned, or procedures for more replicates per sample when mixing is not feasible.
- The design of the current sampling scheme allowed samples to be taken nearby border of the field or surrounding ditches; since their impact on the C stock on farm level is likely limited, it might be valuable to add this as additional requirement to the sampling design.

4.4 Recommendations

The current study describes the development and evaluation of a sampling protocol for accurate assessment of C stocks on farm level. The protocol is tested and evaluated for three farms in the Netherlands with varying site properties and land management history. We showed that the proposed protocol has the potential to be applied across the Netherlands (and even outside the Netherlands given the use of open source available satellite data and soil databases). Soil C stocks on farm level can be quantified accurately with a maximum deviation less than 5% and a confidence interval of 95%. Further improvements leading to higher accuracy or lower sampling density requirements can be obtained when:

- PLS models linking satellite data and covariates to spatial variation in soil C levels are replaced by machine learning algorithms.
- The database used to train these PLS models are extended with production data from other farms being analysed for soil C stocks given the proposed procedure.
- The uncertainty on the Handheld sensor declines due to continues improvement of underlying deep learning models.
- Other data sources reflecting organic matter management on field level are available.

The aforementioned conclusions are derived from a limited set of farms (n = 3). We recommend to extend this pilot with a minimum of ten farms that vary in soil properties, crop rotation and location across the Netherlands to cover the potential variation in drivers controlling C build-up and soil organic matter decomposition. In the current phase it might still be valuable to use multiple sampling densities to further validate the conclusions of the current study.

Given that sample taking, homogenising and scanning in the field requires time, it might be valuable to assess whether soil surface scanning can replace the soil sampling protocol. This requires statistical functions that predicts how C levels decline over depth and was not part of the current study.

Soil sampling monitoring is aimed at detecting SOC content and stock changes from an initial baseline condition. Since changes in SOC are quite slow, it might be valuable to extend the sampling protocol with a simulation module that predicts the expected change in carbon stocks due to sustainable soil management practices (Ros, 2020). In that case it is also possible to account for business as usual management practices, where a reduction in soil C decline might be valorised equally as an increase in soil C. A multi-model ensemble approach, using multiple models to make predictions of SOC stocks, is the preferred option. It might be valuable to explore this in conjunction with the further development of the sampling monitoring.

Literature

Adhikari K & AE Hartemink (2017). *Soil organic carbon increases under intensive agriculture in the Central Sands*, Wisconsin, USA. Geoderma Regional 10, 115–125.

Brus DJ (2019) *Sampling for digital soil mapping: a tutorial supported by R scrips*. Geoderma 338, 464-480.

De Gruijter J, McBratney A, Minasny B, Wheeler I, Malone B & U Stockmann (2016). *Farm-scale soil carbon auditing*. Geoderma 265, 120–130.

De Gruijter JJ, Wheeler I & BP Malone (2019). *Using model predictions of soil carbon in farm-scale auditing - A software tool.* Agricultural Systems, 169, 24–30.

Minasny B & AB McBratney (2006). *A conditioned Latin hypercube method for sampling in the presence of ancillary information*. Computers and Geosciences, 32(9), 1378–1388.

Ros GH (2020) *Evaluation C monitoring protocols*. NMI note, 25 december 2020, 8 pp.

Smith P et al. (2019) *How to measure, report and verify soil carbon change to realize the potential of soil carbond sequestration for atmospheric greenhouse gas removal***.** Global Change Biology 26, 219-241.

Taghizadeh-Mehrjardi R, Nabiollahi K & R Kerry (2016). *Digital mapping of soil organic carbon at multiple depths using different data mining techniques in Baneh region, Iran*. Geoderma 266: 98-110.

Yang L, Qi F, Zhu AX, Shi JJ & YM An (2016). *Evaluation of integrative hierarchical stepwise sampling for digital soil mapping*. Soil Sci. Soc. Am. J. 80 (3), 637–651.

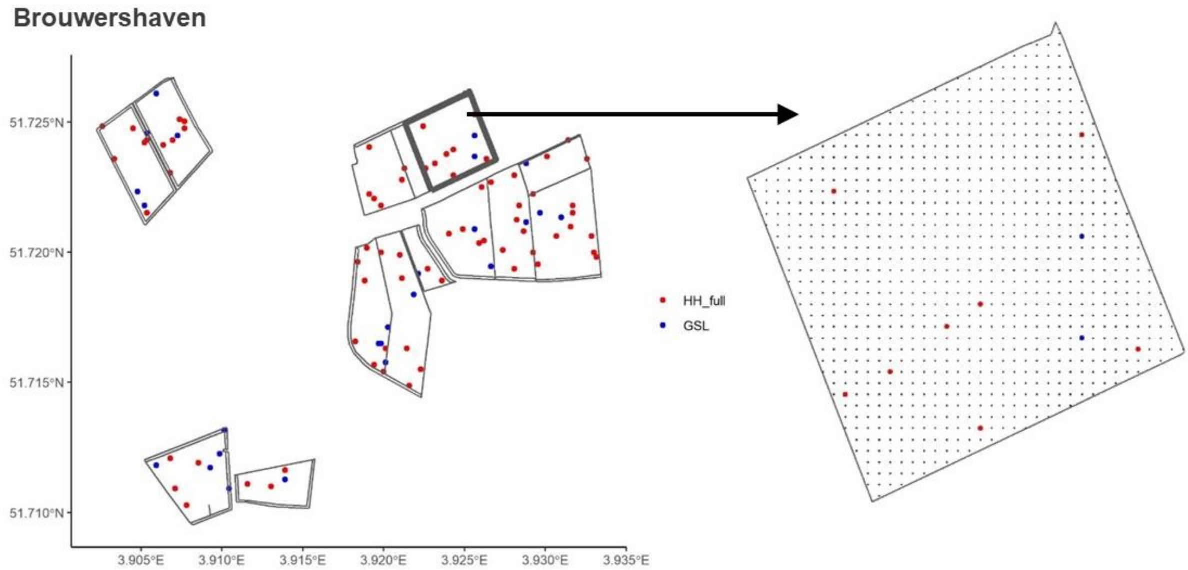
Yang L, Li X, Shi J, Shen F, Qi F, Gao B, Chen Z, Zhu A-X & C Zhou (2020) *Evaluation of conditioned Latin hypercube sampling for soil mapping based on a machine learning method*. Geoderma 369, 114337.

Zhang Y & AE Hartemink (2021). *Quantifying short-range variation of soil texture and total carbon of a 330-ha farm*. CATENA 201: 105200. https://doi.org/10.1016/j.catena.2021.105200

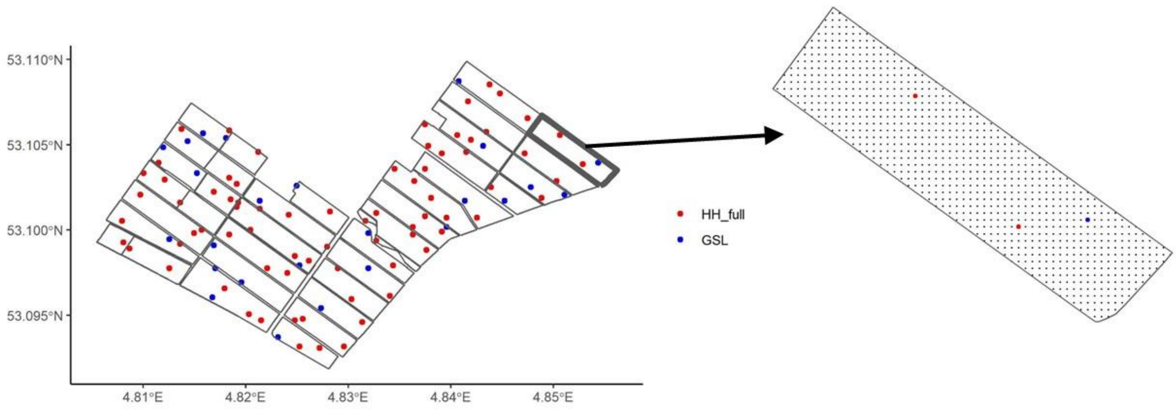
Appendices

Appendix I. Potential and selected sampling points

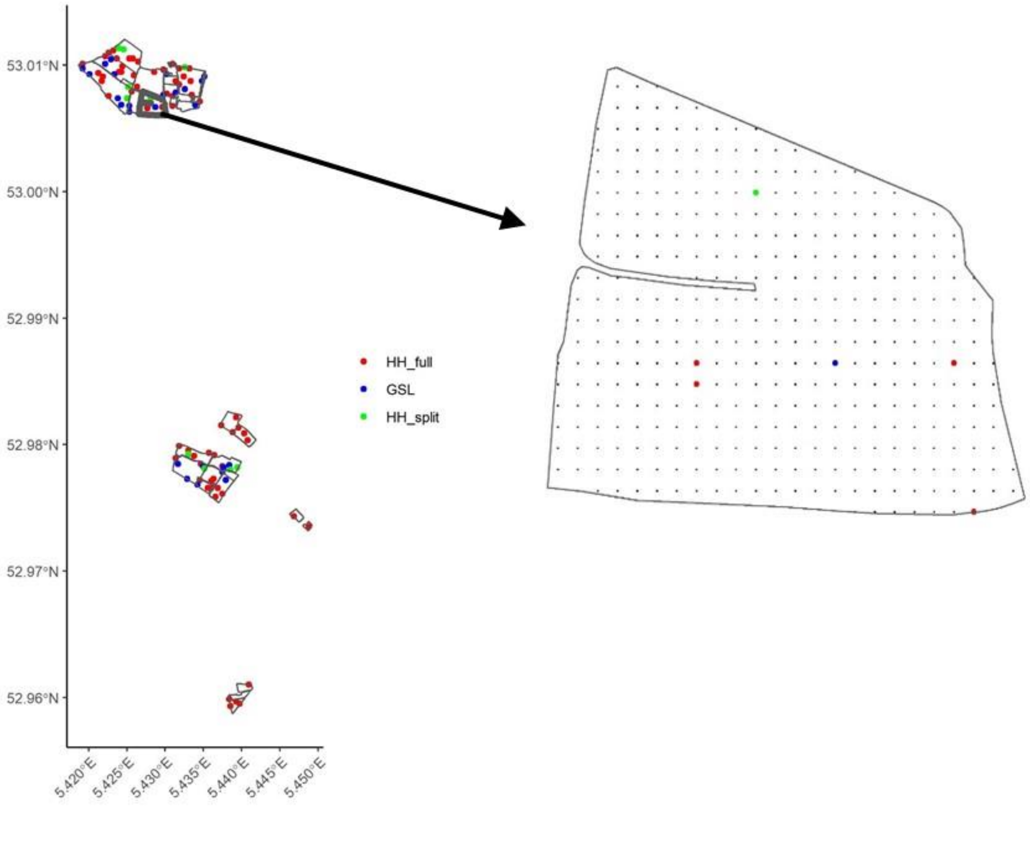
100 sampling locations selected with cLHS are shown below. Different colours of points represent the sampling type (HH_full: soil of 0-30cm depth is scanned, GSL: soil of 0-30cm depth is scanned and analyzed with GSL, HH_split: soil is scanned for 0-10, 10-30, and 0-30 cm separately). For each farm, one field was zoomed up to show all potential sampling points (small black points) which are regularly located on a 10m grid.



Texel



Friesland



Appendix II. List of covariables

Table 1. List of all collected covariables. Covariables used for the PLS model is marked with *.
CGLS: Copernicus Global Land Service, FAPAR: Fraction Absorbed Photosynthetically Active Radiation

Parameter	Source	Description	Format	Resolution
A_AL_CO_PO	Eurofin Agro	Al-occupation [%]	Polygon	
A_CA_CO*	Eurofin Agro	Ca-occupation [mmol+/kg]	Polygon	
A_CA_CO_PO*	Eurofin Agro	Ca-occupation [%]	Polygon	
A_CEC_CO*	Eurofin Agro	Cation Exchange Capacity [mmol+/kg]	Polygon	
A_CEC_CO_PO	Eurofin Agro	CEC-occupation [%]	Polygon	
A_CLAY_MI*	Eurofin Agro	Clay content [%]	Polygon	
A_CN_FR*	Eurofin Agro	C/N ratio	Polygon	
A_COS_FR	Eurofin Agro	Organic C fraction	Polygon	
A_CS_FR*	Eurofin Agro	C/S ratio	Polygon	
A_C_IF	Eurofin Agro	Inorganic C fraction	Polygon	
A_H_CO_PO	Eurofin Agro	H-occupation [%]	Polygon	
A_KZK_MI*	Eurofin Agro	Carbonic lime [%]	Polygon	
A_K_CC*	Eurofin Agro	Plant-available K [mg K/kg]	Polygon	
A_K_CO*	Eurofin Agro	K-occupation [mmol+/kg]	Polygon	
A_K_CO_PO*	Eurofin Agro	K-occupation [%]	Polygon	
A_K_KG*	Eurofin Agro	K-number	Polygon	
A_MG_CC*	Eurofin Agro	Plant-available Mg [mg Mg/kg]	Polygon	
A_MG_CO*	Eurofin Agro	Mg-occupation [mmol+/kg]	Polygon	
A_MG_CO_PO*	Eurofin Agro	Mg-occupation [%]	Polygon	
A_NA_CC*	Eurofin Agro	Plant available Na	Polygon	
A_NA_CO*	Eurofin Agro	Na-occuptation [mmol+/kg]	Polygon	
A_NA_CO_PO*	Eurofin Agro	Na-occuptation [%]	Polygon	
A_N_PMN*	Eurofin Agro	Microbial activity [mg N /kg]	Polygon	
A_N_RT*	Eurofin Agro	Total N content [mg N / kg]	Polygon	
A_OS_GV*	Eurofin Agro	Organic matter content [%]	Polygon	
A_PH_CC*	Eurofin Agro	Acidity	Polygon	
A_P_AL*	Eurofin Agro	P [mg P2O5 / 100 g]	Polygon	
A_P_CC*	Eurofin Agro	Plant available P (PAE) [mg P / kg]	Polygon	
A_P_WA*	Eurofin Agro	P-water (Pw) [mg P2O5 / l]	Polygon	
A_SAND_MI*	Eurofin Agro	Sand content [%]	Polygon	
A_SILT_MI*	Eurofin Agro	Silt content [%]	Polygon	
A_SLIB_MI	Eurofin Agro	Slib content [%]	Polygon	
A_S_CC	Eurofin Agro	Plant available S	Polygon	
A_S_RT*	Eurofin Agro	Total S content	Polygon	
A_C_BB	Eurofin Agro	Bacterial biomass	Polygon	
A_C_FB	Eurofin Agro	Funghi biomass	Polygon	
A_C_MB	Eurofin Agro	Microbial biomass	Polygon	
A_FB_FR	Eurofin Agro	Fungi / barcterial ratio	Polygon	
A_AL_OX*	NMI	Al-oxalate	Polygon	
A_FE_OX*	NMI	FE-oxalate	Polygon	
A_P_OX*	NMI	P-oxalate	Polygon	
A_P_VG*	NMI	P-saturation [%]	Polygon	
A_CN_OF*	NMI	C/N ratio	Polygon	
A_K_HCL	NMI	K content [%]	Polygon	
A_CU_CC*	NMI	Plant available Cu	Polygon	
A_CU_HNO3*	NMI	Cu content [%]	Polygon	
A_CO_CC	NMI	Plant available Co	Polygon	
A_CO_AA	NMI	Co content [%]	Polygon	
A_ZN_CC*	NMI	Plant available Zn	Polygon	
A_B_HW*	NMI	B hot water extraction	Polygon	
A_B_CC*	NMI	Plant available B	Polygon	
A_NA_HCL*	NMI	Na in HCl extraction	Polygon	
A_MN_CC	NMI	Plant available Mn	Polygon	
A_MO_CC*	NMI	Plant available Mo	Polygon	
A_SE_CC	NMI	Plant available Se	Polygon	
A_FE_CC*	NMI	Plant available Fe	Polygon	
B_BT_AK*	NMI	Agricultural soil type	Polygon	
D_CP_STARCH	BRP	Fraction of starch potatoes in cultivation rotation between 2009-2019	Polygon	

Parameter	Source	Description	Format	Resolution
D_CP_POTATO*	BRP	Fraction of (non-starch) potatoes in cultivation rotation between 2009-2019	Polygon	
D_CP_SUGARBEET*	BRP	Fraction of sugarbeets in cultivation rotation between 2009-2019	Polygon	
D_CP_GRASS*	BRP	Fraction of grassland in cultivation rotation between 2009-2019	Polygon	
D_CP_MAIS*	BRP	Fraction of maize in cultivation rotation between 2009-2019	Polygon	
D_CP_OTHER*	BRP	Fraction of other crops in cultivation rotation between 2009-2019	Polygon	
D_CP_RUST*	BRP	Fraction of rest crops in cultivation rotation between 2009-2019	Polygon	
D_CP_RUSTDEEP	BRP	Fraction of deep-rooting rest crops in cultivation rotation between 2009-2019	Polygon	
bd50.code*	Bodemkaart Nederland	Dutch soil code	Polygon	1:50.0000
bd50.gwt.org*	Bodemkaart Nederland	Groundwater table	Polygon	1:50.0000
rivm.ccyclus*	RIVM	Carbon cyclus	Raster	100m
b_rwt_dank	RIVM	Most common rain worms specie	Raster	100m
b_rwa_dank*	RIVM	Abundace of rain worms	Raster	100m
fapar_2014.2018_01*	CGLS	FAPAR in January, 2014-2018	Raster	300m
fapar_2014.2018_02*	CGLS	FAPAR in February, 2014-2018	Raster	300m
fapar_2014.2018_03*	CGLS	FAPAR in March, 2014-2018	Raster	300m
fapar_2014.2018_04*	CGLS	FAPAR in April, 2014-2018	Raster	300m
fapar_2014.2018_05*	CGLS	FAPAR in May, 2014-2018	Raster	300m
fapar_2014.2018_06*	CGLS	FAPAR in June, 2014-2018	Raster	300m
fapar_2014.2018_07*	CGLS	FAPAR in July, 2014-2018	Raster	300m
fapar_2014.2018_08*	CGLS	FAPAR in August, 2014-2018	Raster	300m
fapar_2014.2018_09*	CGLS	FAPAR in September, 2014-2018	Raster	300m
fapar_2014.2018_10*	CGLS	FAPAR in October, 2014-2018	Raster	300m
fapar_2014.2018_11*	CGLS	FAPAR in November, 2014-2018	Raster	300m
fapar_2014.2018_mean*	CGLS	FAPAR yearly mean, 2014-2018	Raster	300m
fapar_2014.2018_median*	CGLS	FAPAR yearly median, 2014-2018	Raster	300m
fapar_2014.2018_sd*	CGLS	FAPAR yearly standard deviation, 2014-2018	Raster	300m
sentinel_B2*	CGLS	Band 2 (Blue) of Sentinel 2 mosaic of 2017	Raster	10m
sentinel_B3*	CGLS	Band 3 (Green) of Sentinel 2 mosaic of 2017	Raster	10m
sentinel_B4*	CGLS	Band 4 (Red) of Sentinel 2 mosaic of 2017	Raster	10m
sentinel_B5*	CGLS	Band 5 (Vegetation Red Edge) of Sentinel 2 mosaic of 2017	Raster	20m
sentinel_B6*	CGLS	Band 6 (Vegetation Red Edge) of Sentinel 2 mosaic of 2017	Raster	20m
sentinel_B7	CGLS	Band 7 (Vegetation Red Edge) of Sentinel 2 mosaic of 2017	Raster	20m
sentinel_B8A	CGLS	Band 8A (Narrow NIR) of Sentinel 2 mosaic of 2017	Raster	20m
sentinel_B11*	CGLS	Band 11 (SWIR) of Sentinel 2 mosaic of 2017	Raster	20m
sentinel_B12*	CGLS	Band 12 (SWIR) of Sentinel 2 mosaic of 2017	Raster	20m
ahn3_5m_dtm*	AHN3	Terrain height from Actueel Hoogtebestand Nederland 3	Raster	5m

Appendix III. Protocol for sample collection

1) Equipment list

- Auger and spatula
- Sample bag
- Agrocares Scanner
- Styrofoam boxes
- Android phone with the Carbon/Global Application
- Sticks + waterproof pen + plastic tape (to mark the sampling location)

2) Sampling procedure

Each sample are taken at a depth of 0 – 30 cm and from a location determined by Agrocares. Below are the steps to follow to collect a sample. There are 3 different sampling procedures depending on the type of sampling:

a) HH_full: Soil of 0-30 cm is scanned
b) GSL: Soil of 0-30 cm is scanned and analyzed with GSL
c) HH_split: Soil is scanned for 0-10, 10-30, and 0-30cm separately

1. Choose the spot where you will collect the sample according to the GPS locations that were given to you. If there is no suitable spot around the given XY-coordinate (i.e. within 2m radius from the point), skip the location (see note *2 for the criteria of good sample location).
2. Switch on the external gps device and note the exact GPS location and the sample number/barcode. As the accuracy of the location is quite important, confirm that the gps signal is stable.
3. Collect the sample from a bare spot if possible.
4. If the soil surface is covered with plants or plant debris – remove them before you take the sample.
5. If the soil is very hard: Start by breaking the soil crust using the spiral auger
6. Drill down with the auger to 30 cm.
7. Remove the top 2cm of soil from the sample auger, in order to remove any plant debris that might have fallen into the drill hole.

[HH_full & GSL]

8ab. Put the sample in an empty and clean white bucket and break the particles to find stones and roots.

9ab. Remove the stones (>2mm), big roots and organic material particles (>2mm) from the sample (you can eventually use a 2 mm sieve). Take a picture of the soil.

10ab. Scan the sample with the AgroCares Scanner (see the chapter on how to scan a sample below)

[GSL only]

11b. Take a clean sample bag and attach a barcode/label to it (e.g. with a paper clip).

12b. Place the sample in the sample bag. See note *1 for the minimum required soil weight.

13b. Push the air out of the plastic bag, fold it a few times, close it by twisting the yellow ends of the sample bag and make sure no soil can escape.

14b. Place the sample bag in a cool box or a well insulating styrofoam box. Wet soil needs to be kept at 10 degrees Celsius. This box can be either cooled electrically or by using ice-packs depending on the circumstances. A barcode/sample number should be given to the sample to later on identify the sample when it is additionally analyzed in the GSL. The sample will be put into a sample bag together with the barcode/notes and stored in a cooler for transport.

15b. Clean all your tools to avoid rusting and contamination of the next sample.

[HH_split]

8c. Put the upper 10cm soil and lower (10-30cm) soil separately in an empty and clean white bucket and break the particles to find stones and roots.

9c. Remove the stones (>2mm), big roots and organic material particles (>2mm) from the sample (you can eventually use a 2 mm sieve). Take a picture of the soil.

10c. Scan the sample with the AgroCares Scanner (see the chapter on how to scan a sample below), separately for 0-10cm and 10-30cm depth. After scanning, combine both soil and mix well, and scan again.

11c. Clean all your tools to avoid rusting and contamination of the next sample.

*1: Minimum weight

The required sample weight after setting the scale without weight to 0g should be:
If dry – minimum of 600 g
If humid – minimum of 750 g
If wet – minimum of 850 g If the sample weight is less than the minimum required – drill a second sample hole just 5 cm away from the first one and repeat the procedure. Proceed until you have enough sample material.

*2: Choosing the final sample location

General remarks on good sample locations:

- Do not take samples within homesteads.
- Do not take samples on visibly contaminated spots (oil/fuel, rubbish, nutrient sources like heaps or bags of compost /plant debris/fertilizer…).
- Do not take a sample where the topsoil was obviously removed.
- Do not take samples near trees: Do not collect samples under tree crowns.
- If it is not possible to sample at the location within a radius of 2 metres, skip the location

3) Scan a sample

Scanning equipment

Scan the samples 5 times during the day with the AgroCares Scanner and the Carbon/Global Application. During the day you need to scan the samples with the AgroCares scanner and the GS soil app on your phone. You need:

- Scanner
- Sample cup
- Phone with Carbon/Global App

The sampling electrical equipment needs to be charged daily to ensure its good functionality while sample taking.

Scanning steps

First make sure that the app is working on your Android Phone.

3.1) Turn on the scanner

Push on the scan button for 3 seconds to turn on the scanner. When the scanner is turned on, 2 led lights start blinking. The green “scan” led and the yellow/green “battery” light blink for about 10 seconds to perform a system check of the device. When scanner is ready to connect, the Battery icon (green LED) is continuously on and the Bluetooth icon starts blinking (blue LED).

3.2) Connect the scanner via Bluetooth

The scanner must be connected with the phone via Bluetooth. Therefore, it is important to have the Bluetooth connection activated on the phone. You can activate your Bluetooth in the device settings of your phone, or in the dropdown menu. Please check the manual of your phone when this proves difficult. The scanner is ready to connect with your smartphone when the battery icon (green led) is continuously on and the Bluetooth icon (blue led) starts blinking. To connect the scanner with your phone you open the app and select in the bottom menu “account”. Find the scanner ID in the list of “paired devices”. If you haven’t connected the scanner before with this phone push the orange bar “search for devices” and find your scanner. The name of your scanner ID SC_xxxX can be found on the side of your device.

3.3) Scanning your soil

When the scanner is connected with your phone you can start to scan your soil. There are several ways in the app that lead to the scanning process. Select “scan” in the home menu. Or select “scan” in the below menu of the app. Or select “scan field” when you are inside your clients information. Select one of the three options and wait until the green LED light on the button of the scanner stays continuously green like in the image below. If the scanner button does not become continuously green, then retry by re-entering the scan menu.

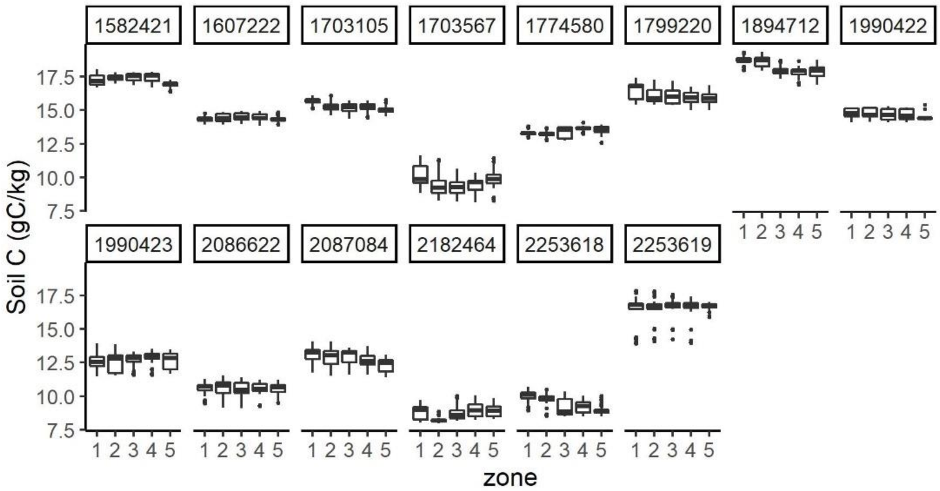
Every first scan of the day needs a complete calibration. This includes scanning the white side (background) and yellow side (standard sample) of the calibration cap. The complete calibration is also asked when a scanner is paired for the first time with a phone, after 20 soils processed within the same day, or when the scanner is unused for more than 12 hours. The app always indicates when a complete calibration is needed.

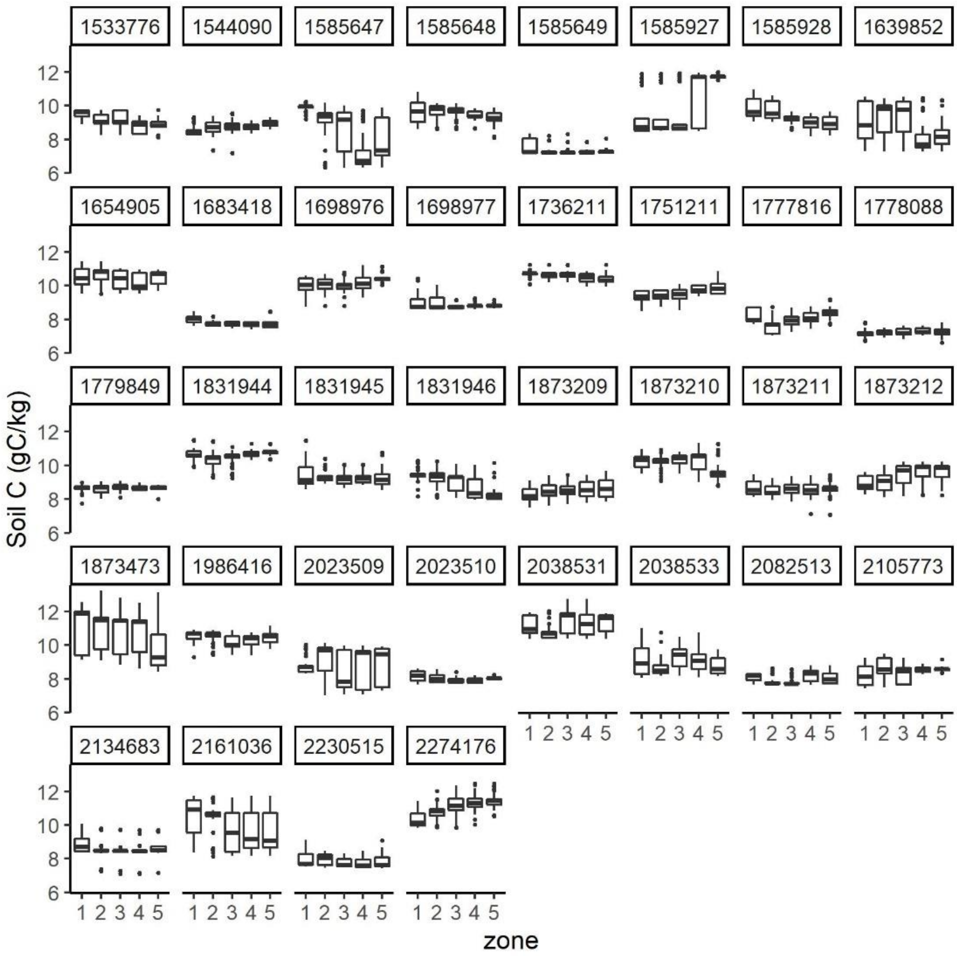
When it is your first scan the app will ask for the white background first. Place the scanner on the background and ensure that the pins lock correctly on the pinholes. Once this is done, press the scan button for 1 second. Don’t press for too long this will turn off the scanner! Upon releasing the scanner button, you will hear a beep meaning it is scanning. Also, the green light will be blinking indicating the scanning process has started. When the scans light is constantly green again you can make a second scan. Now turn around the calibration cap and scan the other side with the yellow standard sample.

Now you can start scanning your soil. Your sample needs to be scanned 5 times. Each scan takes 30 seconds. Make sure that the soil of your sample is well mixed in a bucked and that additional roots, stones and organic matter are removed from your sample. This will improve the quality of your measurement. Take for each scan a full scoop of soil in the grey sample cup and press the scanner against the soil. Push the button ‘scan’ for 1 second. They grey sample cup ensures a reliable EC measurement and a standardized procedure for all scanner users. This cup should always be used, do not replace the sample cup by scanning directly on the fields soil or by using a bucket. Make sure that the sensor head of the scanner stays clean between each scan to guarantee the best performance of the device.

Once the 5 scans have been successfully made the app directs you automatically to the screen where you can select your customer, field and crops to complete your order.

Appendix IV. C content per field and zone for Bioscope maps





1533776
1544090
1585647
1585648
1585649
1585927
1585928
1639852
1654905
1683418
1698976
1698977
1736211
1751211
1777816
1778088
1779849
1831944
1831945
1831946
1873209
1873210
1873211
1873212
1873473
1986416
2023509
2023510
2038531
2038533
2082513
2105773
2134683
2161036
2230515
2274176
Soil C (gC/kg)
zone

soil for life

Nutrient Management Institute BV
Nieuwe Kanaal 7c
6709 PA Wageningen

tel: (06) 29 03 71 03
e-mail: nmi@nmi-agro.nlnmi
website: www.nmi-agro.nl