Ministry for Primary Industries
Manatū Ahu Matua

Design of an on-farm soil carbon benchmarking and monitoring approach for individual pastoral farms

MPI Technical Paper No: 2020/02

Prepared for Theresa Wilson, International Relations, Ministry for Primary Industries

by Paul Mudge; Stephen McNeill; Carolyn Hedley; Pierre Roudier and Matteo Poggio, Manaaki Whenua – Landcare Research; Brendan Malone and Jeff Baldock, CSIRO; Paul Smith, Pipinui Trust; Sam McNally and Mike Beare, Plant and Food Research; Louis Schipper, The University of Waikato.

ISBN No: 978-1-99-002535-8 (online)

ISSN No: 2253-3923 (online)

August 2020

New Zealand Government

Disclaimer

Requests for further copies should be directed to:

Publications Logistics Officer
Ministry for Primary Industries
PO Box 2526
WELLINGTON 6140

Email: brand@mpi.govt.nz
Telephone: 0800 00 83 33
Facsimile: 04-894 0300

This publication is also available on the Ministry for Primary Industries website at
http://www.mpi.govt.nz/news-resources/publications.aspx

Reviewed by:	Approved for release by:
Norman Mason	Peter Millard General Manager – Science
Manaaki Whenua – Landcare Research	Manaaki Whenua – Landcare Research
MWLR Contract Report:	LC3549

Contents

Glossary

Auditing: In the context of soil carbon, verifying an estimate of the soil organic stocks at a point in time. See also: **benchmarking**.

Benchmarking: here refers to quantifying soil organic stocks in a CEA at time zero (a baseline).

Bulked sample: A sample in which the sampling units are pooled and homogenised. Sometimes called a composite sample.

Carbon estimation area (CEA): a defined area of land where soil organic carbon stocks will be estimated. For example, in the context of this report, a CEA could be a whole farm, a group of paddocks or a single paddock.

Confidence interval: A range of values, within which the true parameter value (e.g. the mean) is contained with a specified probability. A 95% confidence interval implies that the estimation process will produce an interval within which the true value would be contained with a probability of 0.95. Note that the stated probability level refers to properties of the interval.

Monitoring: determining how soil organic stocks in a CEA change over time.

Power: The probability of a hypothesis test of finding an effect if there is an effect to be found. A power analysis can be used to estimate the minimum sample size required so that a specified change in soil organic carbon stocks can be measured, given a desired significance level, effect size, and statistical power

Random sampling: A sample of items selected from a population in such a way that each sample of the same size is equally likely to be chosen in the sample.

Significance level: The probability of rejecting the null hypothesis when it is true. For example, a significance level of 0.05 indicates a 5% risk of concluding that a difference in soil organic carbon stocks has been detected when there is no actual difference.

Soil organic Carbon concentration (SOC): The mass of organic carbon per unit mass of oven dried soil. Generally expressed in units of grams of carbon per 100 grams of oven dried soil, or percent of the mass of the oven dried soil.

Soil organic Carbon stocks (SOCS): The mass of organic carbon in a sample of known bulk density. Generally expressed in megagrams of carbon per hectare ($Mg.ha^{-1}$) for a nominated depth (in this report, 0-30 cm).

Standard deviation: a measure of the amount of variation that a set of values has from the mean of the values.

Stratum: an area of land within a CEA (plural is strata).

Stratification: the process by which a CEA is divided into different strata, with land in each stratum being relatively homogenous in terms of biophysical features (e.g. soils, topography), land management regimes, and soil organic carbon stocks.

Type I error: In statistical hypothesis testing, a Type I error is where true null hypothesis is rejected. This is also known as a "false positive" decision. The Type I error rate or probability is the chance that a Type I error is made.

Type II error: In statistical hypothesis testing, a Type II error is where a false null hypothesis fails to be rejected. This is also known as a "false negative".

Executive summary

Background and objectives

Individual farmers and primary industry organisations in New Zealand are becoming increasingly interested in the risks and opportunities associated with sequestering carbon in agricultural soils and thus how soil organic carbon stocks (SOCS) and stock changes can be measured. However, there is currently no standardised SOCS measurement system recommended for use on New Zealand farms. **The primary purpose of this report is to provide recommendations for the design of a SOCS benchmarking and monitoring programme for individual pastoral farms in New Zealand.** The expertise required and costs associated with implementation of farm-scale SOCS measurements are also outlined.

Existing SOCS benchmarking and monitoring systems

There are three existing SOCS benchmarking and monitoring protocols that are relevant at the farm-scale: **1)** the Food and Agriculture Organization's 'Measuring and modelling soil carbon stocks and stock changes in livestock production systems' method (FAO 2019); **2)** the Australian Governments (2018) 'Carbon Credits (Carbon Farming Initiative – Measurement of Soil Carbon Sequestration in Agricultural Systems) Methodology Determination'; and **3)** VERRA's Verified Carbon Standard, Soil Carbon Quantification Methodology. We believe the Australian 2018 methodology is the most relevant to farm-scale SOCS measurements in New Zealand and it is also robust, having been implemented in national legislation and well supported by recent literature. Therefore, we used the Australian 2018 method to guide the approach recommended in New Zealand, while at the same time considering the FAO (2019) and VERRA methods, plus other literature and taking into account any unique issues specific to New Zealand landforms and management.

We believe that currently there is not enough New Zealand data to parameterise models (such as used in Australia and Canada) to confidently predict how SOC will change in response to changes in management for **specific individual farms** across New Zealand (due to the diversity of climate, soil type and management regimes). However, SOCS data and modelling capability is rapidly increasing in New Zealand and this may be an avenue to pursue in the future.

Recommended approaches for farm-scale SOC benchmarking and monitoring in NZ

Define objectives and the carbon estimation area(s) (CEA)

The first step in the development of a SOCS benchmarking and monitoring programme is to clearly articulate the objectives including a consideration of the costs and potential benefits. Going through this process will help confirm the desire to proceed, and the objectives can influence study design and sampling intensity. The CEA also needs to be clearly defined – the whole farm, a group of paddocks (i.e. land management unit) or an individual paddock.

Stratification

The next step is to assess whether the CEA should be broken into different strata (e.g. based on soils and management) within which SOCS and stock changes will be relatively uniform compared with those across the whole CEA. Stratification can help enable more precise estimates of SOCS and stock changes for a given sampling effort. We recommend using one of two general approaches for stratification: 1) a relatively simple approach based on land management units (LMUs) as defined in farm environment plans (FEP); and 2) a more complex approach that relies on some prior knowledge of SOCS and multiple spatial data layers. Choice of stratification method depends on study objectives and availability of existing data, expertise, and money.

Number of soil samples required

Estimating the number of soil samples required is challenging and depends on how precise estimates of SOCS and stock changes need to be and the variability and rates of change of SOCS. Unfortunately, before results are obtained from soil sampling in a CEA, it is difficult to estimate the variability of SOCS, since SOCS vary between farms, and therefore it is difficult to calculate how many samples would be required without additional information. The best approach is to conduct a pilot study where soil samples are collected from across the CEA and analysed for SOCS. Once an estimate of the mean and variance of SOCS has been determined, a power analysis can be used to estimate the overall number of sampling sites required to determine baseline SOCS and changes in SOCS for a given level of statistical certainty.

Soil sampling, processing, and analysis

We recommend using a choice of three soil sampling methods to determine SOCS: 1) deep continuous cores; 2) short cores from pits; and 3) quantitative pits. The method chosen is determined by on-site soil conditions (e.g. stony soils, sandy soils) and the method used for benchmarking (baseline) sampling should also be used for subsequent sampling rounds. Sampling must be to a minimum of 30-cm depth and changes in SOCS must be calculated on an equivalent soil mass basis. We recommend soil processing and especially organic carbon analysis be undertaken by a laboratory with accreditation from International Accreditation New Zealand (IANZ). Following this recommendation will ensure that the concentration of SOC in a soil sample will be accurate and results between benchmarking and subsequent monitoring rounds consistent and comparable.

Expertise required

A range of expertise is required to design and implement a SOCS benchmarking and monitoring programme. A key attribute required is the ability to pay **attention to detail** and repeatedly follow protocols precisely. Soil sampling for SOCS needs to be conducted more carefully than for standard soil fertility testing because both accurate measurement of the concentration of carbon and total mass of soil in a defined sample volume are required. As we are attempting to demonstrate changes as small as 2–5% of the stocks, even small errors with sampling, sample processing, and analysis will mask true changes or indicate untrue increases or decreases. Currently, the expertise required for study design and implementation of the necessary protocols for SOCS benchmarking and monitoring can be provided by scientists and associated technicians from Crown Research Institutes (CRIs) or universities. Alternatively, many rural professionals (e.g. farm consultants) with some specific training related to processes influencing SOCS and stock changes, soil sampling and processing, and subsequent data analysis and interpretation could design and implement a SOCS benchmarking and monitoring programme for individual farms. Additional specialist statistical skills might be required for estimation of the number of samples, and spatial modelling skills would be needed if the detailed stratification option were chosen. Development of standardized easy-to-use spreadsheets could simplify and streamline the data analysis and interpretation process and would help potential aggregation of SOCS data into industry and national databases for wider analyses.

Costs

The cost to design and implement a SOCS benchmarking and monitoring system will vary considerably between farms, depending on a range of factors, such as the objective (e.g. the precision with which SOCS and stock changes need to be determined) farm size, topography, spatial variability, and whether individual or composite samples are analysed. The costs will also depend on who is designing and implementing the system and the associated charge out rates. An example is provided in Section 5.6 of this report.

1 Introduction

Soil organic carbon is critical for overall soil health, and because soils contain over twice as much carbon as the atmosphere, any increase or decrease will impact atmospheric CO_2 concentrations and thus climate. Carbon is continuously cycled in soil and therefore stocks can change over relatively short time scales in response to changes in land use and management (and changes in temperature and precipitation regimes). In general, losses of carbon can occur faster than gains. Concern over climate change is prompting nations, organisations, and individuals to implement actions to curb greenhouse gas (GHG) emissions and sequester carbon.

However, there is currently no standardised **system** recommended for use in New Zealand, meaning there is a risk that poor study design and sampling methods may lead to data being of limited value. There is, therefore, an urgent need to reach consensus on the most appropriate method(s) to benchmark and monitor changes in soil organic carbon stocks (SOCS) at the farm scale in New Zealand. It is also important to clearly outline the **expertise** required, as well as the **costs** associated with implementation of farm-scale SOCS measurements. Development and documentation of such a system will enable landowners/managers to make more informed decisions on whether to and how to implement SOCS benchmarking and monitoring. Use of a consistent approach will enable easy comparison between farms and scaling to industry, regional and national initiatives (and vice versa).

The primary purpose of this report is to provide recommendations for the design of a SOCS benchmarking and monitoring programme for individual pastoral farms in New Zealand.

Here, **benchmarking** refers to quantifying SOCS at time zero (a baseline) and **monitoring** is determining how SOCS change over time. While the focus of this report is on pastoral farms, the methods could be easily adapted for other land uses.

It is important to note that this report only covers the technical aspects of SOCS benchmarking and monitoring (e.g. study design and sampling) and does not specifically address how derived data could be used in policy, regulatory or marketing contexts.

In this report we focus mainly on requirements for a statistically robust method for SOCS benchmarking and monitoring which could then be modified (e.g. reduce the sampling intensity) if less robust results were required. However, it must be emphasised that the approach chosen will impact the **inference** that can be drawn from results. For example, SOCS monitoring with a few samples on **one area** of a farm cannot provide a quantitative estimate of changes for the **whole farm**.

1.1 Report outline

The report will begin with an overview of the basic principles of soil carbon benchmarking and monitoring and a summary and discussion of existing SOCS benchmarking and monitoring systems. This will be followed by a more detailed discussion of the technical issues and options that need to be considered before recommending a system for New Zealand.

Section 5 is the key part of the report and provides recommendations covering overall study design, sampling methods, sample processing and analysis, the expertise required and the estimated costs associated with the implementation of farm-scale SOCS benchmarking and monitoring in New Zealand. The recommendations section is written as a *summary* of the key steps and in some cases refer to other documents which have the specific details required for implementation. Data and learnings from previous implementation of SOCS benchmarking on two case-study farms with contrasting topography and management were used to inform our recommendations and provide examples (Malone et al. 2018). The case-study farms were the Massey University No.1 Dairy and Tuapaka sheep and beef farms.

2 Basic principles of soil carbon benchmarking and monitoring

As noted in the introduction, we define **benchmarking** as quantifying (or auditing) SOCS at time zero (a baseline) and **monitoring** as quantifying how SOCS change over time. To be fully quantitative, the benchmarking and monitoring needs to be representative of a defined spatial area such as a single paddock, a group of paddocks (land management unit/block) or the whole farm. Here we define the spatial area of interest as a **carbon estimation area (CEA)**. The key questions that need to be addressed for soil carbon benchmarking and monitoring are therefore:

1 What is the spatial mean SOCS for a specific CEA (e.g. farm)?
2 What is the temporal trend of the spatial mean of the SOCS for a specific CEA?

The key steps for SOC benchmarking and monitoring are:

1 Define the objectives for the benchmarking and monitoring.
2 Define and delineate the CEA (e.g. whole farm or a block within the farm).
3 Assess whether the CEA needs to be broken into homogenous strata (e.g. river flats, hills).
4 Collect representative soil samples from each stratum, conduct C analyses on samples and quantify baseline SOCS for the CEA at some initial **time.**
5 Quantify SOCS in same CEA at some later **time** (1–5 years after the initial time).
6 Calculate SOCS **change** through time.
7 Repeat steps 4–6 for times 2, 3, 4… (at 1–5-year intervals).

Modelling is another approach that can be used to predict changes in soil carbon in response to changes in land use and management (e.g. Kirschbaum et al. 2017), and can provide rich information concerning the state and likely trend of SOCS. However, it is our view that currently there is not sufficient New Zealand data to parameterise models to confidently predict how SOC will change in response to changes in management for specific, individual farms across New Zealand (due to the diversity of climate, soil type and management regimes). However, SOCS data and modelling capability is rapidly increasing in New Zealand and this may be an avenue to pursue in the future.

3 Summary of existing SOC benchmarking and monitoring systems

Despite the importance of soil carbon, there are few functioning national, regional, or farm-scale soil carbon monitoring systems in place, and until recently few approaches were based on direct measurements of changes in SOCS over time. Here we provide a brief overview of some of the systems currently in place around the world to provide background and determine their strengths and weaknesses in relation to potential adoption/adaption for use at the farm-scale in New Zealand.

3.1 United Nations Framework Convention on Climate Change (UNFCCC)

In their review of proximal sensing for soil carbon auditing, England and Viscarra Rossel (2018) provide an overview of the policy and reporting instruments that are concerned with the topic of soil organic carbon accounting. In summary, the United Nations Framework Convention on Climate Change (UNFCCC), and later the Kyoto Protocol, set up a system of national communications and national inventory reporting to be compiled by parties and published by the UNFCCC. To estimate GHG emissions and to monitor changes in C stocks, including soil organic C, the International Panel on Climate Change (IPCC) developed a tiered methodology that relates data on land use and management activities to emissions and storage factors to estimate fluxes from the activities (IPCC, 2006). The three-tiered approach depends on the scale, capability, and availability of data. Using the tiered approach, individual countries can interact with the required methods in the appropriate manner given the available information they have, to fulfil their reporting obligations. Such reporting obligations instigate efforts to develop policies and schemes that seek to reduce or offset emissions through a range of activities, including improved land management practices.

Current IPCC inventory approaches are unlikely to be applicable at the farm scale because they use global or country-specific parameters that apply to broad classes of land use and/or climate. Moreover, where current IPCC methods use national data, the samples on which they are based were not designed to be used at the finer farm scale. However, Malone et al. (2018) showed that SOCS data and model predictions from New Zealand's National Soil Carbon Monitoring system could be utilised to inform sampling design and intensity for farm-scale SOCS benchmarking.

3.2 Canadian systems

As summarised by England and Viscarra Rossel (2018), in Canada, the Government of Alberta amended the Climate Change and Emissions Management Act (CCEMA) in 2007 to require industries with substantial emissions to report and reduce their emissions to established targets. There are three options to meet the targets, one of which is emission offsets. The Alberta Offset System operates under a set of standards, known as the 'Offset Quantification Protocols and Guidance Documents'. The development of protocols involves the inclusion of expert engagement, rigorous peer review, defensible scientific methodologies, and documented transparency. Under the Alberta Offset Scheme, the Conservation Cropping Protocol quantifies SOCS change following changed land management from conventional cropping to conservation cropping (reduced tillage and summer fallow). The protocol uses Canada's National Emissions Tier 2 methodology, which developed carbon sequestration coefficients (emission factors) based on measuring and modelling local crop rotations, soil/landscape types, and inter-annual climate variation for geo-specific polygons in the national eco-stratification system.

A second Canadian scheme in Saskatchewan, the Prairie Soil Carbon Balance (PSCB) Project was developed for quantifying and verifying the direct measurement of SOCS changes in response to a shift from conventional tillage to no-till, direct-seeded cropping systems (McConkey et al. 2013). It was not designed to monetize soil carbon offsets but has been supported by farm groups with an interest in securing financial recognition for GHG mitigation. A regional monitoring project was established to measure the temporal change in soil carbon storage under agricultural cropland on the Canadian Prairies (Ellert et al. 2001). The sampling scheme reported in Ellert et al. (2001) for monitoring was designed to maximize the ability to detect changes in SOCS over time by ensuring that exact sample locations can be relocated, limiting horizontal variability of SOCS.

3.3 Australian systems

In Australia under the Emission Reduction Fund, eligible emissions reduction activities are included in 'methodology determinations', or 'methods' for short. Emissions reduction methods set out the rules for estimating emissions reductions from different activities. These methods ensure that emissions reductions are genuine – that they are both real and additional to business-as-usual operations.

In relation to methods associated with soil carbon auditing, two are currently in use:

1. The Carbon Credits (Carbon Farming Initiative—Measurement of Soil Carbon Sequestration in Agricultural Systems) Methodology Determination 2018 (https://www.legislation.gov.au/Details/F2018L00089)
2. Carbon Credits (Carbon Farming Initiative—Estimating Sequestration of Carbon in Soil Using Default Values) Methodology Determination 2015 (https://www.legislation.gov.au/Details/F2016C00263)

The 2018 method subsumes an earlier legislated method, *Carbon Credits (Carbon Farming Initiative) (Sequestering Carbon in Soils in Grazing Systems) Methodology Determination 2014* (2014 method: https://www.legislation.gov.au/Details/F2014L00987). The 2014 method was based on the direct measurement of changes in soil carbon stock over time, in response to changes in grazing systems management. Collection and analysis of soil samples over time generate the data to estimate soil carbon stock change. The abatement activities covered by the 2014 soil carbon method are included within the scope of the 2018 method. However, the 2018 method can be applied to a wider range of agricultural systems than the 2014 method. The 2018 method was designed to enhance the ability to detect a temporal change in soil carbon stocks and potentially lower soil analysis costs by including additional methods of soil carbon analysis. Advancements in the 2018 method compared with the 2014 method include:

- an improved soil sampling strategy that can use information related to spatial variation in soil carbon stocks to stratify the area included in a project to reduce the uncertainty of soil carbon estimates and enhance the ability to detect management induced temporal changes;
- an increased range of eligible farming systems including cropping, grazing and horticultural production systems;
- allowing the use of soil amendments containing biochar;
- allowing management activities involving the addition or redistribution of soil using mechanical means including clay delving, clay spreading or water ponding;

- an additional measurement option allowing for the ability to estimate carbon stocks using in-field or laboratory sensors and associated models as well as the combustion techniques;
- a 10-year baseline period; and
- use of a land management strategy, to be developed or reviewed by an independent person.

The second method, *Estimating sequestration of carbon in soil using default values* (default value method), is based on the use of default rates for soil carbon stock change over time, in response to changes in specified management practices for cropping systems. These default values were predicted using simulation results obtained by applying the Full Carbon Accounting Model (FullCAM) developed for, and used in, the Australian National Greenhouse Gas Inventory.

Before the 2018 method was introduced, a limited uptake of either the 2014 or default value soil carbon methods had been identified. There are several reasons for this, and a few are discussed here. The scope of the 2014 direct measurement method was limited to grazing lands and did not include cropping soils. The 2018 direct measurement method can be applied to a wide range of agricultural systems. With respect to the default value method, it only requires demonstration of adoption of well-defined changes in management practice and requires no actual measurement of soil carbon change. Although this made the method relatively inexpensive to implement, it was very restrictive in the definition of baseline conditions and what could be done during a project. These restrictions were required because the default values were modelled assuming such restrictions were in place. The decreased flexibility in future management practices that can be applied by land managers within a default values project has limited the development of projects of this type. It should also be noted that while these management activity-based methods are comparatively cheaper to implement than direct measurement methods (Conant & Paustian 2002), such methods present technical challenges considering the diverse combinations of climate, soil type and management regimes. As pointed out in de Gruijter et al. (2016), it is inevitable that not all combinations will be covered or parameterized and support for emerging managements will have a temporal lag in incorporation as data over time is required. At present there are approximately 17M tonnes of CO_2 equivalent (CO_2e) contracted from soil-based projects running under the 2014 and 2018 methods.

3.4 New Zealand systems

Currently, national scale changes in New Zealand's SOCS for mineral soils are quantified and reported using a relatively standard Tier 2 IPCC approach based on predicted gains/losses associated with changes in land use (McNeill et al. 2014). A simple Tier 1 approach is used for the small area of organic soils. These approaches are well established and provide useful estimates of changes in SOCS at the national scale. However, the system is not based on direct measurements of SOCS change over time, but rather relies on historical SOCS data that were collected over many decades and not with the intent of calibrating a national model (i.e. samples were not from randomly selected locations). The system also requires the assumption that SOCS remain constant if the broad land uses (e.g. high producing grassland) do not change, although recent evidence suggests that specific management regimes within high producing grasslands can impact SOCS (e.g. Mudge et al. 2017).

Key limitations of the current New Zealand soil carbon monitoring system could potentially be eliminated if direct measurements of SOCS change over time were made for different land uses or management regimes. Recently, McNeill et al. (2019) completed statistical analyses (based on historic SOCS data) to determine how many sampling sites would be required for monitoring to detect SOCS changes of different magnitudes for specified target areas within New Zealand's managed agricultural land. Results revealed that a **minimum** of 377 monitoring sites established to represent the broad land uses and management regimes of Cropland (78 sites), Horticulture (92), Dairy (71), Drystock on flat-rolling land (76), and Drystock in hill country (60) would enable detection of SOCS changes of 2 $Mg.ha^{-1}$ within each target area (on average across all sample sites), should changes of this magnitude occur between monitoring cycles (recommended to be 3–5 years apart). This analysis revealed that it is tractable to directly monitor changes in SOCS under broad land use and management regimes in New Zealand at the national scale. A subsequent project will explore how SOC sampling design can be optimised for synergies across multiple scales (e.g. national, industry and farm) and how the resulting data (and physical samples) can be stored, harmonised and shared efficiently for analysis, modelling and accounting purposes.

3.5 FAO-LEAP report

The FAO-LEAP guidelines for 'Measuring and modelling soil carbon stocks and stock changes in livestock production systems' (FAO 2019) provide a set of requirements, recommendations, and options for the layout, sampling, and quantifying soil carbon stocks and stock changes, and also provides supporting discussion and examples of the likely effect of farm management (e.g. land use intensity, fertilisation) and climate (e.g. annual rainfall), on SOCS and stock changes. The guidelines cover standard direct soil sampling and modelling techniques, but also include sections on how to determine change in SOCS by quantifying the full carbon budget of a site (e.g. with the aid of the eddy covariance technique) and how changes in SOCS could potentially be integrated into life cycle assessment (LCA) of goods and services.

The FAO (2019) direct soil sampling guidelines provide general advice concerning sampling design, but they do not prescribe a sampling scheme. For example, a pre-sampling of the area of interest is suggested to estimate the mean SOCS and variability, which can then be used to estimate the sample size, but other information can be used to determine the sampling effort, such as modelling or assumed values. While this provides some flexibility for a given farm to implement, there is a risk that different methods might produce different results. Moreover, the FAO (2019) guidelines suggest a sampling approach where the baseline SOCS is established at some date, and then the change is measured by repeating the measurements at the same site after a fixed time period (e.g. 5 or 10 years). As pointed out by Lark (2009), this approach provides the best (i.e. most accurate) estimate of SOCS at the start and end of the measurement period but does not necessarily produce the best estimate of SOCS change. Slightly more accurate results can be obtained by regular measurements over the five or ten-year period, albeit with an increase in field effort (and therefore, cost). As a result, there is a risk that using FAO (2019) recommendations will produce an estimate of SOCS change that is less accurate but also less expensive than an approach designed to optimise accuracy.

The full carbon budget approach based on the eddy covariance technique was deemed to be too expensive for routine monitoring and is largely used for research purposes and to calibrate/validate models (e.g. Kirschbaum et al. 2017). It was concluded that changes in SOCS could readily be incorporated into LCA once consensus on the methods for estimating and reporting changes was reached.

3.6 VERRA

VERRA's 'Verified Carbon Standard' is the world's most widely used voluntary GHG programme and has standards for quantifying different types of GHG emissions across many sectors, including soil carbon in agricultural systems (VERRA 2018). VERRA's Soil Carbon Quantification Methodology (VM0021) was developed in 2012 and is part of a modular system with different modules providing specific methods for parts of the benchmarking and monitoring process such as determining project boundaries (VMD0020), stratification (VMD0018), quantifying soil carbon stocks (VMD0021) and also determining the net change in GHG emissions resulting from 'project activities' (VMD0035).

In general, the VERRA soil carbon method (VM0021) follows similar methods to those outlined in FAO (2019) and the 2018 Australian method, but with a few notable differences. First, the VERRA method suggests taking separate soil samples for SOC and bulk density analysis, while the other two methods explicitly recommend that both these variables be determined on the same soil sample. VERRA also makes no provision for use of newer sensor technologies, while there is a requirement to have permanent sampling plots marked with an in situ relocatable marker (i.e. metal stake). Overall, the VERRA Soil Carbon Quantification Methodology (VM0021) provides a relatively robust system to quantify changes in SOCS and net GHG emissions associated with implementation of a new land management 'project' and provides a pathway to potentially acquire 'voluntary' carbon credits should New Zealand farmers want to pursue this option.

3.7 Conclusions – guiding principles for NZ farm-scale design

- Both Canadian systems have a cropping focus that differs from this report, which is focussed on pastoral systems
- We believe there is insufficient data available in New Zealand to parameterise a model (e.g. the Alberta Offset System or the FullCAM model in Australia) for accurate prediction of changes in SOCS for specific individual farms across New Zealand (due to the diversity of climate, soil type and management regimes).

- Direct measurement methods are generally preferable because of their flexibility to account for emerging management practices at the farm scale, independence of established management assumptions relating to carbon fluxes, provision of site-specific information to the landholder and the generation of robust data pertaining to SOCS change. However, implementation costs will likely be higher than modelling approaches (see section 5.6).
- The direct SOCS measurement systems outlined in the FAO, VERRA and Australian 2018 methods are all similar. For example, all require stratification of the project area and follow relatively standard methods for soil sampling, processing, and data analysis.
- **We believe the Australian 2018 methodology is the most relevant to farm-scale SOCS measurements in New Zealand and is robust, having been implemented in national legislation and well supported by recent literature**. Based on this assertion we use the Australian 2018 method to guide the approach recommended in New Zealand, while at the same time considering the FAO 2019 and VERRA methods, plus other literature, and any unique issues specific to New Zealand conditions.
- We will also endeavour to condense the recommended approach into a succinct summary so farmers, rural professionals and industry representatives can quickly get an idea of what is required for SOCS benchmarking and monitoring.

4 Technical discussion of options for SOC benchmarking and monitoring at farm-scale

This section provides a technical description and discussion of the issues and options for farm-scale SOC benchmarking and monitoring. The description first looks at the measurement aspects underpinning the quantification of soil carbon stocks. Second, the sampling aspects of a soil carbon auditing project are summarised. Most sections have a very brief summary outlining the key points and recommendations which will be included in Section 5.

4.1 What are we measuring?

Using relatively simple notation adapted from Poeplau et al. (2017), and also adapted into the Australian 2018 method, SOCS at a specific point can be measured using the following equation:

$$\mathrm{SOC_{stock}} = \mathrm{SOC_{conc,\,fine\,soil}} \times \mathrm{BD_{whole\,soil}} \times \mathrm{Thickness} \times \left(1 - \mathrm{MassFraction_{gravel}}\right) \quad \text{Equation 1}$$

$\mathrm{SOC_{stock}}$ is the SOCS of the investigated soil over a given soil thickness. The units of $\mathrm{SOC_{stock}}$ are conventionally in Mg C ha^{-1} to a given fixed depth or a specified cumulative mass of soil (more on this further on). $\mathrm{SOC_{conc,fine\,soil}}$ is the gravimetric content of SOC in the fine soil expressed on an oven dry equivalent basis (%), where the fine soil refers to all dried and ground soil material of the collected sample that passes through a 2 mm sieve. $\mathrm{BD_{whole\,soil}}$ is the corresponding dry bulk density of the whole soil, and $\mathrm{MassFraction_{gravel}}$ corresponds to the mass fraction of the sample that does not pass through a 2 mm sieve (i.e. the gravel and rock fraction). Note that this equation negates the need to estimate both the volume fraction of the gravel and rocks, and the bulk density of the fine soil fraction. If we were to measure these then $\mathrm{SOC_{stock}}$ would be measured using:

$$\mathrm{SOC_{stock}} = \mathrm{SOC_{conc,\,fine\,soil}} \times \mathrm{BD_{whole\,soil}} \times \mathrm{Thickness} \times (1 - \mathrm{rock\ fragments\ fraction}) \quad \text{Equation 2}$$

where $BD_{fine\ soil}$ is calculated using the following equation:

$$\mathrm{BD_{fine\,soil}} = \frac{\mathrm{mass_{sample}} - \mathrm{mass_{rock\,fragments}}}{volume_{sample} - \frac{mass_{rock\ fragments}}{\rho_{rock\ fragments}}} \quad \text{Equation 3}$$

$mass_{sample}$ and $mass_{rock\ fragments}$ correspond to the total masses of the oven dried sample, and rock fragments respectively. $volume_{sample}$ is the total volume of the sample (cm^3), and $\rho_{rock\ fragments}$ is the rock fragments density which is commonly found to be 2.65 g cm^{-1}. Of course, measurement of rock density could be considered, where methods involving the use of a hydrostatic scale (which is based on Archimedes principle) could be used instead of settling with the common value of 2.65 g cm^{-1} (Mehler et al. 2014).

A simplification of Equation 1 and 2 as shown in Poeplau et al. (2017) is:

$$MS = \frac{mass_{fine\ soil}}{volume_{sample}} \times thickness$$

$$SOCstock = SOCcon_{fine\ soil} \times MS$$ Equation 4

where MS here is the fine soil mass (< 2 mm) in Mg.ha^{-1}.

Equations 1, 2, and 4 assume that the $SOCcon_{fine\ soil}$ is expressed on an oven-dry equivalent basis where the value derived by the analyser has been corrected for water content (θm) of air-dried soil, otherwise the full equation becomes:

$$SOCstock = \left(SOCcon_{fine\ soil} \times (1 + \theta m)\right) \times BD_{whole\ soil} \times Thickness \times \left(1 - MassFraction_{gravel}\right)$$

Equation 5

For reference, the following version of the above equation also gives all the units and correction factor for units

$$\begin{array}{c}\text{Soil carbon}\\ \text{stocks}\\ \text{(Mg C/ha)}\end{array} = \left[\begin{array}{c}\text{Carbon}\\ \text{content}\\ \text{(g C/kg AD soil)}\end{array} \times (1+\theta_m)\right] \times \begin{array}{c}\text{Soil}\\ \text{layer}\\ \text{thickness}\\ \text{(cm)}\end{array} \times \begin{array}{c}\text{Bulk}\\ \text{density}\\ \text{(Mg/m}^3\text{)}\end{array} \times \left[1 - \begin{array}{c}\text{Mass fraction}\\ \text{of gravel}\\ \text{(>2mm)}\end{array}\right] \times 0.10$$

Equation 6

4.1.1 Fixed depth vs. Equivalent Mass

Historically, SOCS have been quantified to a fixed depth. For example, the common reporting depth as set by Intergovernmental Panel on Climate Change (IPCC, 2006) is 0–30 cm. Such fixed-depth practices are in place throughout the world. It is widely accepted now that using fixed depths from an auditing standpoint is problematic because changes in management (or any other change in the system) may change the soil bulk density that in turn will lead to different soil masses if a fixed depth is used (von Haden et al. 2020; Wendt & Hauser 2013). Basically, there is no correspondence in comparing SOCS values based on different soil masses. If pursued in this way, such comparisons will lead to misleading reporting of improved or lost SOCS when in fact the opposite may be true or there has been no change at all. Using the equivalent soil mass (ESM) approach largely solves this issue.

ESM can be derived by adjusting SOCS based on changes in bulk density (see example at: https://www.agric.wa.gov.au/soil-carbon/measuring-and-reporting-soil-organic-carbon). This method would work if there is an increase in bulk density, where the depth of soil required for ESM for the follow up round of sampling would be less than that at the previous round of sampling. If bulk density has decreased however, a bulk density correction would give misleading outcomes because the measured soil carbon only corresponds to the depth of measurement and not actually to the ESM. The Australian 2014 method demonstrates a practicable approach to avoiding this issue, where an ESM is selected after the sampling has occurred to a set depth. In this case, sampling is conducted to 30 cm and soil masses of all the collected samples are determined, followed by a setting of the ESM which was determined to be the lower 10th percentile of all soil masses. This ensures that in 90% of cases, no extrapolation of SOCS occurs on unmeasured soil material. If the mean of the ESM was used, then 50% of the samples would be subjected to extrapolation which is not a desirable outcome. Once the ESM is set, estimation of SOCS ensues for each depth that is then corrected for the ESM. These correction equations are detailed in the 2014 method.

Alternatively, the SOCS at the given ESM can be achieved by fitting a depth function such as a cubic spline through the corresponding soil masses and SOCS that were measured at given depth intervals (see figure 2, Wendt & Hauser 2013). Implementing such a method will therefore require several measurements to be made in sequential depth increments in order to fit an appropriate cubic spline function. This could be a barrier where for cost and efficiency no more than two depths are sampled for SOCS measurement.

Summary:

1 Soil carbon stocks should be calculated on an equivalent soil mass (ESM) basis.
2 An ESM should be set after the initial baseline sample, and a conservative target is set (10th percentile) that is then adhered to throughout the life of the monitoring program.

4.1.2 Estimation of spatial mean SOC stock

If the aim is to measure the mean SOCS in a farm, then in most cases we are not interested in SOCS at one point in the landscape, but rather the spatial mean SOCS over some extended area (the CEA, such as a paddock, farm block or whole farm). Direct measurement of SOCS therefore requires some form of representative sampling from within the CEA. In this sense, a representative sample is designed to accurately reflect the characteristics of the wider CEA. For instance, if a whole farm is being sampled, then points from all over the farm are required covering the range of terrain and land use that are likely to be encountered.

There are two general approaches in sampling: design-based and model-based (Brus & de Gruijter 1997; de Gruijter et al. 2006). Design-based sampling is ideal for deriving global estimates, which for soil carbon auditing is the mean soil carbon stock in a CEA. In design-based systems, sampling locations are selected by probability sampling (random sampling with known inclusion probabilities), and the inference associated with these data (estimation of stocks) is based on the sampling design process used to select the sampling locations. Model-based sampling on the other hand is useful for applications where one may want to create a map of soil carbon stocks across a CEA. The main priority in model-based sampling is to select samples that optimise some criteria such as minimisation of a kriging variance, optimising spatial or geographical coverage (or both), each of which adds an element of purposiveness to the exercise, and thereby excludes design-based inference using this approach. Despite these limitations in model-based sampling, design-based sampling can be used for model-based inference, such as for mapping. According to Viscarra Rossel et al. (2016) there is no published literature on sampling methods that are designed for both design- and model-based estimation.

Sampling site placement

Once a general approach to sampling has been adopted (design- or model-based), a method of determining the actual spatial placement of sampling locations needs to be considered. The key requirement is that sample locations are chosen by random selection rather than purposefully selected for convenience. Random selection of soil sampling sites ensures that SOCS estimates are unbiased.

It is possible to randomly allocate sampling sites over the whole CEA (e.g. a farm), but this is only likely to be a useful approach if the distribution of SOCS is homogenous over the CEA. It is more likely that some areas of the farm will have higher SOCS while others will have lower values depending on soil topography and land management. For example, Figure 1 shows the distribution of 0–10 cm SOCS by soil order (from the New Zealand Soil Classification), showing that Allophanic soils generally have higher SOCS when compared with Brown soils. If a CEA with both Allophanic and Brown soils needed to be sampled, combining these two soil orders would require more samples (i.e. the sampling would be of lower efficiency) than if the Allophanic and Brown soils were sampled separately. For this reason, the first step before sampling is to decide whether stratification of the CEA is required, and if so what the basis for stratification should be (e.g. by soil type, slope class and management). If performed appropriately, stratification will lead to higher precision of mean SOCS (lower sampling variance), and lower cost of estimation when compared with random sampling without stratification. In practice, the degree of stratification that is used is likely to be a compromise between many strata to represent all possible factors important in a CEA, and the additional effort required to sample each of the adopted strata. Options for stratification are discussed in Section 4.3.1.

Once the degree of stratification has been decided, each stratum must be sampled using some form of random sampling approach. The simplest method of probability sampling within a stratum (de Gruijter et al. 2016) is to use Simple Random Sampling (SRS), where a fixed number of sample points is drawn at random and independently from each other within the stratum. When combined with stratification, this sampling scheme becomes stratified simple random sampling (StSRS).

In the simple case where there is no stratification of the CEA, the sampling scheme reduces to simple random sampling (SRS) of the whole target area. We note that SRS without any stratification is seldom used in practice, since SOCS tend to vary by slope class and by soil type, for instance, so there is generally some advantage in stratification.

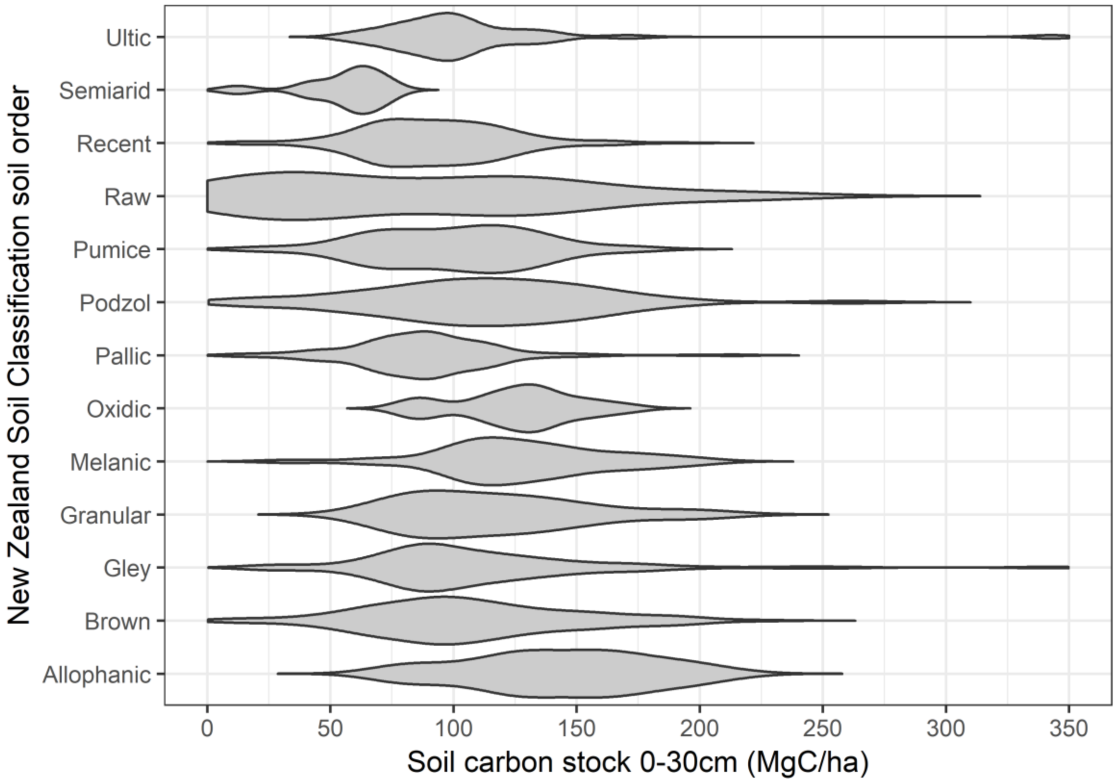

Figure 1: Violin plot of the 0–30-cm soil carbon stocks over New Zealand grassland soils, grouped by soil order from the New Zealand Soil Classification. The width of the violin plot shows the relative size of the distribution of soil carbon.

StSRS of a CEA has lower efficiency when compared with other possible sampling methods, but the method is easy to understand, simple to implement, and widely used. StSRS is the approach adopted in the Australian 2018 Method, where it is compulsory to have at least three strata and a minimum of three samples from each stratum. However, it is indicated that the use of additional strata and collection of additional samples would be beneficial.

One disadvantage with using StSRS is that the selection of samples does not consider the presence of other nearby samples, so there is a tendency for samples to be closely spaced, and the CEA is less efficiently covered (McNeill et al. 2018). Another general approach to sampling within each stratum is to force the random samples to be spread across the stratum so that all areas within the stratum are covered, while still maintaining randomness. This general class of methods is known as balanced sampling methods. Two common balanced sampling methods are Generalized Random Tessellation Stratified sampling or GRTS (Stevens & Olsen 2004) and Balanced Acceptance Sampling or BAS (Robertson et al. 2013). Combining balanced sampling with stratification produces stratified balanced sampling.

Another common method of spatial sampling is to generate a regular or systematic grid and retain samples where they happen to lie within the stratum. This method (grid sampling) is simple to understand, since all that is required is a random starting point and selection of the regular sampling distance. However, the method has several important disadvantages:

1. Grid-based sampling selects points based on a random initial point and an adopted grid sampling distance. This approach is incompatible with the design-based sampling approach that we advocate (de Gruijter et al. 2006).
2. Since grid sampling cannot have two points located closer than the grid sampling distance, this method of sampling cannot recover SOCS at a spatial scale finer than the size of the sampling grid.

For the above reasons, we do not recommend systematic or grid sampling. An example of four common stratum sampling methods is shown in Figure 2, where 100 samples are allocated within a single stratum using SRS, GRTS, BAS, and systematic or grid. These examples can be easily expanded to cover more than one stratum.

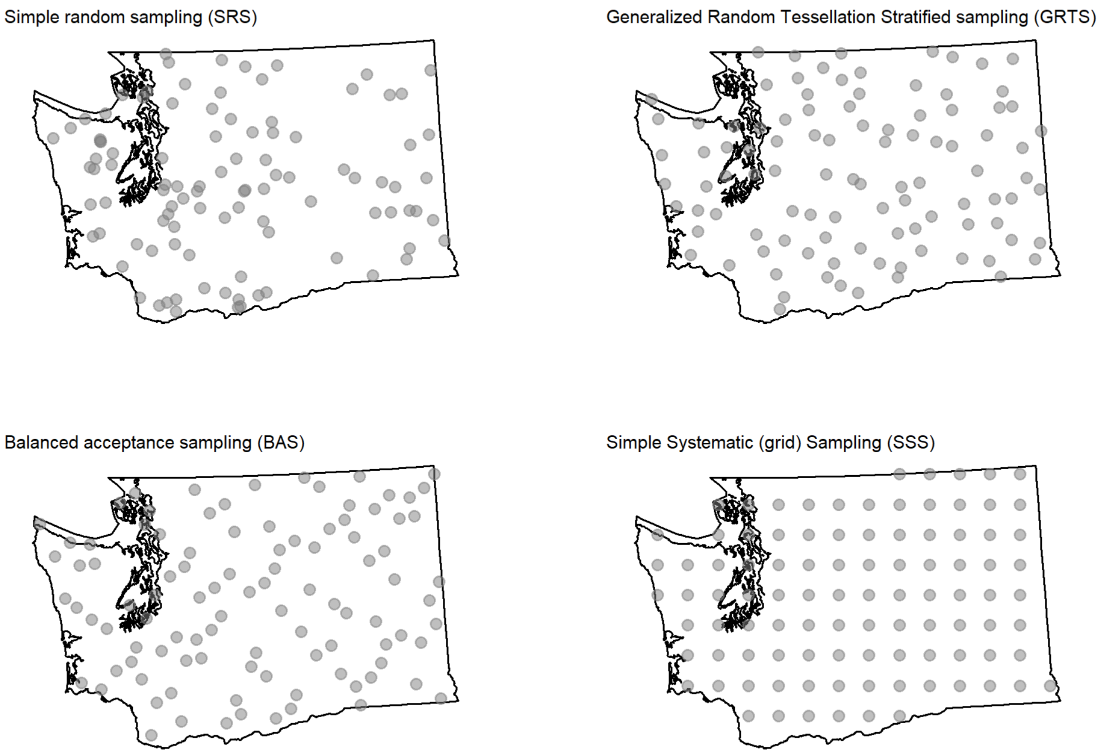

Figure 2: Example of spatial sampling of a single stratum with 100 samples using simple random sampling, two variants of balanced sampling (GRTS and BAS), and systematic or grid sampling.

Once the degree of stratification has been decided, the choice between SRS, balanced sampling, and other approaches represents a pragmatic compromise between efficiency, ease of understanding, and practicality. There is no doubt that SRS is the easiest method to understand, but it has lower efficiency compared with other methods. Variants of balanced sampling tend to be highly efficient (Benedetti et al. 2017), especially for small sample sizes, but balanced sampling is more conceptually difficult to describe. The cost of the field effort required to estimate SOCS to a specified accuracy is broadly proportional to the number of field samples, and the number of samples required is also inversely proportional to sampling efficiency (i.e. a more efficient sampling scheme requires fewer samples to achieve a pre-specified estimated SOCS accuracy when compared with a less efficient scheme).

Summary:

1 Design-based approaches are the simplest and most common approach used for SOCS benchmarking (measuring mean SOCS) and monitoring (measuring SOCS change).
2 Stratification will likely increase the accuracy of mean SOCS estimates within a CEA (i.e. for benchmarking), but the degree and type of stratification depends on the farm. Examples include stratification by different combinations of landform, slope class, soil order and management.
3 In terms of specific sampling site placement within a stratum, simple random sampling (SRS, or StSRS within stratification) is the simplest and most flexible method, but balanced acceptance sampling will be more efficient, although slightly more complicated.

4.1.3 Analysing individual vs composite soil samples

Once stratification and the method to determine sampling site locations are determined, there are two ways of deriving an estimate of the spatial mean SOCS: single samples, and composite or bulked samples. The primary purpose of compositing is to reduce the number of samples that need to be analysed (reducing laboratory costs), while still capturing the spatial variability of SOCS in the estimate of the spatial mean. However, in addition to estimating the spatial mean, one also wants to know the variance about that estimate to be able to communicate the certainty of the estimate. If all samples collected randomly from within a stratum sample are bulked, that effectively leaves only one sample to measure from the stratum, and therefore no way to derive a measure of variance. Therefore, more than one composite sample is essential, and these are formed by bulking an equal number of randomly selected (pre-chosen) samples, either within or across strata. Bulking within strata is where short range variability can be dealt with where soil samples within an area are collected then composited (Lark 2012). Within a stratum, several composited samples are collected and can be treated as either individual samples or used as further composites across strata.

As pointed out in de Gruijter et al. (2016), bulking (compositing) is a common practice in sampling for soil testing, with the basic assumption that soil carbon is additive and analysing a bulk sample gives the same mean result as averaging the values of the individual samples. Bulking across evenly sized strata has been proposed for probability sampling in SOCS estimation by Chappell et al. (2013), which contributed to the recommendations from the Australian 2014 Method. However, it is important to note that one of the criteria of the 2014 method was that it could be applied without any prior knowledge of the spatial distribution of SOCS across the area. In the Australian 2018 method, the requirement of compositing soil samples was relaxed to additionally include auditing using individual samples to be collected from unequal area strata formed on the basis of prior information indicative of the spatial variation in SOCS.

One major advantage of analysing individual samples separately is that the resulting data can be used to refine stratification and sample numbers for subsequent sampling rounds, provided there are sufficient samples to do so. For example, based on prior SOC data for a farm, de Gruijter et al. (2016) calculated that 127 was the optimal sampling number (based on costs and potential returns from SOC sequestration), but after re-stratifying based on data from 50 points for the baseline sampling, only 22 sampling sites were required to achieve the same result. de Gruijter et al. (2016) point out that one advantage of compositing is cost reduction, and at the same time pointed out disadvantages related to an increase the contribution of measurement errors to the total and that fact that sample sizes (total and within strata) can no longer be chosen freely. For example, in a hypothetical case of compositing across 10 strata, the total sample size must be a multiple of 10. This implies that the sampling variance cannot be minimized as effectively as in non-composite sampling. A few other technical disadvantages were identified, but de Gruijter et al. (2016) firmly believed that once sampling and measurement costs and then the measurement errors of spatial variability of SOC were taken into account, the advantages of compositing do not outweigh the possible disadvantages. In section 5.4.5, we provide the relevant equations for estimation of spatial means using both individual and composite samples.

Summary:

1 If resources (i.e. time and money) are not limited, it is best to not composite samples, because this provides the richest and most flexible data.

2 If resources are limited, analysing composite samples will likely be the best option. However, there is likely to be a trade-off between reduced costs in the first sampling round against the potential reduction in sample numbers in subsequent rounds due to refined sampling design if individual samples are analysed.
3 Section 5.6 provides cost estimates for the two different options

4.1.4 Sampling in both space and time

Incorporating the temporal component into a benchmarking and monitoring program necessitates not only addressing the question of where to sample, but also when. Or in other words, some form of space-time sampling is required. From a monitoring perspective, landholders could potentially be credited for positive increases in stocks from their established baseline stocks. Intuitively and perhaps naively (without considering spatial and temporal correlation between sample times), one would want to establish around three measurement time points (inclusive of the baseline) where stocks are measured before establishing a linear trend line to quantify the rate of SOCS increase (or decrease). Building on this intuition, two time points would allow one to establish an initial estimate of change, and perhaps a coarse estimate of trend, but with three or more time points, one is able to estimate the trend with better precision. As established in de Gruijter et al. (2006), there are several basic types of space-time designs. We will discuss a few of these that are relevant to SOCS monitoring in Section 4.3.3, but it is important to note that sampling to estimate the spatial mean through time and to estimate the temporal trend are differing goals that warrant perhaps different space-time or hybridised space-time sampling approaches to achieve each in an optimal sense (Brus & de Gruijter 2011, 2014; Chappell et al. 2013).

4.2 Specific measurement details

This sub-section will provide a short summary of the analytical methods available for the measurement of the components used to estimate SOCS: soil carbon concentration and bulk density. Although not explicitly required we also discuss the methods for measuring rock and gravel content.

4.2.1 Soil carbon concentration

Soil carbon concentration can either be directly measured or inferred using spectroscopy techniques that rely on a calibration relationship to be formed between direct measurements and the associated spectra of samples. The most common spectroscopy techniques for SOCS acquire spectra within the visible, near and mid-infrared regions of the electromagnetic spectrum.

Several analytical approaches to soil carbon estimation exist and can be distinguished as either wet or dumas oxidation methods. Wet oxidation methods include the Walkley-Black method (Walkley and Black 1934) and its derivatives (Heanes 1984). Dumas oxidation or dry combustion methods use purpose-built analysers such as those manufactured by LECO (https://www.leco.com/) or Elementar (https://www.elementar.de/en.html). There are relative advantages and disadvantages to each analytical method (Nelson & Sommers 1982), although Conyers et al. (2011) recommended against using Walkley-Black due to measurement variability issues, but found the Heanes and Leco methods to be equivalent in their capture of SOC. The Dumas oxidation methods also capture the inorganic component of soil C, whereas the Heanes method does not. Therefore, for Dumas oxidation pre-treatment of carbonate containing soils is required to remove the inorganic C from the soil samples before analysis to measure only SOC. From a practical viewpoint, dumas methods are preferred over wet oxidation methods because they are procedurally simpler and require less interaction with potentially dangerous chemicals.

Increasingly, alternative ways to estimate soil carbon concentration are being developped (Smith et al., 2020). Some expansive reviews on the use of visible near- infrared spectroscopy to predict soil properties and, in particular, soil carbon concentration exist, with important contributions being Bellon-Maurel and McBratney (2011), Soriano-Disla et al. (2014), and Viscarra Rossel et al. (2006). Recently, mid-infrared spectroscopy has also been suggested (Baldock et al, 2014, Dangal et al, 2020). The accuracy of these methods to predict soil carbon concentration is contingent upon the quality of the calibration model and the validity of the contributing datasets (spectral libraries) that went into constructing the model. While a soil spectral inference model could exhibit high predictive skill based on goodness-of-fit statistics with reference to the model calibration data and even an independent validation dataset, if the spectral library used for calibration does not represent the local variability of soil carbon in the target universe, one will incur misleading values, and ultimately a high difficulty in assessing whether SOCs are increasing or not. The scanning protocol, spectral library and chemometric procedures are key to prediction, and the relative advantages of 'globally' calibrated vs 'locally' calibrated spectral inference models is widely discussed (e.g. Lobsey et al. 2017). Method development in the local calibration model from large spectral libraries (e.g. Ramirez-Lopez et al. 2013, Shi et al. 2015) provide compelling insights and evidence to indicate that global models often fail to capture accurately the local, site-specific characteristics of soil. The Australian Govt 2018 method is cognisant of this issue and stipulates that spectrally inferred SOC estimates be derived from locally calibrated models. For example, the method states that it is a minimum data requirement to have either, a) 40 soil samples with spectra when the total number of samples in an auditing area is ≤ 200 or, b) 20% of the total number of soil samples with spectra in an auditing area when the total number of samples in an area is ≥ 200. Alternatively, a combinatorial approach using both local spectral libraries as well as spectra extracted from compiled libraries for more robust predictions. An algorithm such as the RS-Local (Lobsey et al. 2017) is an example of one such method that exploits this combinatorial modelling approach.

The type of spectroscopy used also has a dramatic effect on the prediction performance. Mid-infrared spectroscopy typically yields superior results, and with a large enough library can challenge the reference measurement techniques (Dangal et al, 2020). Visible near-infrared, while still quantitative, has higher levels of uncertainty associated with its predictions. However, the processing costs associated with Vis-NIR are much lower than those required for MIR. Recently, authors have also suggested both approaches can be combined (Ng et al., 2019).

Soil matrix effects, and soil moisture are a source of heterogeneity that if not contained will affect the quality of soil spectral predictions (Ge et al. 2014, Roudier et al. 2017). Managing these issues is largely dealt with by drying the soils and grinding then to a consistent size fraction, then subjecting the samples for infrared analysis in a controlled laboratory setting. The popularity of soil infrared spectroscopy is largely built around the experience that such techniques are more efficient in time and cost terms than traditional wet chemistry measurements. Taking the instruments into the field is yet another time- and cost-saving idea, and method development in this space continues to evolve to address the issues that affect the quality of measurements (e.g. variations in soil water content, matrix effects). In particular, methods such as external parameter orthogonalisation (Minasny et al. 2011) and direct standardisation (Ji et al. 2015) that were designed for other purposes, can be adapted to soil spectral model inference applications, to correct for or remove the attenuation affect that water imparts on collected infrared spectra of soil. Of course, the capability to achieve such a transformation or standardisation is contingent on purpose-built soil spectral libraries, where spectra have been collected from soil in both field and air-dried/ground conditions. If such standardisation is possible, then commissioning spectral libraries that have been built using air dried/ground soils can be used for spectral inference of soil properties on soils in their field condition. By all accounts, treatment of the field-collected spectra, to make them appear as if they were collected in a laboratory setting, does result in improved predictions of soil properties (Minasny et al. 2011).

Despite the potential benefits of Vis-NIR infrared spectroscopy and its potential adaptation specifically in a SOC auditing and monitoring programme, widespread adoption has yet to be realised outside research institutions in Australia, although in New Zealand commercial laboratories are employing these techniques. Research into field application of these measurement instruments is also accelerating e.g. (Hedley et al. 2015, Roudier et al. 2015, Viscarra Rossel et al, 2017). There are continued developments that suggest we are at the beginnings of operationalisation and mainstream application. First, in regard to the technology, there are more manufacturers in the game, which, together with lowering component costs, means business competition will increase, leading to a decrease in capital investment costs for the end-user (Tang et al., 2020). Second, the continued developments in portability and robustness of the technology will ensure a reasonable life span and robustness needed for in-field applications. Moreover, the miniaturisation of the technology will see more units manufactured at cheaper costs and on-boarded with existing technologies such as mobile phones and generic sensing probes (McLaughlin et al. 2017; Wang et al. 2019). All the while, continued improvements in computation, modelling, and data processing, in particular from the soil science and soil spectroscopy, together with the building of representative soil spectral libraries, show great potential for future adoption of these methods.

4.2.2 Bulk Density

As is well practised and established, a known volume of soil is extracted using a coring device, then the oven dried mass of that soil divided by the volume of the core will provide a measure of bulk density that is expressed in terms of g/cm^3 or Mg/m^3. Relative to other measurable soil properties, bulk density is much less frequently measured (Searle 2014) due to the time and effort necessary to perform the measurement, particularly at some depth below the soil surface, as this would generally require a significant amount of excavation to be able to insert and extract the soil coring implement.

Hydraulic and mechanical coring machines can generally insert a probe to a depth of 1m or more without the need for soil excavation. Segments of the soil core can then be measured for bulk density via the conventional method. While this approach is from experience, an efficient method, estimates of core volume are difficult to control, particularly when working in dry and crumbly soils, which can obviously lead to imprecise estimates (Malone et al. 2018). There is also anecdotal evidence to suggest that such forceful probe penetration will inadvertently compress the soil from its otherwise original condition, particularly in loamy and clayey soil at the wetter end of available water content, leading to inaccurate estimates of bulk density. Despite this however, the ESM approach will deal with this issue for most soils.

Soil spectral inference modelling has been demonstrated to be an efficient way to estimate bulk density (Moreira et al. 2009). The spectral modelling workflow for bulk density is the same as it would be for any other target soil variable. However, there is some apprehension about spectrally inferring such soil physical quantities as there is no direct relationship between, in this case, soil density and the absorbance features of a given soil spectrum (Minansy et al. 2008). Rather it is more likely that good spectral model results are the outcome of exploiting established indirect relationships between bulk density and soil properties such as soil carbon content, soil texture, cation exchange capacity, and to some extent the clay mineralogy, which we know can be quite well estimated via soil spectral inference. In that regard, the implementation of pedotransfer functions (Bouma 1989) may offer a workable solution to efficient measurement of bulk density as demonstrated in Tranter et al. (2007), Benites et al. (2007), and Rodriguez-Lado et al. (2015).

Non-destructive approaches for bulk density measurement include gamma radiation attenuation methods (Davidson et al. 1963). Purpose-built instruments, such as that used by Holmes et al. (2012), are internally calibrated and provide direct measures once corrections are made to remove the attenuating effect of soil moisture. Lobsey et al. (2016), developed a bulk density sensor using the principals of gamma radiation attenuation. They manually estimated mass attenuation coefficients of soil and water, and developed an independent measure of soil moisture based on vis-NIR spectral inference. With all the required attenuation information, they were able to predict bulk density. In Holmes et al. (2012), the limitation of this technique was the depth of penetration, which had a maximum is 30 cm. Such instruments are commonly used in engineering and road building applications. While their adoption for the method for SOCS measurement concluded that the non-destructive approach makes the effort of bulk density estimation much easier to achieve, the limitation of depth needs resolving if auditing and monitoring is required beyond 30 cm. The technique proposed in Lobsey and Viscarra Rossel (2016) gets around this technical issue. Finally, the performance of these non-destructive techniques in soils containing high amounts of coarse fragments was only really addressed in Holmes et al. (2012), who conclude that in soils with >20% volume proportion of coarse fragments, soil bulk density estimates can be quite uncertain. However, they concluded, at least in their study area, that most of the uncertainty of SOCS estimation was dominated by uncertainty related to measurement of soil carbon concentration, and bulk density uncertainty contributed only a small amount to the overall uncertainty.

4.2.3 Coarse Fragment content

For estimation of SOCS, coarse fragments are removed entirely from the collected sample so that SOC concentration measurement only deals with the fine soil fraction.

Coarse fragments are any materials that do not pass through a 2-mm sieve, and can include both plant root and rock, sand, and gravel materials. It is generally reported that the density or specific gravity of rock materials is 2.65g/cm^3 and for plant roots it is close to 1g/cm^3. Note that for the above equations, plant roots are not explicitly accounted for, but can easily be done by inserting the respective terms, as they are for dealing with the coarse fraction.

A possible solution, at least in terms of addressing the labour intensity involved in the recovery of coarse fragments from a core of soil, was proposed in Viscarra et al. (2016). They developed a modular core sieving system to enable rapid wet-sieving of soil cores. Using manual weighing and image analyses techniques, they were able to report on rock fragment proportions more efficiently than would have been achieved using a conventional recovery approach. This work appears to be still in a research phase, but it represents the first known attempt to make more efficient and automate what continues to be a rate limiting task in SOCS measurement.

Ultimately, gravel only becomes an issue when spectral predictions underpin the derivation of SOCS. This is largely because calibration models for SOC content are based on the fine fraction of soil. If scans are being done in the field, it would become a big issue when gravel reaches a content where it potentially dominates the window of the NIR scan. Some more consideration to this issue needs to be given as it largely seems unsolved.

Summary:

1 At present we recommend using standard laboratory procedures for determining fine soil mass (< 2 mm), coarse fraction content (gravel/stones) and soil carbon content (see Section 5.4 for details).
2 However, sensor techniques are evolving rapidly and provision for their use (subject to strict quality control measures) are already included in the Australian Government 2018 and FAO 2019 methods.

4.3 Sampling design for on-farm SOC monitoring

This section will deal exclusively with the ‘where’ (location) and ‘when’ (time) components of SOCS benchmarking and monitoring. The location component, as summarised earlier, involves design-based sampling and inference, and involves consideration of stratification variations, the intensity of stratification and the intensity of sampling. The time component is the consideration of sampling strategies to detect changes and the trend of SOCS. To some degree, when to sample is largely governed by the existing land management practices and the interaction between edaphic properties (i.e. properties influenced by the soil) and recent weather (e.g. soils with high clay content exposed to significant rainfall). This means that at some periods of the year it may not be possible to get into field to sample. Similarly, SOCS varies between seasons, so it is important to take measurements within a specified window of time to avoid confounding of the SOCS measurement with seasonal variation. These seasonal factors mean that synchronous sampling through time is most suitable for SOCS monitoring.

4.3.1 How many samples and how many strata?

A practical question when deciding to monitor soil carbon on a farm is: how many samples will be needed to detect a specified change of soil carbon in the future? Prior information concerning the distribution of the mean level of soil carbon on the farm is needed to answer this question. In other words, it is impossible to estimate the required sample size without knowing something about the distribution of soil carbon for the farm.

If all farms in New Zealand were the same, then one set of samples (e.g. from a research farm) would suffice to give estimates for the sample size. However, soil carbon for different farms can differ widely in the overall mean level as well as the variation around that mean level. Therefore, prior information concerning the soil carbon distribution must be inferred by carrying out a preliminary field effort to gather some samples of soil carbon in representative locations over the farm. We call this a pilot study. The principal assumption made in this approach is that the distribution of the SOCS derived from the pilot study will be the same (or at least, very similar to) the distribution of the “true” (but unknown) soil carbon distribution over the farm. In practical terms, this assumption means that samples for the pilot study must be randomly spread all over the farm – preselecting the sites on the basis of convenience could lead to a biased estimate of the mean soil carbon for the farm.

How many samples?

The total numbers of samples to take from the CEA is conventionally determined as the largest number that can be afforded. While affordability is a considerable driver of decision making, if not enough samples are taken to provide a reliable estimate of the mean SOCS or the change in mean SOCS, the resulting data may be of limited value. Therefore, a key determinant of the number of samples required is the accuracy or confidence with which the SOCS and stock change needs to be known. This can only be determined by the landowner/manager, but as a general rule, we believe that designing a system whereby repeat measurements of SOCS could detect a change of 2–5% (should such a change occur) would be a good starting point. A 2–5% change in SOC corresponds to a change of about 2 or 5 $Mg.ha^{-1}$ for the top 30 cm of New Zealand's agricultural soils.

A good understanding of the spatial variation of soil carbon is invaluable information to get a sense of how many samples will be needed. There are two different types of problem relevant here:

1. If only a pilot study has been carried out, this will provide an estimate for the mean and standard deviation of SOCS for a small number of samples (perhaps 5–15). Using this information, plus some key assumptions concerning the variability of changes in soil carbon, a power calculation can be used to estimate the number of samples that will be required to detect a pre-specified change in soil carbon stocks. The basis for the power calculation is described elsewhere (Cohen 1988, Bain and Engelhardt 1992). Standard software provides appropriate methods, or a statistical calculator can be used (Faul et al. 2007).
2. If a set of soil carbon estimates has already been gathered from two different dates for a specified number of sites, perhaps because of historical surveys, then a power analysis is not relevant. The issue in this case is to determine the change in soil carbon stocks that can be detected using the paired set of samples. The common statistical tool for achieving this is by estimating the Minimum Detectable Difference (MDD), which is the smallest detectable difference between means when the variation, significance level, statistical power, and sample size are specified (FAO, 2019; Saby et al. 2008; Singh et al. 2012).

Sample size prediction, which is chiefly achieved using a pilot study, depends for its accuracy on certain assumptions concerning SOCS and its change over time. Some of these assumptions are easily verified, such as the spatial variability of SOCS around a specified sampling point – the value can be established by measuring the difference in SOCS between sets of closely spaced points. Other assumptions, such as the variability in the SOCS for a specified location over time, cannot easily be verified, simply because there are few long-term time series of SOCS from which this information could be estimated. In the latter case, reasonable estimates can be provided using historic SOCS measurements from simple estimates of SOCS change, but it is possible that the estimates are imperfect – the resulting sample size estimates will reflect this uncertainty.

Both the power and MDD analysis require adoption of values for the significance level or Type I error rate (sometimes called the significance level) and the power, or 1-"Type II error rate". It is conventional to use a Type I error rate of 0.05, and a power of 0.8 or so (the latter corresponding to a Type II error rate of 0.2), but these figures are strongly dependent on the nature of the problem.

The Type I error rate, or significance level, is the probability of rejecting the null hypothesis given that it is true. In other words, it is the probability, after analysis, of deciding that the true mean SOCS difference is not equal to zero, when in fact there is no difference at all. It is important that this probability is as low as possible since if α is too high the SOCS may be falsely flagged as having changed. A figure of 0.05 or even smaller is appropriate.

The Type II error rate (1-"Power") is the probability of failing to reject the null hypothesis given that it is false. In other words, it is the probability after analysis, of deciding that there is no SOCS difference between the two dates, when in fact there is a difference of the mean between the two dates.

There is usually a trade-off between the Type I and Type II error rates; for a given set of SOCS sample data, reducing the Type I error rate generally results in an increase in the Type II error rate (and consequently a decrease in the power), and vice versa. In the context of detecting whether the mean SOCS has changed, it is desirable to avoid false positive errors, so a low Type I error rate is preferred. Consequently, a higher Type II error rate is usually tolerable. A Type II error rate of as high as 0.2–0.4 may be quite acceptable. We note that there are no firm guidelines for the adoption of Type I and Type II error rates for SOCS change estimation other than convention (FAO, 2019; Saby et al. 2008; Singh et al. 2012).

Grealish et al. (2011) provide simulation outcomes of sample size optimisation of a dataset produced by McKenzie et al. (2002), along with visual plots of the relationship between statistical power and sample size, with sample completeness (i.e. proportion of required samples actually acquired). An important conclusion from this visual summary is that data completeness (i.e. maintaining the required minimum sample size) is critical to maintain the pre-specified power of the study.

de Gruijter et al. (2016) approach the sample size optimisation problem using a Value of Information approach (Morgan et al. 1990). Here, the sample size can be determined to maximize the expected profit for a farmer, by financial quantification of the value of the sample data and the costs to collect the data. Data value depends on the precision of the estimate of SOCS change, as the tradable amount of carbon is determined largely by the certainty of estimates between two-time intervals. A high level of uncertainty means a likely low level of tradable carbon, so sampling should be increased accordingly. To increase the sampling, one also needs to be cognisant of the associated costs. As could be imagined, the maximisation of financial gain entails an iterative exercise under different stratification and sampling number settings. The compelling feature of the value of information approach is that it considers real socio-economic factors such as the market cost of carbon and the dollar amount of collecting a sample, features that are not considered in the MDD approach. But, like the MDD approach, there is an implied requirement of understanding what the variation of soil carbon is within the target universe, and this information is seldom available at the farm scale.

Getting a measure of the spatial variation of SOCS is challenging in circumstances where there is little available data. For example, in the establishment of baseline stocks, there might be next to nothing available to plan for the initial audit. It has been suggested that consulting published literature on soil carbon variograms is useful, and good examples exist by way of Bishop and Pringle (2012) who provided a meta-analysis of paddock scale soil carbon variables. The diversity of the variograms in terms of nugget, sills, and ranges is potentially problematic, however, in terms of site specificity. One may be lucky to find a variogram that was constructed near their CEA, but there is little to go on in terms of the limit of extrapolation about which variograms to use where, so their utility has to be questioned, particularly in a soil auditing context. Existing soil carbon maps of the target universe are invaluable. Their availability while skirting the site-specificity issue, is probably much below that which would be expected, given much growth and activity in digital soil mapping in recent times (Arrouays et al. 2014). Much of the growth in digital sol mapping has been in global, national, and regional programmes, with comparatively less effort (outside of research studies) in field- and farm-scale settings. With the availability of a growing number of national soil mapping products, there have been some efforts to investigate their applicability to farm-scale applications. This work has centred on spatial downscaling and has had mixed outcomes in a soil carbon auditing framework because the uncertainties in the national mapping propagate through the farm-scale mapping, in some cases leading to inappropriate sampling recommendations (Malone et al. 2018).

A general recommendation would be to conduct a pilot survey of the CEA. Ideally, the sampling would be optimised in terms of spatial and/or geographic coverage. Some prior information could be gleaned to inform sample size, or a sampling density of at least *x* samples per unit area could be stipulated. This pilot survey might come at a significant high cost (Chappell et al. 2013), but in the case where no other information is available it is essential and therefore of intangible value for subsequent sampling efforts and ongoing monitoring. In this case the expected benefits of a pilot survey are likely to outweigh the initial costs.

How many strata and how to define strata?

As noted in (de Gruijter et al. 2016), the mean level of SOCS within each subunit tends to be related (correlated), and the standard deviation of each subunit is often less than the standard deviation of the pooled SOCS samples from the whole farm. When this is the case (i.e. the standard deviation of the strata is less than the standard deviation of the SOCS for the farm as a whole), there is an advantage to stratification in terms of sampling efficiency.

Although conceptually simple, there are several potential difficulties with this approach:

1. It is not obvious how the stratification should be decided. Possible choices include stratification by farm plan, by landform, or by slope classes. However, there is no guarantee that the stratification chosen will give an overall reduction in the required minimum number of samples. For example, using Massey Tuapaka as an example, stratification by landform does reduce the estimated SOCS standard deviation within each landform class, but the overall required minimum number of samples is not reduced as a result.
2. The number of pilot samples is increased in the stratification procedure, which increases the cost.
3. Once a method for stratification has been decided, the second set of samples measured some years later must also be stratified in the same way. Therefore, a stratification method must be used that is likely to be stable and relevant over time.

The optimisation of the number of strata is slightly less fraught than that for sample size. This is partly because within each stratum there will be a minimum sample requirement, meaning that increasing the number of strata could have a multiplying effect on the total sample size in the case of equal sized strata. In some instances, such as the Australian 2018 method, a minimum of three strata are required although it is recommended that additional strata are incorporated where possible. In a statistical testing sense, one can go about defining the optimal number of strata using comparative methods against a null hypothesis (no strata). For example, in the case of the commonly used k-means clustering, the total within-cluster sum of squares can be used as a metric to compare and optimise the number of strata. This general concept is at the heart of the approach used in de Gruijter et al. (2016).

Other possible metrics for choosing the number of strata, as well as methods for assessing cluster purity are discussed at length in Odeh et al. (1992). In the special case where the CEA is large and symmetrical, the CEA may be stratified using a grid of strata to ensure equal areal sizes. Where the CEA is an irregular shape, compact geographical stratification is invaluable (Walvoort et al. 2010). Where a suite of environmental covariates is used for stratification, equal sized strata may be generated via generalization of the k-means clustering algorithm. This procedure is implemented in the FuzMe software created by Minasny and McBratney (2002), but to date examples of its application are rare.

It is useful to frame the discussion about stratification as in de Gruijter et al. (2016), which is in terms of what is known about the spatial variation of soil carbon on the farm. For example, if nothing is known of SOCS variability, compact geographical stratification may be implemented. This makes sense because an important motivation is to maximise spatial coverage. The stratification may be improved if there is some ancillary information available that generally describes the environmental spatial variation across the CEA. Examples of such ancillary information include maps of elevation (and subsequent derivatives), remotely sensed data (e.g. NDVI as an indicator of vegetation greenness), yield data, and any other data that may have been collected via proximal soil sensing such as electrical conductivity or gamma radiometrics. With a multivariate suite of ancillary data, stratification would entail a k-means classification procedure. However, it is important to recognise that for the stratification to be useful for quantifying SOCS within the CEA, the variation(s) in the ancillary information must be related to the variation in SOCS. In the absence of such a relationship, imposing a stratification based on that information might lead to a reduced ability to detect changes in SOCS relative to imposing equal area strata.

It is said that stratification is best achieved if a digital soil map of SOCS is available for use as a univariate source of information, for example via the *cum-root-f* method of Dalenius and Hodges (1959) or other approaches mentioned in de Gruijter et al. (2016). It is argued that a digital soil map of SOCS variation embodies the quantitative relationships between covariates (ancillary information) and measured SOCS. With the univariate source of information, stratification of a SOCS map involves locating strata boundaries along the variable distribution. As an aside, it would also be relatively easy to divide the map of soil carbon into spatial equal area size strata by ensuring the number of grid cells belonging to each stratum is kept the same.

When using a SOCS map as a stratifying variable, there is an implied assumption that the digital map is error free. This is incorrect, as it is known that many different sources of uncertainty are propagated through to spatial predictions of phenomena, including soil variables and characteristics. Exploiting digital soil map uncertainties, de Gruijter et al. (2015) proposed an alternative method that can define a strata configuration along the spectrum of compact geographical and univariate SOCS map stratification. They called this method *Ospats,* and it uses a raster of predicted values and associated error variances for deriving the sampling strata. By taking into account the prediction variations, *Ospats* is able to produce stratifications that represent transitions between the 'knowing nothing about SOCS variation' (high mapping uncertainty) and 'knowing a lot about SOCS variation' (low mapping uncertainty) situations. The advantages of the continuum approach to stratification is that the configuration can adapt to inclusions of new data and improvements in soil mapping, ultimately making the SOCS monitoring effort more efficient and cost effective over time. What this approach does not satisfy, like many aspects previously discussed, is the data poor scenario. Essentially, a map with associated prediction variance is a mandatory requirement and is not easy to achieve, at the scale of the CEA. The spatial downscaling approach briefly summarised earlier is one such way of using relatively common national scale maps. The usefulness of this approach, however, is contingent on having relatively accurate national-scale mapping, which is not always guaranteed, given that such maps are created to address questions at spatial scales that do not necessarily correspond to those at the scale of a CEA, which could be a paddock or farm, etc. (Malone et al. 2018).

While analytical methods such as those discussed in the previous paragraphs can be used to determine the number of strata, simpler methods of stratification based on landforms and farm management can often be practical and better suited to data and resource/expertise-poor situations. However, some implied knowledge in terms of where to place strata boundaries and due diligence in the areal calculation of strata are required. Unless they have been split as separate CEAs, stratification boundaries could be based on topographical position or setting, different soil types or lithologies, and in a crude sense, differences in vegetation and land management practices. In New Zealand, farm environment plans (FEP) usually contain maps that break the farm into land management units (LMUs) based on the physical properties of the land and management as outlined in the previous sentence. FEP would therefore provide a good starting point for stratification, although the chosen approach is at the discretion of the landowner or implementer of SOCS benchmarking and monitoring.

As mentioned in Section 4.1.3, the decision whether to analyse individual soil samples or composite samples will impact on whether strata boundaries can be refined through time. Once samples have been composited from across a stratum, boundaries cannot change.

The process for calculating SOCS for individual strata and full CEAs (i.e. farms) is outlined in Section 5.4.5.

Summary:

1 Decisions on whether stratification is desired, and if so, the method used and the number of strata and soil samples are ultimately the decision of landowner/manager and will be influenced by their specific objectives, skills and available resources (e.g. existing spatial data layers and money).
2 If stratification is required, we recommend using farm management units or blocks defined in farm environment plans as a starting point for stratification.
3 Where boundaries for stratification are not clear, automated methods of stratification based on specialized software (e.g. FuzMe or *Ospats*) can be used.
4 There is potential to downscale New Zealand’s national SOCS map and use other finer scale spatial layers to delineate strata (e.g. Malone et al. 2018). However, this process would currently require scientists’ input.

4.3.2 Spatial Trend

This section will describe the sampling considerations for soil monitoring in general. The section builds on the previous discussion of design-based inference of spatial means to include new discussion about inferring the temporal trend of spatial means. In terms of SOCS, our questions include: Are the SOCS increasing, decreasing, or remaining stable? If changing, what is the rate of change? Our interest in changes can be motivated by the fact that SOCS are a tradable commodity, and financial reward is a possible motivator. Or changes can reflect alterations in soil function and building or maintaining SOCS is a way to sustain the overall soil resource. Key contributions by de Gruijter et al. (2006) and Lark (2009) provide deeper insight and guidance about monitoring natural resources in general and are summarised here.

In monitoring SOCS, we have already established that because of coordinating sampling with other land use management practices, a sample for auditing purposes would be more appropriately conducted within a given window of time, for example, before or after seasonal sampling, or the top or bottom end of a cropping cycle, etc. These events preclude randomised sampling through time in favour of synchronous sampling patterns. This opens the question of sampling regularity, which is very much dependent on available resources, but a 1–10-year re-visit time interval is commonly thought to be appropriate for SOCS monitoring (Australian Government 2018; Wheeler 2014).

Although this report specifies temporal sampling with the assumption of two sampling points in time, this does not necessarily have to be the case. In practise, for a given CEA it is worthwhile considering the relationship between how often the CEA is sampled (the sampling frequency) and the period of time over which one would want to define the impact of management on SOCS. For example, if it is important to make definitive statements on SOCS change over (say) five years, then it makes sense to sample the CEA more often over the five-year period, such as three times or preferably more. This increased frequency of sampling ensures that the temporal trend is a true representation of the impact of management. If only two points are used, at the start and end of the interval, the observed changes in SOCS can be affected by exogenous effects (e.g. extended periods of drought) that has an impact on SOCS. By measuring SOCS regularly over time, it is more likely that the measured temporal trend is related to management rather than external factors. Since these external factors can generally not be predicted beforehand, it is difficult to incorporate them into the statistical power design for the required minimum sample size. However, increasing the number of sampling times will increase costs.

4.3.3 Space-time sampling designs

de Gruijter et al (2006) refer to five basic types of synchronous space-time sampling designs:

1. *Independent synchronous (IS)* – where the sampling locations shift over time within the CEA.
2. *Static synchronous (SS)* – where the sampling locations remain fixed over time.
3. *Supplemented panel (SP)* – a compromise between both independent and static synchronous designs (two panels) where only some of the sampling locations are re-visited at the sampling times.
4. *Serially alternating (SA)* – another hybrid design where, after some fixed number of sampling times no new locations are selected, but existing locations are revisited in the same order.
5. *Rotational (R)* – a design method similar to *SA* but like *SP* starts with two or more panels, which are then sequentially rotated out and in again at the consecutive sampling times.

Brus and de Gruijter (2013) provide some guidance on what space-time sampling is best suited to estimation of the mean or the trend, using simulation. This evaluation is framed in the context of the persistence of the spatial pattern of the target variable, such as SOCS, and the performance is evaluated on the basis of the covariance matrix of the parameters of a linear model for the change of the spatial means over time. The general advice is:

1. When there is strong persistence of spatial patterns, *SP* space-time sampling is preferred.
2. When there is no persistence of spatial patterns, *IS* or *SA* can be used for estimation of either the mean or the trend.
3. When there is moderate persistence of spatial patterns, the advice is more complicated:
    a. *IS* is the best method for estimation of the mean,
    b. *SS* is the best method for estimation of the trend,
    c. *SP* or *SA* are good compromise choices when both the mean and trend are required.

What these results generally confirm is that revisiting sample sites reduces the sampling variance, and thus improves the overall precision of the spatial mean estimate of change. This is because by re-visiting the same sites, the temporal correlation is taken into account, thus reducing the variance. Therefore, the *SS* approach of re-visiting and resampling near the same sites should require less samples (and hence cost) to detect a given level of change in SOCS than sampling new independent sites (the IS approach). However, a potential disadvantage of returning to the same sites is an envisaged loss of market confidence if ‘gaming’ of known monitoring sites occurs (Wheeler 2014), since it is important to ensure that changes in SOCS are real and not the product of preferential site treatment (fraud).

It is important to note that these recommendations assume that the spatial pattern of SOCS is known when the space-time sampling method is to be designed. In an ideal world, this would be true, but in practice there might be insufficient information to describe this spatial pattern. In this latter case, hybrid designs, such as SP or SA might be preferred. Alternatively, IS or SS could be used where the mean or trend is required, respectively.

An additional consideration is that the choice among space-time sampling schemes has been made in terms of the variance of the mean or trend components of SOCS, not the total cost. If minimum cost is an additional requirement, then the various designs would need to be evaluated for their cost structure.

On balance, the hybrid *SP* design first, has the ideal statistical properties for a monitoring system and second, to some extent provides some protection against gaming the system, by ensuring that a certain proportion of the samples are randomised across the CEA. Operational choices would include what proportion of the total samples would remain fixed and what spatial sampling designs should be considered. Intuition says that if *SP* is used, then perhaps random sampling (e.g. stratified SRS) should be used for both panels (fixed and changing sample locations). For the fixed sampling portion, a spatial balance could be achieved through compact geographical stratification. For the changing sample location panel, stratification could entertain any of the options described above, including stratification based on the target variable. This option permits the configuration of sampling strata to change, which could be realised by using such algorithms as OSPATS, to ensure maximum efficiency in the sampling effort. In a data poor situation, while the geographical stratification is certainly doable for the fixed sampling location panel, the practitioner could entertain any sort of stratification variable/s that is achievable as long as the areal properties of each strata are recorded and well defined.

Summary:

1 The *static synchronous* approach would be the simplest, cheapest (due to the least number of samples required), and most easily understood method for monitoring changes in SOC stocks at the farm scale. However, this approach could be susceptible to fraudulent practices because SOCS at known sampling locations could be deliberately altered (for example by the addition of manure). This risk could be avoided by implementing the *independent synchronous* approach or the risk reduced with the *supplemented panel* approach, but more sampling sites will likely be required using these approaches. The *static synchronous* approach also precludes the option of refining stratification through time because sampling locations are fixed.
2 The final choice among the various space-time sampling schemes will depend on whether it is important to reduce the variance of the estimated mean SOCS, or the trend in SOCS over time, or resource cost, or some combination of these factors. Resolving the best design in this sense is likely to require simulation, as well as some field information.

4.3.4 Quantifying SOCS changes over time

Given the discussion above, confidence in the statistical inference of the spatial means through time may be increased by considering the temporal autocorrelation. The usual method for estimating the temporal trend of spatial means is by ordinary least squares regression where it was demonstrated in Chappell et al. (2013). From a statistical viewpoint, this approach can only be justified, however, in the case of an *IS* design, where no use is made of temporal autocorrelation between co-located observations at previous sampling times. From a practical standpoint, this approach limits the scope of space-time sample designs that could be entertained.

Where spatial- and/or temporal-correlation is present, one approach that can be used is generalised least squares (GLS), described in detail by Brus and de Gruijter (2011).This versatile approach ensures that, irrespective of the space-time sampling design that is selected, the proper statistical inference is used to process the data. For an experienced practitioner, this novel development provides the necessary flexibility in selection of sampling designs to suit a specific monitoring context. However, this added flexibility comes at a cost in terms of the required knowledge and expertise to fit the GLS model.

Summary:

1 When quantifying changes in SOCS over time, ordinary least squares regression is the simplest approach, but is limited to independent synchronous space-time sampling.
2 GLS can be used for more complex space-time sampling methods, but this approach is more computationally demanding, and requires at least some information on the degree or spatial and/or temporal correlation of SOCS in the CEA.

5 Recommended approaches for farm-scale SOC benchmarking and monitoring in NZ

Here we recommend and document approaches for benchmarking and monitoring SOCS at the farm scale in New Zealand. The recommended approaches are generally consistent with other national and international approaches (i.e. those discussed in sections 3 and 4 of this report) but written in the briefest way possible and adapted to the New Zealand context where relevant. Also included is an outline of the expertise required and a case study of the estimated costs associated with implementation of farm-scale SOCS benchmarking and monitoring in New Zealand.

More detail is provided on the sampling design aspects compared with physical soil sampling and processing methods because these are relatively well defined and documented elsewhere.

5.1 Define objectives for implementing SOCS measurements

It is important to clearly articulate the objectives for implementing a farm-scale SOCS benchmarking and monitoring system and estimate the costs (see Section 5.6) and potential benefits before taking measurements. Going through this process will help confirm the desire to proceed, and the objectives can influence study design and sampling intensity.

For purposes of illustration in these recommendations, we have nominally made the following simple objective that we think is realistic in relation to benchmarking and monitoring SOCS on a pastoral farm in New Zealand:

- To design and implement a benchmarking and monitoring system to detect a SOCS change (gain or loss) of 2–5 $Mg.ha^{-1}$ should such a change occur.

This objective would enable a SOCS change of 1 $Mg.ha^{-1}$ y^{-1} to be detectable after 2–5 years or a change of 0.5 $Mg.ha^{-1}$ y^{-1} to be detectable after 4–10 years. On average, SOCS in New Zealand's managed grasslands are approximately 100 $Mg.ha^{-1}$ in the top 30 cm, so a change of 2 or 5 $Mg.ha^{-1}$ would be a change of about 2 and 5% respectively. For context, differences in average SOCS of between 5 and 10 $Mg.ha^{-1}$ have previously been observed between different grassland management regimes in New Zealand (Barnett et al. 2014; Mudge et al. 2017) and observed average SOCS changes over time in some grassland soils across New Zealand have been between 0 and 1 $Mg.ha^{-1}$ y^{-1} (Parfitt et al. 2014; Schipper et al. 2014). Note that generally we are more interested in obtaining an accurate estimate of the change in SOCS rather than the mean SOCS at any given point in time.

Attempting to determine a change in SOCS as small as 2–5% means that any unnecessary errors with sampling, soil preparation and analysis can easily mask detection of true changes. Consequently, we recommend very close attention is given to all steps of the sampling design described below and will necessitate that soil sampling, processing and analysis is implemented by well-trained technical staff.

5.2 Study design

5.2.1 Define and delineate the CEA

An important step in the study design process is to decide on and then clearly define and delineate the CEA as follows:

1. Decide whether to implement SOCS benchmarking and monitoring on the whole farm, or a selected area of the farm.
2. Produce an accurate map of the CEA excluding areas where SOC will not be measured (e.g. buildings, wetlands, forested areas, races):
   a. The map should be produced in a Geographic Information System (GIS) programme to enable accurate spatial quantification of different areas of the farm (e.g. strata). This will also help with the random generation and positioning of soil sampling sites.

5.2.2 Stratification

Stratification is used to delineate areas (strata) *within* the CEA (i.e. a farm) where SOCS and stock changes will be relatively uniform and usually helps enable more precise estimates of SOCS and stock changes for a given sampling effort (and cost). If the whole CEA is homogenous (e.g. in terms of soils, topography, and management) then stratification may not be required. Historic land use should also be considered because this could strongly influence SOCS and stock changes (e.g. previous cropping would have likely reduced SOCS which could increase rapidly under permanent pasture).

There are a number of different ways to stratify the CEA, but here we recommend two general approaches: 1) a relatively simple approach for when existing data, funding, and skills are limited, and 2) a more complex approach that relies on prior knowledge of SOCS and multiple spatial data layers. If the whole CEA is homogenous but stratification is still desired (e.g. to determine potential change in SOCS for different parts of the farm) then strata could simply be defined as adjacent paddocks or blocks of paddocks (called 'compact geographical stratification'). There is potential to refine stratification following a pilot study (see Section 5.2.4) and following the benchmarking sampling, but this requires that soil samples are not composited so each spatial sampling location has an associated SOCS value (see Section 5.2.5).

Simple stratification based on farm environment plans

Farm environment plans (FEP) are already required in most regions of New Zealand as part of Regional Council regulations (e.g. Canterbury Water 2019; Beef and Lamb n.d.). FEPs generally require a relatively detailed map that breaks the whole farm into different land management units (LMUs) based on the physical properties of the land/climate (e.g. soils, slope, aspect) and land management (e.g. effluent block or cropping block). Figure 3 shows a simple theoretical example of farm-scale stratification based on LMUs. The method by which LMUs are delineated will likely vary in complexity between different farms, but SOCS and stock changes will likely be more similar within, than between LMUs and therefore LMUs should provide a good basis to stratify for SOCS benchmarking and monitoring. If existing LMUs contain land with distinctly different soil types, topography or management practices, then these areas should be further divided for the purposes of SOC benchmarking and monitoring. If soils on the farm are poorly characterised, more detailed soil mapping may be warranted. If the CEA of interest is only a small portion of the farm, then it may be more appropriate to stratify by individual paddocks (finer scale LMUs) or even areas within paddocks.

It is important to note that although LMUs may provide a reasonable basis for stratification, there is no guarantee that stratification using this approach will be suitable. As noted in de Gruijter et al. (2016), the mean level of SOCS and changes in SOCS within each stratification unit might be expected to be related (correlated), and the standard deviation of each stratification unit might be expected to be less than the standard deviation of the pooled SOCS samples from the whole farm. When this is the case, there may an advantage to stratification by LMU, but it might turn out that this is not the case.

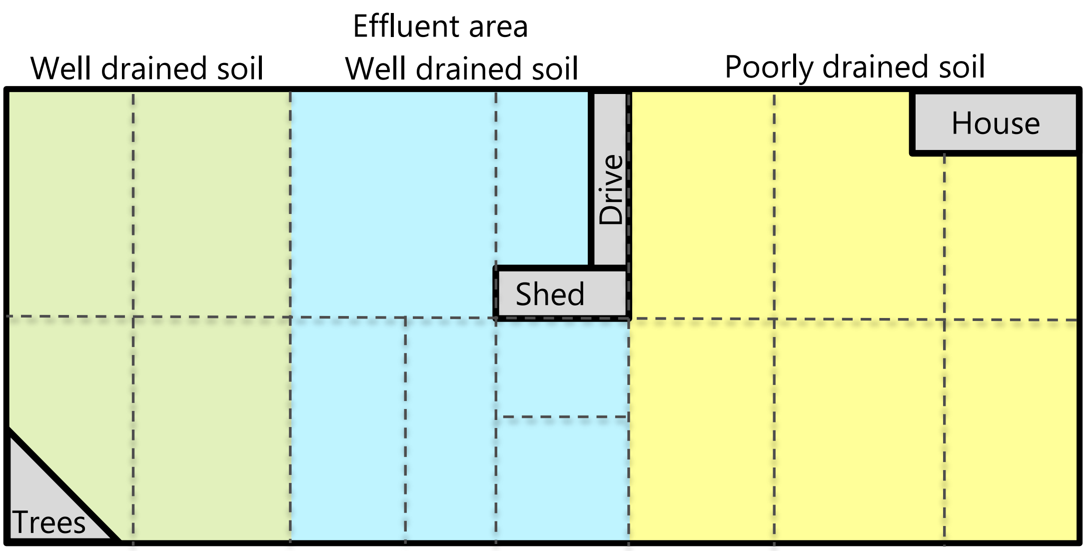

Figure 3: Illustration of a simple farm divided into three strata (different background colours) based on soil properties and drainage. Dotted lines are paddock boundaries. The grey areas represent a house, drive, shed and forested area which are excluded from the carbon estimation area.

Complex stratification based on detailed prior data and multiple spatial layers

Malone et al. (2018) describe a method where New Zealand's national scale SOCS map/model can be downscaled using high resolution local data (soil type map; digital elevation map and derived terrain attributes such as slope, aspect, wetness; apparent electrical conductivity; gamma radiometrics) and used as the basis to reduce the sampling effort for SOCS benchmarking. The downscaled SOCS map combined with local data can then be used to help delineate strata on individual farms where SOCS should be relatively uniform. The method was applied to two farms in the Manawatu region: 1) the Massey University 476-ha Tuapaka sheep and beef farm with 99 ha of flats and 365 ha of hill country, and 2) the Massey University 143-ha No. 1 Dairy Farm, which is situated on alluvial Recent soils bordering the Manawatu River. Figures 4–7 illustrate the stratification process for these two farms, with more specific details provided in Malone et al. (2018).

The reliability of stratification using this approach depends on how accurate the national-scale SOCS map is for the farm in question and the amount and accuracy of higher resolution farm scale data. If there are differences in land management within a proposed CEA, this would need to be considered when deciding how to stratify for SOCS monitoring. For example, stratifying using the covariates noted above (e.g. soils and slope) might span two different management regimes where SOCS might be expected to increase in one but decrease in another. This would likely result in no change in SOCS being detected, even though changes in SOCS did occur.

In general, this more detailed approach to stratification should provide strata that are homogenous in terms of SOCS (and likely stock changes), which should improve the precision with which SOCS and stock changes can be determined. The method can also incorporate a procedure to optimise the number of strata and sampling sites through time. This process of stratification (and other similar approaches) currently requires a high level of expertise. We therefore only recommend using this stratification approach if the combination of topography, soils and management were complicated and the benefits associated with precise SOCS benchmarking and monitoring were expected to be large relative to costs. However, a long-term goal would be to develop a web tool that interrogates a database of nationally available data layers and automates the stratification of any one farm.

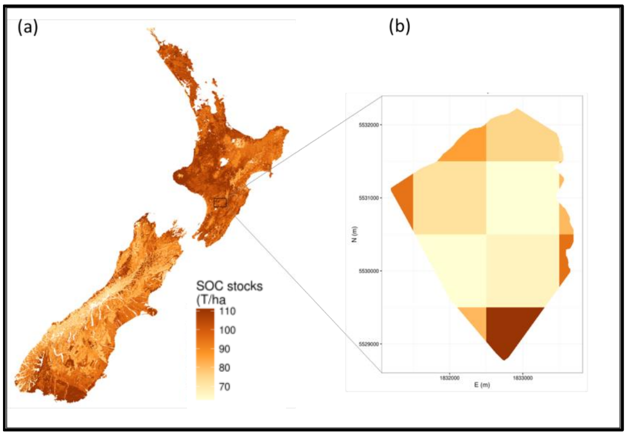

Figure 4: (a) A national SOCS model (Stephen McNeill, pers. comm.; NZAGRC 2016) was used to derive (b) the baseline map and a weighted mean soil carbon stock estimate (93.77 Mg.ha^{-1} to 30 cm; 90% confidence interval 54.4 – 133.15) for Massey University Tuapaka Sheep and Beef farm.

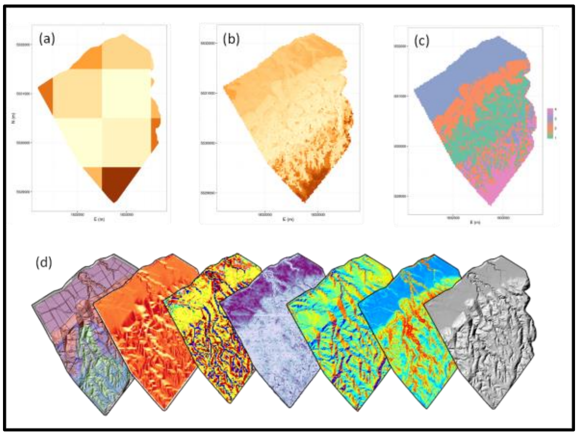

Figure 5: (a) Stratification was achieved by downscaling the (a) baseline SOCS map to produce (b) a disaggregated SOCS map, which was then optimally stratified (c) to guide a soil sampling campaign. Panel (d) shows the high resolution covarying data layers (parent material, global irradiation, landform elements, wetness index, aspect, slope and elevation) used to guide the process of downscaling the national scale SOCS map.

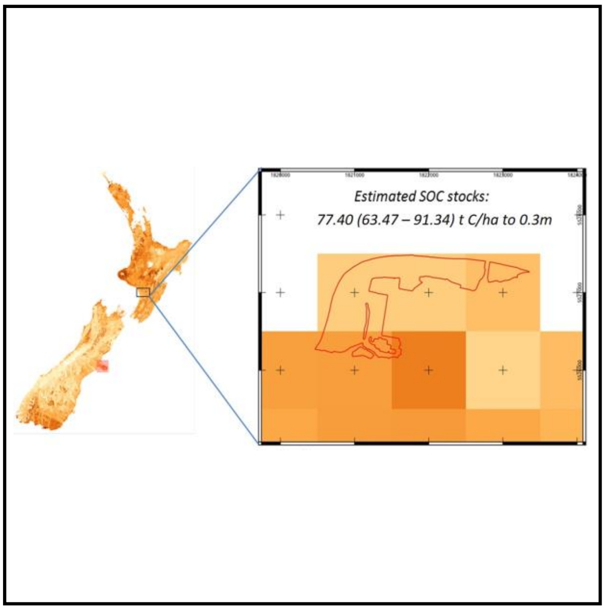

Figure 6: A national SOCS model (NZAGRC, 2016) (left) was used to derive a coarse scale SOCS map of the Massey University No. 1 Dairy farm (right).

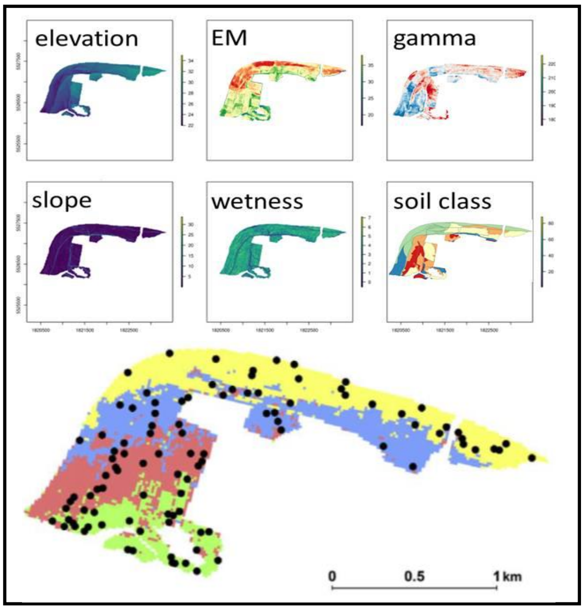

Figure 7: Covariate data layers (top) were used to downscale the national SOCS model and this finer scale map was then stratified to assist a stratified random sampling method for benchmarking SOCS and for future monitoring exercises. The black points are the random sampling locations in each stratum.

5.2.3 Procedure for determining soil sampling sites

Once the CEA (Section 5.2.1) and strata have been defined and delineated (Section 5.2.2), three further sampling design issues must be decided:

1 The procedure for random allocation of soil sampling sites within each stratum (this Section),
2 The number of soil sampling sites required within each stratum (Section 5.2.4), and
3 Whether to conduct analyses on individual or composite soil samples (Section 5.2.5).

The procedure for allocating sampling sites within each stratum can be any method where sites are determined randomly, while also requiring that there is a defined probability for every point within the stratum being selected. The simplest method that meets these requirements is simple random sampling (SRS), and when this is combined with stratification the result is stratified simple random sampling (StSRS). An alternative is to use one of the spatial balanced sampling methods within each stratum, such as balanced acceptance sampling (BAS) (Robertson et al. 2013) (see Section 4.1.2). The balanced sampling methods have greater efficiency when compared with SRS; that is, fewer samples are required for a required accuracy of the CEA SOCS. However, SRS is easier to understand and simpler to implement in software, although reference implementations are well-known for all methods (McDonald 2016).

If implementing StSRS, random soil sampling locations in each stratum need to be determined before sampling using a random number generator that produces random numbers that could include all possible latitude and longitude (or eastings and northings) values within the stratum. The Australian Government (2018) method stipulates that the precision of each soil sampling location is a minimum of five decimal places if latitude and longitude are used and three decimal places if northings and eastings are used. Figure 8 provides an example of potential random sampling site layout on a simple farm divided into three strata.

5.2.4 Number of soil sampling sites required

The number of soil sampling sites required in each stratum and the whole CEA for benchmarking or monitoring is dependent on how precise the estimate of the SOCS mean or change in the SOCS mean needs to be, as well as the spatial variability of SOCS and the variability of the rate of change in SOCS through time (temporal variability). The number of sites is also influenced by the type of monitoring design used. Unfortunately, it is not possible to ‘know’ for sure how many sampling sites are required for SOCS monitoring until samples are collected from two sampling dates. Indeed, the FAO (2019) and Australian Government (2018) methods provide examples of how to calculate minimum detectable differences in SOCS change once data from two or more points in time are in hand. However, it is useful to get an indication of the likely number of sampling sites required to meet the objective(s) of the study before investing in a sampling campaign. This can be achieved using a statistical power analysis, usually with data collected from a smaller pilot study. Here we use a subset of data from an existing SOCS survey from Massey Dairy 1 and Tuapaka (e.g. Figure 7) to show how a pilot study can be used to estimate the number of sampling sites required to detect a change in SOCS of 2 or 5 Mg.ha^{-1} when different types of monitoring design are used.

We recommend using the static synchronous (SS) study design, where the same sites are revisited in each monitoring round, since this requires the smallest number of samples to detect a specified change in SOCS. However, for completeness, below we also describe the independent synchronous monitoring design and Section 4.3.3 provides a more detailed discussion on the pros and cons of different monitoring designs.

Independent synchronous monitoring design

For the independent synchronous (IS) monitoring design (outlined in Section 4.3.3), samples are collected at random locations (sites) across the farm. New independent random sites are sampled in the monitoring rounds at later dates. To illustrate how a pilot study could be implemented to determine the required sample size for this design, SOCS data from three of the random soil sampling sites in each of the four strata as shown in Figure 7 were used to provide a total of 12 pilot sampling sites spread across Massey Dairy 1. Note that 100 samples are available from previous research studies on Massey Dairy 1 but gathering this many samples for a pilot study is not realistic, which is why only 12 were used as an example. Using this pilot set of samples, the estimated mean SOCS was 65.45 Mg.ha^{-1}, and the estimated standard deviation was 12.70 Mg.ha^{-1}. Using an estimated temporal standard deviation of 0.6 Mg.ha^{-1}, based on the assumption that random annual changes in SOCS at any one sampling location will generally[1] be within the range of $\pm$1 $Mg.ha^{-1}$ y^{-1}, the spatial-and-temporal SOCS standard deviation is 12.71 $Mg.ha^{-1}$ (calculated from the pooled variance). This figure was used to estimate the required minimum sample size using a two group, two tail, t-test power analysis with a significance level of 0.05 and a power of 0.8. The result was that 636 sampling sites would be required for both the benchmarking and first monitoring round to detect a SOCS change of 2 $Mg.ha^{-1}$ in the top 30 cm. The same calculations for the more variable hill country Massey Tuapaka farm (with a standard deviation of 22.94 $Mg.ha^{-1}$) showed that 2067 sites would be required for each sampling date (i.e. 4134 in total) to detect a SOCS change of 2 $Mg.ha^{-1}$.

Carrying out farm-scale SOCS monitoring with 636 (Dairy 1) or 2067 sites (Tuapaka) is probably not practically tenable. To reduce the soil sampling effort to a more practical level we have a few choices such as: 1) increasing the minimum size of the SOCS change we want to detect; for example for Massey Dairy 1, detecting a 5-$Mg.ha^{-1}$ change instead of 2 $Mg.ha^{-1}$ would require 103 sites when using the IS design; 2) changing the study design to use static synchronous sampling as described in the next subsection; or 3) using detailed stratification methods. As noted in Section 5.2.2, stratification can provide a way to reduce the sampling effort in cases where SOCS differ markedly between different parts of the farm but the method requires detailed prior information and computation to determine whether stratification will be beneficial. For Massey Dairy 1, stratification as illustrated in Figure 7 did not usefully reduce the number of monitoring sites required, due to the relatively uniform nature of SOCS on the farm. However, an analysis for Massey Tuapaka farm gives strikingly different results, improving the precision of the estimated mean SOCS by 26% for an allocation of 42 samples across the farm (Malone et al. 2018). The reduction in the sampling effort as a result of optimal stratification means that fewer samples would be required to detect a specific change in SOCS, or alternatively, a smaller change in SOCS could be detected for a specified sampling effort.

Static synchronous SOCS monitoring design

Since the required number of sampling sites needed for the IS monitoring design is impractical (i.e. 636 for Massey Dairy 1) to detect a SOCS change of 2 $Mg.ha^{-1}$, we now consider the static synchronous (SS) monitoring design. For the SS monitoring design, samples are still collected at random sites across the farm for the benchmark sampling, but for the subsequent monitoring rounds samples are collected in close proximity (e.g. spaced 5 metres apart) to the sites sampled in the benchmarking phase (Figure 8). This design exploits the fact that SOCS close together are usually more similar than those further apart (i.e. are spatially correlated). A power analysis for the SS study design requires an estimate of the standard deviation of the difference in SOCS between samples collected near each other (e.g. paired sites within 5 metres) as well as the estimated temporal variability of SOCS change.

The original survey of Massey Dairy 1 was designed with the IS monitoring design in mind and did not have soil samples collected in close proximity, with most closely spaced points still over 30 metres apart. There are few data available in New Zealand where unbulked soil samples have been taken with a spacing of less than 10 metres and SOCS determined. Based on the few data that are available, the standard deviation of differences in SOCS from paired sites less than 10 metres apart is in the range 5–10 $Mg.ha^{-1}$. Applying these figures to a paired sample, two tail, t-test power analysis with a significance level of 0.05 and a power of 0.8, the required minimum sample size is shown in Table 1 below.

[1] If the random changes in SOCS are Gaussian distributed and have a 90% confidence interval of [-1,1] $Mg.ha^{-1}$, then the implied standard deviation is approximately 0.608 $Mg.ha^{-1}$. We have used a figure of 0.6 in calculations.

Table 1. Estimated minimum number of samples for a two-sided t-test for paired SOCS samples, for various standard deviation of SOCS differences and two different specified changes in SOCS over time. The power estimate is based on a significance level of 0.05 and a required minimum power of 0.8.

Standard deviation of SOCS differences	Change in SOCS to be detected (significance level 0.05, power 0.8)	
Mg.ha^{-1}	2 Mg.ha^{-1}	5 Mg.ha^{-1}
5	52	11
7	99	18
10	199	34

Table 1 shows that knowing the true value of the standard deviation of the SOCS differences for closely spaced samples is critical for estimation of the number of samples required for monitoring with a specified level of change in SOCS. Doubling the standard deviation of SOCS differences for closely spaced samples results in a requirement to quadruple the number of samples for 2 Mg.ha^{-1} change, or triple the sampling requirements for a 5 Mg.ha^{-1} change. This emphasises the need to adhere to FAO guidelines of 5–15 SOCS locations for a pilot survey (FAO 2019), and for careful field practise.

Static synchronous SOCS monitoring with sample bulking

Based on the calculations above (Table 1), SS sampling is a more efficient approach for SOCS monitoring for Massey Dairy 1 than IS. The sample analysis requirements (and thus costs) could be reduced further by bulking (or compositing) samples. If taking this approach, the actual soil sampling for the benchmarking and monitoring rounds remains almost identical to when analysing individual samples, with the only differences being that random groups of individual samples from within each stratum are bulked together and thoroughly homogenised before analysis (e.g. see Figure 8), and the number of individual sampling locations required changes slightly. The number of sampling sites required increases when samples are composited because the power analysis is conducted using the standard deviation of the differences in SOCS of bulked samples, of which there are logically fewer than for all the individual samples. The standard deviation of the difference of paired SOCS for any bulking factor can be estimated as $\mathrm{Stdev\_ind}/\sqrt{N_B}$, where $\mathrm{Stdev\_ind}$ is the standard deviation of the differences between individual paired samples and N_B is the bulking factor (the number of soil samples bulked together). FAO (2019) recommend that at least five cores should be collected to form a composite sample within a stratum.

Assuming a nominal figure of 5 Mg.ha^{-1} for the standard deviation of the differences of paired SOCS, applying a bulking factor of 5 (i.e. 5 samples bulked together as described above) to the data from Massey Dairy 1, reduces the standard deviation of SOCS difference to $5/\sqrt{5} = 2.24$ Mg.ha^{-1} for the spatial standard deviation of SOCS differences between bulked paired samples, assuming the bulked samples are independent. The pooled spatial-and-temporal SOCS standard deviation is now 2.32 Mg.ha^{-1}, and the required minimum number of bulked samples is 13 to detect a change of 2 Mg.ha^{-1}. In this case, a total of 65 (13 bulked samples consisting of 5 cores each), or 13 more than if all individual soil samples were analysed. Further discussion on the advantages and disadvantages of bulking samples is provided in Section 5.2.5.

Summary of how to conduct a pilot study

The variability of SOCS on a particular farm is generally not known beforehand, so a pilot study needs to be undertaken to establish these values from which the number of sampling sites required to detect a given change in SOCS can be estimated using a power analysis. The recommended way to establish a pilot study when using the static synchronous monitoring approach is outlined in the steps below and illustrated in Figure 8. Steps for the independent synchronous sampling design are the same, except paired soil samples at each location are not required and the power analysis utilises the standard deviations across the sample area.

1 Using the random approach outlined in Section 5.2.3, select a minimum of five soil sampling locations within each stratum defined within the CEA (Section 5.2.2). FAO (2019) recommends 5–10 samples per stratum for a pilot study.

2 At each of the soil sampling locations, generate a new random sampling location spaced 5 metres[2] from the initial location using a random bearing from 0 to 360 degrees (Figure 8).
3 Collect a soil sample at each of the locations determined in the previous step observing the protocols set out in Section 0 and process samples to determine SOCS for each sample as per the protocols in Section 5.4.
4 Calculate the standard deviation of the difference in SOCS measured in the two soil samples at each location.
5 Estimate the standard deviation for the temporal change in SOCS (generally not known, but a figure of 0.6 $Mg.ha^{-1}$ is suggested).
6 Calculate the pooled standard deviation for SOCS differences incorporating both spatial and temporal components, using:
$\sqrt{\text{Standard deviation of SOCS differences}^2 + \text{Standard deviation of temporal SOCS change}^2}$.
7 The pooled standard deviation is combined with the specified change in SOCS to be detected (e.g. 2 $Mg.ha^{-1}$) a significance level, and required minimum power, and used in a paired t-test to estimate the required minimum sample size that will be required to detect the change in SOCS, given the specified parameters. Appendix 1 gives the mathematical development for this method, including when bulking samples.

The method described above is particularly suitable for estimating the sampling effort required for monitoring SOCS change. The method can also be used to estimate the sampling effort required to estimate the mean SOCS for a farm (i.e. benchmarking), although this method of pilot study requires twice as many samples as strictly required for benchmarking. The simplest pilot study to estimate the sample size for benchmarking retains steps 1–3 above, but only one SOCS sample at each location. This simpler procedure yields an overall mean SOCS over the farm, as well as an estimate of the standard deviation of the independent samples.

The mean SOCS for the farm has a precision that depends on the number of samples in the pilot study; more samples provides a greater degree of precision for the mean SOCS estimate. If the benchmarking study requires an estimate of the mean SOCS over the farm with a precision that is greater than that provided by the small pilot study (this will generally be true) then it is possible to estimate the number of additional samples that would be required to achieve this precision. Appendix 1 provides the mathematical development for this method.

It is important to note that the nature of the pilot study is specific to the requirements for the farm, such as whether the aim is benchmarking (i.e. estimating the mean SOCS to some required precision) or monitoring (i.e. measuring the change in SOCS over time). Acquiring two samples for each location in the pilot study gives the greatest flexibility, since the required sample size for the mean SOCS as well as the change in SOCS (using the SS approach) can both be determined from the pilot study. The compromise in the latter case is that twice as many samples are needed in the pilot study.

[2] A figure of five metres has been chosen at the time of writing to represent the likely repeatable distance that a location can be determined using consumer-grade satellite location systems (e.g. GPS). Over time, this distance is likely to reduce as satellite navigation improves location accuracy and consumer-grade navigation also improves.

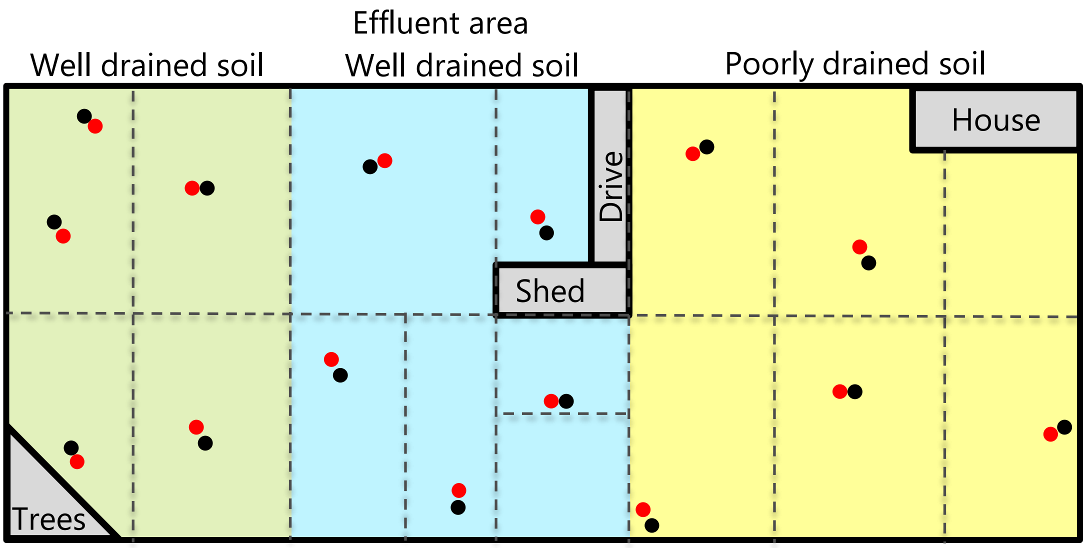

Figure 8. Example layout of a pilot study for a farm with five sample pairs in a stratum. The second element of each pair (red dot) is located a distance of 5 m from the first, in a random quarter-circle direction from North. Note, a similar paired sampling design would be implemented for the full study, with the black dot representing the benchmark sampling and the red dot the first monitoring sampling. The full study would have more samples (e.g. Figure 9).

Cautions

The above example for Massey Dairy 1 shows the kind of calculations required to estimate the required minimum sample size. The calculations themselves, however, rely on several key assumptions:

1 The sample locations are randomly placed over the different strata within the CEA
2 The pilot study samples are spaced far enough apart that they can be considered independent samples of SOCS
3 The distribution of the difference in SOCS is Normal (Gaussian), with no outliers.

Violations of any of the above assumptions may give biased estimates in the sample size calculations. For example, if the sample locations in the pilot study are not random, the estimated mean SOCS might be biased. If outliers are present in the differences of the SOCS then the standard deviation of the difference might be incorrect.

Finally, since the calculations depend on estimating a standard deviation of the SOCS differences, it is important to remember that many more sample points are required to estimate a standard deviation of a specific property (e.g. SOCS or SOCS differences) to a specified relative precision than are required to estimate the mean of the quantity to the same relative precision.[3] Although a minimum of two samples is required to estimate a standard deviation, the resultant estimation has very poor precision, and many more samples are typically recommended. For this reason, FAO (2019) recommends 5–10 samples per stratum for a pilot study, which is strongly recommended.[4]

---

[3] For example, assume a “true” SOCS mean and standard deviation of 65 and 7 $Mg.ha^{-1}$ respectively. A simulation shows that 15 samples has an estimated 95% confidence interval that is 11% of the mean, but 78% of the standard deviation. Increasing the number of samples to 100 improves the precision of the mean estimate to 4.2%, but the precision of the standard deviation estimate for the same number of samples is 29%.

[4] For two samples x_1 and x_2, the sample standard deviation is $|x_1 - x_2|/\sqrt{2}$, which is as informative as giving the sample range of the two sample values. There is a great deal of variability to this estimate, and adding additional value (e.g. three or more) progressively improves the precision of the sample standard deviation estimate.

5.2.5 Analysing individual vs bulked soil samples

The mean SOCS and stock changes for a CEA or individual strata can be estimated in two ways: 1) analysis of single samples collected from across a stratum and 2) analysis of composite or bulked samples. If resources (i.e. time and money) are not limited, we recommend analysing individual soil samples because this provides the richest and most flexible data. For example, provided sufficient samples are taken in the benchmarking round, analysing samples from each location separately means the resulting data can be used to refine stratification and sample numbers for subsequent sampling rounds, and also provides useful information on the spatial distribution of SOCS. Refinement in this way will likely improve the precision with which SOCS changes can be detected and/or reduce the cost associated with sampling because fewer sampling sites may be required (de Gruijter et al. 2016). Having individual sampling points which have a known SOCS and then change in SOCS after subsequent sampling rounds, will also improve the ability to identify what factors are influencing changes in SOCS (e.g. slope, management).

If soil samples are composited within or across strata, then stratum boundaries cannot be changed in future sampling and data analysis, unless the user is prepared to restart the monitoring process from the beginning. In this case, the initial sampling would act as a pilot study to provide information to better constrain sampling design. However, analysing composite samples can substantially reduce the number of samples that need to be processed and analysed, thus reducing costs. Compositing samples reduces the spatial variability of SOCS in the estimate of the mean, but multiple composite samples are still required to provide an indication of the variance around the mean. As a general rule, FAO (2019) suggest that the mean SOCS of composite samples taken from the same stratum should not differ by more than the SOCS change that is aiming to be detected (e.g. 2–5 $Mg.ha^{-1}$). FAO (2019) recommend that at least five cores should be collected to form a composite sample within a stratum.

Figure 9 illustrates a potential random sampling site layout on a simple farm and explains the difference between individual vs composite sampling. Section 5.6 provides cost estimates for the two different options based on the Massey Dairy 1 case study farm and Section 5.4.5 provides the relevant equations for estimation of spatial means using both individual and composite samples.

Figure 10: Example of a simpler sampling layout than in Figure 9, where multiple cores are taken from small 20 m by 20 m plots and composited for analysis. If very rigorous benchmarking and monitoring of SOCS at the farm-scale is the main objective, the approach outlined in Figure 9 would be preferable. Figure 10 provides an example of a simpler small plot-based sampling layout than that illustrated in Figure 9**Error! Reference source not found.** This approach is easier to implement and will provide an accurate measure of SOCS and potentially changes SOCS for the individual plots. However, even if multiple plots were randomly located within each CEA, the resulting data would not be as spatially representative as the method shown in Figure 9. Therefore, if rigorous benchmarking and monitoring of SOCS at the farm-scale is the main objective, the approach in Figure 9 would be preferable. If a less spatially representative estimate of farm-scale SOCS and changes in SOCS was all that was required, then the simpler small plot approach may be adequate. Resulting data could then potentially be easily incorporated into the national SOCS monitoring system which employs this plot-based sampling approach across the whole country.

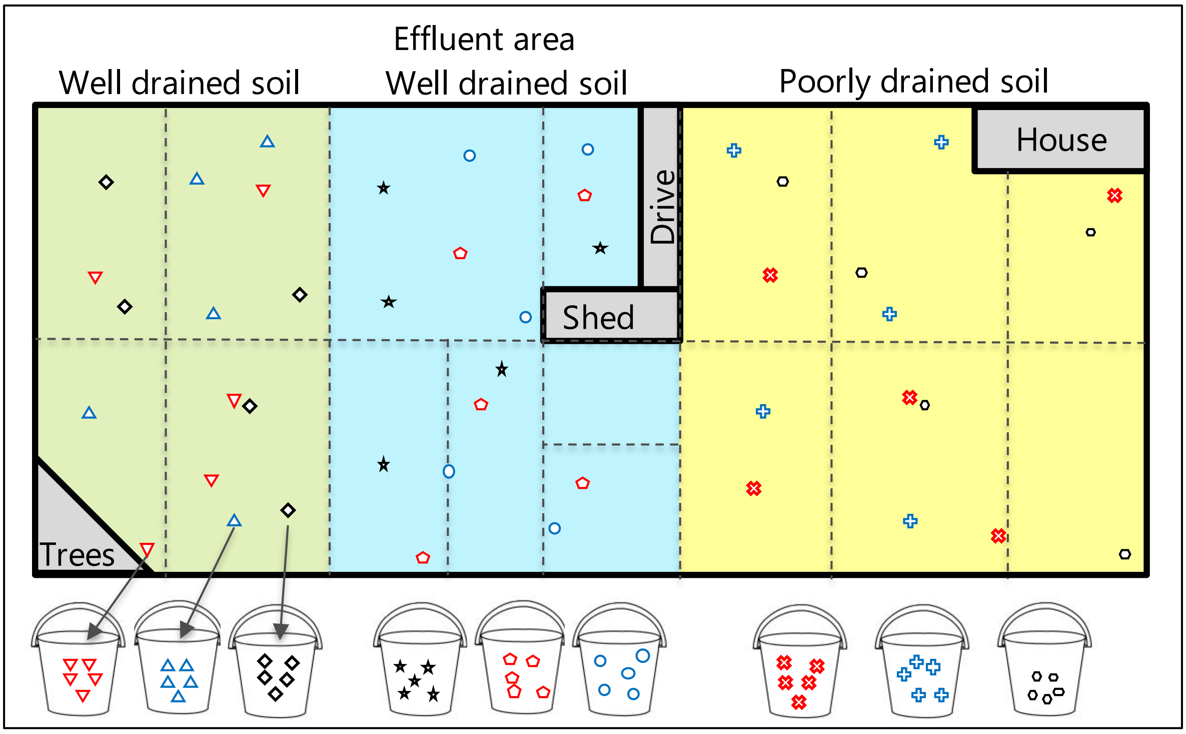

Figure 9: Illustration of a simple farm divided into three strata (different background colours) showing the potential layout of 15 random soil sampling sites in each stratum. Dotted lines are paddock boundaries. If samples were composited, the five samples with the same symbol in each stratum would be bulked together to provide three composite samples per depth for each stratum and nine in total for each depth. If each soil sample was analysed individually it would be a total of 45 samples. The grey areas represent a house, drive,shed and forested area which are excluded from the carbon estimation area.

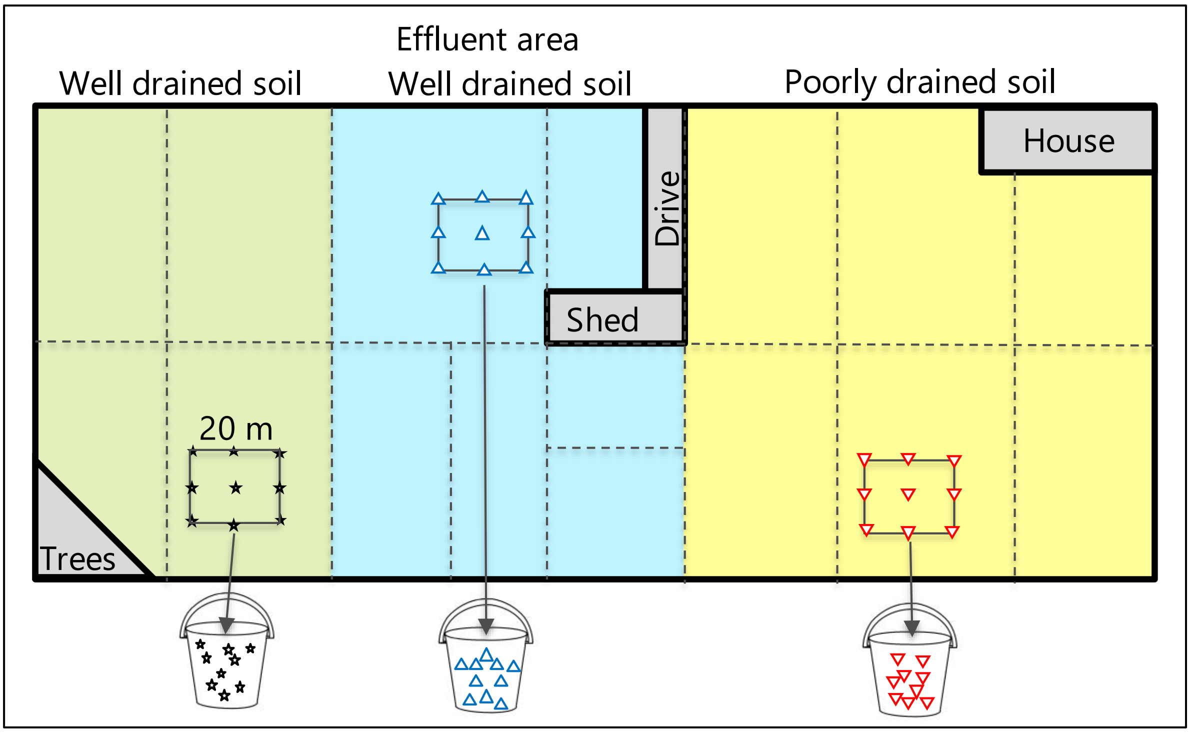

Figure 10: Example of a simpler sampling layout than in Figure 9, where multiple cores are taken from small 20 m by 20 m plots and composited for analysis. If very rigorous benchmarking and monitoring of SOCS at the farm-scale is the main objective, the approach outlined in Figure 9 would be preferable.

5.2.6 Soil sampling depth and number of depth increments to analyse

Following the FAO (2019) and Australian Government (2018) guidelines, soil sampling for SOCS should be to a minimum depth of 30 cm. If sampling is greater than 30 cm deep, SOCS for the 0–30 cm depth must be reported separately. This is consistent with the depth the IPCC uses for national reporting of SOCS and also reduces systemic issues of shallow sampling depths. Soil material shallower than 30 cm is more susceptible to short-term changes associated with land management such as cultivation that can alter the distribution of C within the topsoil and can cause changes in bulk density.

Ideally, we would encourage SOCS to be determined to 100 cm depth where possible. However, from practical experience and routine sampling throughout New Zealand, a depth of 60 cm is often a good compromise. Sampling deeper than 60 cm increases the likelihood of various issues associated with all techniques (e.g. hitting bedrock, difficulty inserting and extracting cores and inability to reach beyond 60 cm to remove soils and beads when sampling stony soils with pits). A depth of 60 cm is twice the recommended standard IPCC depth and is well into the subsoil of most New Zealand soils. Sampling to this depth will therefore capture any SOCS change in both the surface soil and a portion, if not all, of the subsoil.

The decision on the number of soil depth increments to analyse is largely up to the implementer of the SOCS benchmarking and monitoring system. For context, soil sampling for the purposes of New Zealand's national SOCS inventory reporting has been to 30 cm in three 10-cm increments. Analysing multiple depth increments provides important information about where changes in SOCS may be occurring but increases costs and is not required if changes in **total** SOCS for a nominated depth or soil mass is the key question.

5.3 Soil sampling methods

Specific details of soil sampling for SOCS have been extensively documented elsewhere (e.g. Davis et al. 2004; Whitehead et al. 2010) and therefore here we provide only a brief summary of the options available. Plans are currently in place to update documentation of *step-by-step* instructions for soil sampling, processing and analysis (Hedley et al. 2020 updated from Whitehead et al. 2010 and Davis et al. 2004). In future there may be merit in incorporating the details of study design and soil processing/analysis into the same document to provide a comprehensive summary of requirements in one place.

We recommend one of three different soil sampling methods to determine SOCS, with the method chosen determined by on-site soil conditions (e.g. stony soils, sandy soils). The method used for the benchmarking (baseline) sampling should also be used for all subsequent monitoring samplings.

5.3.1 Deep cores

In non-stony soils, the most preferable sampling method from a cost and efficiency perspective is extracting continuous cores (minimum of 30 cm) and dividing these into discrete depth increments if required (Figure 11). Following established Australian Government (2018) methods we recommend that the cutting diameter of the core be **>3.8 cm**. Smaller diameter cores are less suitable for determining bulk density and more prone to compaction. It is important that the core is inserted to the correct depth and that on extraction the core breaks right at the bottom to ensure that a known volume of soil is sampled. Cores need to be checked for compaction upon extraction by measuring the length of core relative to the depth of the hole from which the core was removed. A discrepancy between the two would indicate compaction of the core or a loss of soil from the corer during the extraction process. Avoid transfer of soil material between depth increments. Any transfer of soil between depths will influence the SOCS measurement. Avoid sampling when soils are too dry or too wet, as these conditions increase the risk of soil transfer and loss (too dry) and compaction (too wet). It is also important that no soil material is lost from the sample right through to the determination of total sample and gravel weights in the laboratory (Figure 14).

Figure 11: Nine 4.2 cm diameter cores taken to 60 cm and placed on a cutting board ready to be cut and combined into a composite sample for each 10 cm depth increment.

5.3.2 Pits and short cores

Unfortunately, the deep coring technique is not suitable for stony soils (because the cores cannot be inserted) and some other soils because the samples fall apart/crumble and can jam in the corer during removal, which means a complete core is hard to extract. Sands and some Allophanic soils can be particularly problematic. In this situation it is preferable to use the second method, whereby a soil pit is dug and shorter cores (e.g. 10 cm deep) are used to extract samples in continuous increments through the profile to the desired depth (Figure 12). This method should be able to be used in all soils except stony soils.

Figure 12: Illustration of a 10 cm diameter by 7.5 cm deep core being taken from a pit. The soil will be cut off flush with the bottom of the core and then another core taken (photo by Veronica Penny).

5.3.3 Quantitative pits for stony soils

The third method is required in all stony soils. For this method, a pit of known surface area is excavated with all the soil material removed and screened (~10 mm sieve), with the >10 mm and <10 mm fractions all weighed in the field. A homogenous and representative subsample of the <10 mm material is taken back to the laboratory for analysis. The volume of the pit (or different depth increments within the pit) is determined by lining the pit with a thin plastic bag and back-filling with a known quantity of water, sand or plastic beads. The volume is then calculated using the volume-to-mass ratio of the water, sand or beads. When lining the pit with the plastic bag, ensure any voids are filled as much as possible (press bag into these voids and fill with beads) to capture any volume that was formed from removing rocks at the pit edge. For further information about the specific steps involved in sampling stony soils to determine bulk density, stone content and SOC content, see Hedley et al. (2012), or Davis et al. (2004).

Figure 13: Photos showing how stony soils are sampled using the quantitative pit and bead method (photos from Thomas Caspari and Veronica Penny). See text for description.

5.3.4 Information to record at each soil sampling location

In addition to the soil sampling, we recommend that at a minimum the additional information below be collected/documented:

- Date of sampling and description of soil moisture conditions (e.g. wet or dry).
- Person(s) sampling
- GPS location of each sampling site (accurate +/- 4m).
- Ideally a photo in **three** different directions from the sampling location with notable permanent features of the landscape lined up with the sampling site. This will help relocation of the site if the GPS fails. Alternatively measure distances to key features with a tape measure.
- Photo of each soil core or soil profile if a pit is dug.
- Soil profile description/classification – depending on the level of detail required this may require the services of a professional pedologist.
- Slope measured in degrees.
- Aspect.
- Land use / vegetation at the time of sampling and any known recent changes* such as re-grassing or a fodder crop.

*Note: the longer-term management history of the CEA and individual strata needs to be documented as accurately as possible in the study design phase (i.e. this could impact on stratification). Any changes in land use and management between sampling rounds also needs to be documented.

5.4 Soil processing and analysis

5.4.1 Drying, sieving, and weighing

Once soil samples have been returned to the laboratory, we recommend following the protocols established in the Australian Government (2018) method which is outlined in Figure 14. Soil samples need to be air-dried at ~40°C until a constant weight is achieved and then the total sample mass determined. The sample is then passed through a 2-mm sieve to separate the gravel/stones from the fine soil. The mass of the gravel/stones must be determined. A representative and homogenous sub-sample of the fine fraction is removed for chemical (carbon) analysis and determination of the air-dry moisture content. It is **critical** to stipulate that the **total mass** of the sample is determined in step 1 of Figure 14 and the **total mass** of gravel/stones is also determined in step 3a. We recommend that the soil processing steps (Figure 14) be undertaken at the same laboratory as SOC analysis (Section 5.4.2).

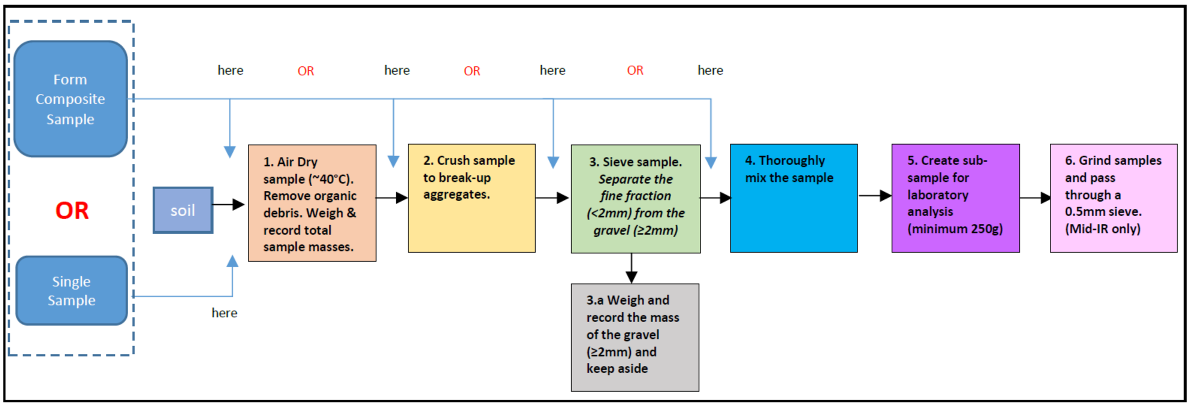

Figure 14: Flow chart outlining the required soil sample processing steps prior to analysis for soil organic carbon concentration and calculation of soil organic carbon stocks (from Australian Government, 2018).

5.4.2 Quantifying SOC concentration of a sample

The SOC concentration needs to be determined by a calibrated and appropriate analytical technique (e.g. dumas combustion or using infrared spectroscopy). We recommend following the protocols adopted for the Australian 2018 methodology as outlined in Figure 15. One specific recommendation we draw attention to is:

> *It is a requirement that analysis of organic carbon content is undertaken by a laboratory that is certified for organic carbon analysis by the Australasian Soil and Plant Analysis Council (ASPAC).*

For New Zealand applications this guidance should also be followed, but modified, such that the organic carbon analysis be undertaken by a laboratory with accreditation from International Accreditation New Zealand (IANZ). Following this recommendation will ensure the concentration of SOC in a given soil sample will be accurate and results between benchmarking and subsequent monitoring rounds consistent and comparable. The same laboratory and analytical method should be used for all subsequent sampling rounds and if a change in method or laboratory is essential, then very strict quality controls must be in place to ensure no systematic bias which could result in an apparent (but not real) change in SOCS. An air-dry archive of each sample will enable checks to be performed or will allow historic samples to be re-run if the analytical technique was to change in the future.

The laboratory must have suitable protocols for dealing with inorganic carbon in samples if present. However, in New Zealand few soils outside of Central Otago contain appreciable amounts of inorganic carbon.

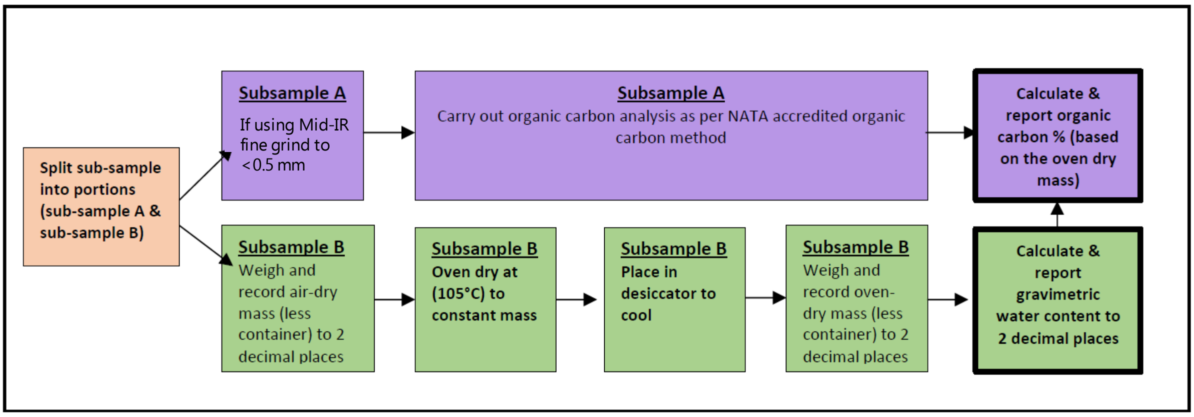

Figure 15: Flow chart outlining the steps required to determine the oven dry soil organic carbon concentration of the fine soil (<2 mm) following sieving as outlined in Figure 14 (from Australian Government, 2018).

5.4.3 Calculating soil mass, stone content and bulk density

The commonly accepted definition for SOCS is the mass of carbon in the portion of soil material that will pass through a 2-mm mesh (fine soil) expressed per unit area (e.g. $Mg.ha^{-1}$) for a specified soil depth or mass (on an oven dry basis). The calculation of SOCS therefore relies on accurate measurement of the mass of fine soil per total sample volume (Poeplau et al. 2017). The simplest way of achieving this is by extracting a soil core (as outlined previously), where the volume will be pre-determined by the core diameter and depth of soil sampled. For pit sampling in stony soils, the volume of the total sample is calculated from the volume of water, sand or plastic beads used to fill the pit (Section 5.3.3).

The air-dry mass of the total sampled soil and gravel is determined in step 1 of Figure 14, and the mass of gravel in step 3a. The fine soil (<2 mm) mass is then determined by subtracting the mass of air-dry gravel/stones from the total air-dry sample weight (this is more accurate than directly measuring the <2 mm material because dust is lost during the sieving process).

The oven dry mass (g) of fine soil ($Mass_{finesoil_{oven\ dry}}$) is calculated by using the air-dry mass (g) of soil ($Mass_{finesoil_{air\ dry}}$) and the moisture content of the air-dry soil (often expressed as a moisture factor, MF) as follows:

$$MF = 1 + \left(\frac{Mass_{finesoil_{air\ dry}} - Mass_{finesoil_{oven\ dry}}}{Mass_{finesoil_{oven\ dry}}} \right) \qquad \text{Equation 7}$$

$$Mass_{finesoil_{oven\ dry}} = \frac{Mass\ finesoil_{air\ dry}}{MF} \qquad \text{Equation 8}$$

If there is a desire to determine whole soil bulk density (i.e. for gravel plus fine soil), then gravel from step 3a in Figure 14 will need to be oven dried and weighed. If a fine soil (<2 mm) bulk density is required, the volume of the gravel (calculated using their respective particle density) will need to be subtracted from the total sample volume.

5.4.4 Calculating SOCS of a sample

Here we provide equations for the simplest method for calculating SOCS for a sample following Poeplau et al. (2017). The SOCS of a sample (*i*) is calculated by multiplying the concentration of SOC of the fine soil by the fine soil stock (FSS_i) (<2 mm) in a given soil layer, reported on a per hectare basis (Mg fine soil ha^{-1}), as follows:

$$FSS_i\left(Mg_{finesoil}\ ha^{-1}\right) = \frac{mass_{finesoil_{oven\ dry}}(g)}{volume_{sample}\ (cm^3)} \times depth_i(cm) \times 100 \qquad \text{Equation 9}$$

where $Mass_{finesoil_{oven\ dry}}(g)$ is from equation 8, $volume_{sample}$ is the volume of the sample (cm^3) and $depth_i$ is the depth of the sample (cm) and 100 converts values to Mg.ha^{-1}

The SOCS $(Mg\ C\ ha^{-1})$ is then calculated as:

$$SOC_{stock_i} = \frac{SOC_{conc\ soil}(\%)}{100} \times FSS_i\left(Mg_{finesoil}ha^{-1}\right) \qquad \text{Equation 10}$$

Where $SOC_{conc\ soil}$ is expressed as a percentage (%) as per the result from the laboratory on an oven dry basis.

It is important to note that when sampling to a fixed depth there is a risk that different masses of soil may be sampled between sampling rounds (e.g. due to compaction). A site that experiences compaction between the first and second samplings would have a greater mass of soil collected and therefore comparison of the SOCS between the two would not be appropriate. Therefore, it is critical that SOCS from all sampling times be expressed on an equivalent soil mass (ESM) basis to provide meaningful comparisons. The preferred way to calculate SOCS on an ESM basis is to utilise cubic spline functions applied to samples collected from at least three contiguous depth layers (e.g. 0–10 cm, 10–20 cm and 20–30 cm). Detailed descriptions and excel based and R based templates for these calculations are provided by Wendt and Hauser (2013) and von Haden et al. (2020) respectively. When the study design involves returning to the same sampling sites for each sampling round, the ESM reference for each sampling site is defined by the sampled soil mass in the initial baseline sampling (e.g. the mass of soil to a given depth for an individual core sample). When the study design involves returning to new independent sampling points in the stratum, or compositing soil samples collected across a stratum, then the ESM reference should be set for the whole stratum.

Here we briefly outline the Australian Government (2018) method for defining the ESM in a fixed, single depth sample (although we calculate the ESM of the fine soil stock (FSS_i) from Equation 9, rather than the whole soil). The ESM should be set after the initial benchmarking sampling by collating the FSS_i in all samples in a stratum (as defined in Figure 9), with the ESM defined as the mass of the 10th percentile $(ESM_{\tau=0.1})$. The 10th percentile is used rather than the mean, so that in 90% of cases there is no requirement for extrapolation of SOCS to unmeasured soil material. Extrapolation would occur in 50% of cases if the mean was used. Using the 10th percentile, rather than the minimum value also avoids the risk of one low value influencing all calculations for the duration of the project. Once the ESM is defined, the SOC_{stock_i} for all sampling rounds is adjusted and expressed on an ESM basis as per the equation below:

$$SOC_{Stock_{ESMi}}\ (Mg\ C\ ha^{-1}) = SOC_{stock_i} \times \frac{ESM_{\tau=0.1}}{FSS_i} \qquad \text{Equation 11}$$

The same method used for ESM calculations must be applied for all sampling rounds.

5.4.5 Calculating SOCS for carbon estimation areas (e.g. a farm)

Section 5.4.4 describes how the SOCS is estimated for a sample within each stratum. If the SOCS mean is $\hat{\bar{Z}}_h$ and the variance is $\hat{V}\left(\hat{\bar{Z}}_h\right)$ over stratum h, where h is from 1.. (H is the number of strata within the CEA), then the overall mean SOCS over the CEA, $\hat{\bar{Z}}_{St}$, with total area A_{CEA} is the area-weighted sum of the means across the CEA (following de Gruijter et al. 2006):

$$\hat{\bar{Z}}_{St} = \sum_{h=1}^{H} \frac{A_h}{A_{\mathrm{CEA}}} \cdot \hat{\bar{Z}}_h \qquad \text{Equation 12}$$

Similarly, the variance of the mean in stratum h is estimated by area weighting using:

$$\hat{V}\left(\hat{\hat{Z}}_{St}\right) = \sum_{h=1}^{H} \frac{A_h}{A_{\mathrm{CEA}}} \cdot \hat{V}\left(\hat{\hat{Z}}_h\right) \qquad \text{Equation 13}$$

In equation 12, $\hat{V}\left(\hat{\hat{Z}}_h\right)$ is the variance of the within-stratum mean, which is estimated as:

$$\hat{V}\left(\hat{\hat{Z}}_h\right) = \frac{1}{n_h(n_h-1)} \sum_{i=1}^{n_h} (z_{hi} - \hat{\hat{Z}}_h)^2 \qquad \text{Equation 14}$$

Where n_h is the number of observations in stratum *I,* and z_{hi} is the observed value within stratum *i.* The 100(1 – α) % confidence interval for$\hat{\hat{Z}}_{St}$ is given by:

$$\hat{\hat{Z}}_{St} \pm t_{1-\alpha/2} \cdot \sqrt{\hat{V}\left(\hat{\hat{Z}}_{St}\right)} \qquad \text{Equation 15}$$

where $t_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the Student distribution with $n-1$degrees of freedom where n is the number of collected samples from the site. In general, when measuring SOCS for a CEA the spatial mean and variance are reported.

5.5 Expertise required for SOCS benchmarking and monitoring

Designing and implementing a SOCS benchmarking and monitoring system for a specific farm will require a range of skills and expertise as outlined in the following sections.

5.5.1 Study design

One skill required at the beginning of the study design process is the ability to engage with farmers and define meaningful objectives that meet their objectives. In most cases, the study design step would require a pilot study for the farm and pedological expertise would be invaluable in assessing the soil pattern. Appropriate study design will also require input from someone with a good understanding of farm systems and soil carbon to inform the stratification process and/or check whether an existing FEP is suitable to use for stratification. The expertise required for stratification will differ depending on the method chosen to define and delineate strata. If the simpler option of using an existing FEP is chosen, only limited Geographic Information (GIS) skills will be required to determine random sampling locations. If the more complex approach is taken to stratification, then more specific spatial modelling skills will be required. Statistical skills will also be required to calculate the likely number of sampling sites required to meet the specified objective (see Section 5.2.4).

5.5.2 Soil sampling

A key attribute required when sampling for SOCS is the ability to pay **attention to detail** and carefully follow protocols. Soil sampling for SOCS needs to be conducted in a more careful manner than for standard soil fertility testing because both accurate measurement of the concentration of carbon and total mass of soil in a defined sample volume is required. Here we are attempting to demonstrate changes as small as 5% of the stocks and so 5% errors with sampling will mask true changes or indicate untrue increases or decreases. Sampling will also be much deeper (≥30 cm) than for typical soil fertility testing (usually 7.5 or 15 cm).

General field team requirements would include:

- Good off-road driving expertise and understanding of on-farm health and safety and biosecurity issues.
- A fitness level for physically demanding soil sampling (coring or pit excavation).
- Ability to use handheld global positioning systems (GPS) to find predetermined sampling locations and accurately record and then download coordinates of sampled positions.

- Ability to identify when predetermined random sampling locations are not suitable (e.g. a sampling site falls in a small area of forest that should have been excluded from the CEA).

More specific expertise required for soil sampling is:

- Ability to decide which soil sampling method is appropriate for each specific location (e.g. pit excavation vs coring).
- Ability to extract intact soil cores with no compaction (or note compaction) and make sure no sample material is lost.
- Ability to excavate pits for stony soils and to determine the volume of the hole and the stone content.
- Ability to clearly label samples.
- Basic understanding of how to describe a sampling site (e.g. slope, aspect, vegetation) and soil profile (at least soil order level – although this may not be essential at the farm scale depending on objectives).

We strongly recommend that if SOCS results are intended to be used for reasons other than general interest, then personnel involved in sampling should be required to do some form of standardised training in the techniques. Any training needs to involve a field component.

5.5.3 Processing and analysis of samples

Although the *soil processing* steps outlined in Section 5.4.1 could be carried out by most people with appropriate equipment (e.g. 2-mm sieve and clean drying systems (40°C)), we strongly recommend that soil samples collected in the field be sent for processing and SOC analysis (Section 5.4.2) in the same laboratory, which has accreditation from International Accreditation New Zealand (IANZ). This is consistent with the Australian Government (2018) method and will ensure good quality control over results and provide confidence that results between benchmarking and subsequent monitoring rounds are comparable.

5.5.4 Data analysis and interpretation

Calculating SOCS for individual samples is relatively simple, but it is more complicated to implement equations to calculate stocks, stock changes and their variances for whole CEAs, particularly one with strata of varying sizes. We recommend that someone with strong statistical skills is involved at some point in the process. This could be to initially set up the analysis system (e.g. an excel workbook), or at least to check the procedures used. Calculations need to be transparent and auditable.

The data analysis and interpretation process could potentially be simplified and streamlined with the development (e.g. for NZ) of standardized easy to use spreadsheets which only require input of key measured parameters (e.g. SOC concentration, fine soil mass) and will then calculate SOCS and stock changes for individual samples, strata and CEAs. This could potentially be associated with a centralised database.

5.5.5 Summary of expertise requirements

In general, many rural professionals (e.g. farm consultants, fertiliser representatives) with some specific training related to processes influencing SOC stocks and stock changes, soil sampling and processing, and subsequent data analysis and interpretation could design and implement a SOCS benchmarking and monitoring programme for individual farms. Additional specialist statistical skills may be required for estimation of the number of samples, and spatial modelling skills would be needed if the detailed stratification option were chosen. Scientists and associated technicians from a few Crown Research Institutes (CRIs) or universities would also have the expertise required to implement the necessary protocols for SOCS benchmarking and monitoring.

Development of standardized easy to use spreadsheets could simplify and streamline the data analysis and interpretation process and would aid with potential aggregation of SOCS data into industry and national databases for wider analyses. Updated documentation (e.g. updated from Whitehead et al. 2010 and Davis et al. 2004) of more detailed *step-by-step* instructions for soil sampling, processing and analysis may also improve consistency of the methods used.

5.6 Estimated costs

The costs to design and implement a SOCS benchmarking and monitoring system will vary considerably between farms, depending on a range of factors such as: the objective (e.g. the precision with which SOCS and stock changes need to be determined) farm size, topography, spatial variability, and whether individual or composite samples are analysed. The costs will also depend on who is designing and implementing the system and the associated charge out rates.

As an example, Table 2 provides a summary of estimated costs for benchmarking and subsequent monitoring of SOCS for the 143 ha Massey University No.1 Dairy Farm (Dairy 1) using the Static Synchronous design. Costs are based on typical Crown Research Institute scientist and technician charge-out rates, and sample preparation and analysis costs based on Manaaki Whenua's IANZ accredited Environmental Chemistry Laboratory. **It is important to note that the costs in Table 2 are only an example based on the specific assumptions/decisions outlined in this section. If different assumptions were made, costs could increase or decrease.**

As outlined in Section 5.2.4, existing SOCS data were used to simulate a pilot study to estimate the minimum sampling effort to detect a change in SOCS of 2 or 5 $Mg.ha^{-1}$. The result from the pilot study was that a minimum of 52 soil sampling sites would be required to detect a change in SOCS of 2 $Mg.ha^{-1}$ between two sampling rounds (using a 5 $Mg.ha^{-1}$ value for the standard deviation of the differences of closely spaced SOCS samples), while 11 sites would be required to detect a change of 5 $Mg.ha^{-1}$ (Section 5.2.4). The cost of soil sampling in the pilot study itself was not included in these calculations. For the purposes of costings, the number of sampling sites was kept at a multiple of four to ensure that an even number of samples were 'collected' from each of four strata for the different scenarios (see Figure 7 and Table 2) (i.e. 52/4 = 13 samples per stratum). The number of samples was increased for the bulking option, so the FAO (2019) recommendation of having a minimum of five samples per bulked sample was followed (i.e. increase the number of samples per stratum to 15, a multiple of five). This yields a total of 60 samples (i.e. 15×4 samples total). Costings also followed the FAO (2019) recommendation that samples from (at least) *"three discrete, contiguous and successive soil layers should be available to describe how bulk density and SOC concentrations change from the surface layer downward",* for example, layers of 0–15 cm, 15–30 cm and 30–60 cm. It was assumed that sampling was via use of deep continuous cores (Figure 11).

The total estimated costs for benchmarking when aiming to detect a change of 2 $Mg.ha^{-1}$ was $9,552 when analysing composite samples, compared with $17,129 if all samples were analysed individually. Total benchmarking costs for a system designed to detect a change of 5 $Mg.ha^{-1}$ would cost $7,493 (using the individual sample option). Costs for each subsequent **monitoring** round were estimated to be the same as the benchmarking round minus the study design value used (e.g. 5,552 for the composite sample option to detect a change of 2 $Mg.ha^{-1}$). Costs for the hill country Massey Tuapaka farm would be at least four times those for Dairy 1.

Table 2: Estimated costs for soil organic carbon stock benchmarking and monitoring for the 143 ha Massey University No. 1 Dairy farm. Costings are based on the estimated number of samples required to detect a change in SOCS of 5 $Mg.ha^{-1}$ between two samplings should such a change occur. Hourly rates are based on typical Crown Research Institutescientist and technician charge out rates, with sample preparation and analysis costs based on Manaaki Whenua's IANZ accredited Environmental Chemistry Laboratory. Note: costs for a pilot study (should it be needed to determine SOCS spatial variability) are not included

Activity	Hrs	$ hr or sample	Individual samples for 2 $Mg.ha^{-1}$ change $ totals	Composite samples for 2 $Mg.ha^{-1}$ change $ totals	Individual samples for 5 $Mg.ha^{-1}$ change $ totals
Study design					
Liaise with landowner, define objectives, check management history, compile relevant datalayers, delineate the CEA and strata, estimate/calculate the number of sampling sites required and determine random sampling locations.	20	200	**4,000**	**4,000**	**4,000**
Number of strata and samples			**n**	**n**	**n**
Number of strata selected	-	-	4	4	4
Number of cores taken per strata	-	-	13	15	3
Total number of cores taken	-	-	52	60	12
Number of depth increments per core	-	-	3	3	3
Number of composite samples per strata	-	-	-	3	-
Total number of soil samples to process and analyse	-	-	**156**	**36**	**36**
Soil sampling					
Travel to and from the farm (assuming 80 km total and $0.9 per km + time cost)	1	130	**202**	**202**	**202**
Hours required to sample soils assuming 0.33 hrs per core to 60 cm, inc. travel between cores	→		*17*	*20*	*4*
Costs for soil sampling		130	2,231	2,574	515
Subtotal – soil sampling			**2,433**	**2,776**	**717**
Sample preparation and analysis					
Air-drying and sieving through <2-mm mesh	-	17	2,652	612	612
Air-dry moisture content	-	9	1,404	324	324
Total carbon (and nitrogen) analysis via LECO furnace	-	31	4,836	1,116	1,116
Bulk density analysis	-	9	1,404	324	324
Subtotal – sample preparation and analysis		**66**	10,296	2,376	2,376
Data compilation and calculations	2	200	**400**	**400**	**400**
Total for study design and benchmarking			**17,129**	**9,552**	**7,493**
Ongoing cost per monitoring time			**13,129**	**5,552**	**3,493**

[1]In the case of Massey No. 1 Dairy Farm, travel would be less than this, but this value has been included because it is estimated to be a typical distance required to be travelled for soil sampling.

6 References

Arrouays, D., McKenzie, N.J., Hempel, J., de Forges, A.R., McBratney, A.B. (Eds.), 2014. GlobalSoilMap: Basis of the global spatial soil information system. CRC Press, London.

Australian Government. 2018. Carbon Credits (Carbon Farming Initiative – Measurement of Soil Carbon Sequestration in Agricultural Systems) Methodology Determination 2018, http://www.environment.gov.au/climate-change/government/emissions-reduction-fund/methods/measurement-of-soil-carbon-sequestration-in-agricultural-systems (last access: 20 June 2019).

Bain, L.J., Engelhardt, M. 1992. Introduction to Probability and Mathematical Statistics. Duxbury.

Baldock, J., Burgess, R., 2017. Soil carbon accounting – the Australian example. Workshop on "Soil Organic Carbon Measurement for Management and Management for Measurement", Available from https://jahnresearchgroup.cals.wisc.edu/wp-content/uploads/sites/223/2017/08/3.-Soil-carbon-accounting-in-Australia-1.pdf. (last access: 9 July 2019).

Baldock, J.A., Hawke, B., Sanderman, J, Macdonald, L.M., 2014. Predicting contents of carbon and its component fractions in Australian soils from diffuse reflectance mid-infrared spectra. Soil Research 51(8), 577-595.

Barnett, A.L., Schipper, L.A., Taylor, A., Balks, M.R., Mudge, P.L., 2014. Soil C and N contents in a paired survey of dairy and dry stock pastures in New Zealand. Journal of Agriculture, Ecosystems & Environment, 185, 34-40.

Beef and Lamb New Zealand, n.d. Canterbury farm environment plan guidelines, https://beeflambnz.com/sites/default/files/factsheets/pdfs/RB4-Canterbury-FEP-guidelines.pdf (last access: 29 June 2019).

Bellon-Maurel, V., McBratney, A., 2011. Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils – Critical review and research perspectives. Soil Biology and Biochemistry 43(7), 1398-1410.

Benites, V.M., Machado, P.L.O.A., Fidalgo, E.C.C., Coelho, M.R., Madari, B.E., 2007. Pedotransfer functions for estimating soil bulk density from existing soil survey reports in Brazil. Geoderma 139(1), 90-97.

Bouma, J., 1989. Using soil survey data for quantitative land evaluation. In: B.A. Stewart (Ed.), Advances in soil science. Springer Verlag, New York, pp. 177-213.

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

Brus, D.J., 2014. Statistical sampling approaches for soil monitoring. European Journal of Soil Science 65(6), 779-791.

Brus, D.J., 2015. Balanced sampling: A versatile sampling approach for statistical soil surveys. Geoderma 253-254, 111-121.

Brus, D.J., de Gruijter, J.J., 2011. Design-based Generalized Least Squares estimation of status and trend of soil properties from monitoring data. Geoderma 164(3), 172-180.

Brus, D.J., de Gruijter, J.J., 2013. Effects of spatial pattern persistence on the performance of sampling designs for regional trend monitoring analyzed by simulation of space–time fields. Computers & Geosciences 61, 175-183.

Canterbury Water, 2019. Farm Environment Plans, https://www.canterburywater.farm/fep (last access: 29 June 2019).

Chappell, A., Baldock, J., Viscarra Rossel, R.A., 2013. Sampling soil organic carbon to detect change over time. CSIRO Publication, Canberra.

Cohen, J. 1988. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. Lawrence Erlbaum, Hillsdale,NJ.

Conant, R.T., Cerri C.E.P., Osborne B.B., Paustian K., 2017. Grassland management impacts on soil carbon stocks: a new synthesis. Ecological Applications 27(2), 662-668.

Conant, R.T., Paustian, K., 2002. Potential soil carbon sequestration in overgrazed grassland ecosystems. Global Biogeochemical Cycles 16(4), 1143.

Conyers, M.K., Poile, G.J., Oates, A.A., Waters, D., Chan, K.Y., 2011. Comparison of three carbon determination methods on naturally occurring substrates and the implication for the quantification of ‘soil carbon’. Soil Research 49(1), 27-33.

Dalenius, T., Hodges, J.L., 1959. Minimum Variance Stratification. Journal of the American Statistical Association 54(285), 88-101.

Dangal, S.R., Sanderman, J., Wills, S., Ramirez-Lopez, L., 2019. Accurate and precise prediction of soil properties from a large mid-infrared spectral library. Soil Systems, 3(1), 11.

Davidson, J.M., Biggar, J.W., Nielsen, D.R., 1963. Gamma-radiation attenuation for measuring bulk density and transient water flow in porous materials. Journal of Geophysical Research (1896-1977) 68(16), 4777-4783.

Davis, M., Wilde, H., Garrett, L., Oliver, G., 2004. New Zealand Carbon Monitoring System Soil Data Collection manual. Caxton Press, Christchurch, New Zealand.

de Gruijter, J.J., Brus, D.J., Bierkens, M.F.P., Knotters, M., 2006. Sampling for Natural Resource Monitoring. Springer-Verlag Berlin Heidelberg, The Netherlands.

de Gruijter, J.J., McBratney, A.B., Minasny, B., Wheeler, I., Malone, B.P., Stockmann, U., 2016. Farm-scale soil carbon auditing. Geoderma 265, 120-130.

Eggleston, S., Buendia, L., Miwa, K., Ngara, T., Tanabe, K., Intergovernmental Panel on Climate Change (IPCC)/Institute for Global Environmental Strategies (IGES) (Eds.), 2006. IPCC Guidelines for National Greenhouse Gas Inventories. Intergovernmental Panel on Climate Change Hayama, Japan.

England, J.R., Viscarra Rossel, R.A., 2018. Proximal sensing for soil carbon accounting. SOIL 4(2), 101-122.

Ge, Y., Morgan, C.L.S., Ackerson, J.P., 2014. VisNIR spectra of dried ground soils predict properties of soils scanned moist and intact. Geoderma 221-222, 61-69.

FAO. 2019. Measuring and modelling soil carbon stocks and stock changes in livestock production systems: Guidelines for assessment (Version 1). Livestock Environmental Assessment and Performance (LEAP) Partnership. Rome, FAO. 170 pp. Licence: CC BY-NC-SA 3.0 IGO.

Faul F., Erdfelder E., Lang A.-G., Buchner A., 2007. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39, 175-191.

Faul F., Erdfelder E., Buchner A., Lang A.-G., 2009. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behavior Research Methods, 41, 1149-1160.

Giltrap, D.J., Hewitt, A.E., 2004. Spatial variability of soil quality indicators in New Zealand soils and land uses, New Zealand Journal of Agricultural Research, 47(2), 167-177.

Grealish, G., Clifford, D., Wilson, P., Ringroase-Voase, A., 2011. National soil condition monitoring program for soil pH and soil carbon. CSIRO Land and Water Science Report 05/11 CSIRO, Canberra.

Ellert, B.H., Janzen, H.H., McConkey, B.G., 2001. Measuring and comparing soil carbon storage. Lewis Publishers, Boca Raton, pp. 131-146.

Heanes, D.L., 1984. Determination of total organic-C in soils by an improved chromic acid digestion and spectrophotometric procedure. Communications in Soil Science and Plant Analysis 15(10), 1191-1213.

Hedley, C., Mudge, P., McNeill, S., Roudier, P., Koele, N., 2020. Reference manual for implementation of a soil organic carbon monitoring programme. Landcare Research Contract Report LC3773. 55 p.

Hedley, C.B., Payton, I.J., Lynn, I.H., Carrick, S.T., Webb, T.H., McNeill, S., 2012. Random sampling of stony and non-stony soils for testing a national soil carbon monitoring system. Soil Research 50, 18-29.

Hedley, C.B., Roudier, P., Maddi, L., 2015. Vis-NIR soil spectroscopy for field soil analysis. Communications in Soil and Plant Analysis 46, 1-18

Hewitt, A.E., Burgham, S.J., Gibson, R.S., 1998. Spatial variability of organic carbon in a tussock grassland, Manorburn, Central Otago, New Zealand, New Zealand Journal of Agricultural Research, 41(4), 613-622.

Holmes, K.W., Wherrett, A., Keating, A., Murphy, D.V., 2012. Meeting bulk density sampling requirements efficiently to estimate soil carbon stocks. Soil Research 49(8), 680-695.

Ji, W., Viscarra Rossel, R.A., Shi, Z., 2015. Accounting for the effects of water and the environment on proximally sensed vis–NIR soil spectra and their calibrations. European Journal of Soil Science 66(3), 555-565.

Jobbágy, E.G., Jackson, R.B., 2000. The vertical distribution of soil organic carbon and its relation to climate and vegetation. Ecological Applications 10(2), 423-436.

Kelliher, F.M., Parfitt, R.L., van Koten, C., Schipper, L.A., Rys, G., 2013. Use of shallow samples to estimate the total carbon storage in pastoral soils. New Zealand Journal of Agricultural Research, 56, 86-90.

Kirschbaum MUF, Schipper LA, Mudge PL, Rutledge S, Puche NJB, Campbell DI 2017. The trade-offs between milk production and soil organic carbon storage in dairy systems under different management and environmental factors. Science of the Total Environment 577, 61-72.

Lark, R.M., 2009. Estimating the regional mean status and change of soil properties: two distinct objectives for soil survey. European Journal of Soil Science 60(5), 748-756.

Lark, R.M., 2012. Some considerations on aggregate sample supports for soil inventory and monitoring. European Journal of Soil Science 63(1), 86-95.

Lobsey, C.R., Viscarra Rossel, R.A., 2016. Sensing of soil bulk density for more accurate carbon accounting. European Journal of Soil Science 67(4), 504-513.

Lobsey, C.R., Viscarra Rossel, R.A., Roudier, P., Hedley, C.B., 2017. rs-local data-mines information from spectral libraries to improve local calibrations. European Journal of Soil Science 68(6), 840-852.

Lohr, S.L., 2010. Sampling: Design and Analysis. Brooks/Cole, Boston, MA

Malone, B., Hedley, C., Roudier, P., Minasny, B., Jones, E., McBratney, A., 2018. Auditing on-farm soil carbon stocks using downscaled national mapping products: Examples from Australia and New Zealand. Geoderma Regional 13, 1-14.

McBratney, A.B., Minasny, B., de Gruijter, J.J., 2015. Optimizing Stratification and Allocation for Design-Based Estimation of Spatial Means Using Predictions with Error. Journal of Survey Statistics and Methodology 3(1), 19-42.

McConkey, B., Haugen-Kozyra, K., Staley, D., 2013. Prairie soil carbon balance project summary: Soil organic carbon change on direct-seeded farmland in Saskatchewan, In Proceedings of: Soils and Crops 2013 Conference, Saskatoon, SK, Canada.

McDonald, T., 2016. SDraw: Spatially Balanced Sample Draws for Spatial Objects. R package version2.1.3. https://CRAN.R-project.org/package=SDraw.

McKenzie, N., Henderson, B., McDonald, W.S., 2002. Monitoring soil change. Principles and proactices for Australian conditions. CSIRO Publication, Canberra.

McLaughlin, M., Soriano Disla, J.M., Janik, L., Fruzangohar, M., 2017. New Infrared Instrumentation For Soil Analysis Field Applications. Grains Research & Development Corporation, Urrbrae, SA.

McNeill SJE, Golubiewski N, Barringer J 2014. Development and calibration of a soil carbon inventory model for New Zealand. Soil Research 52(8), 789–804.

McNeill SJ, Odgers N, Lilburne L, Carrick S, Hainsworth S, Fraser S, Hedley C, Roudier P, Grealish G 2018. Spatial sampling tools for representative soil sampling. In: Farm environmental planning – Science, policy and practice. (Eds L. D. Currie and C. L. Christensen). http://flrc.massey.ac.nz/publications.html. Occasional Report No. 31. Fertilizer and Lime Research Centre, Massey University, Palmerston North, New Zealand. 8 p.

McNeill S., Mudge, P., Hedley C., Roudier P., Schipper, L., 2019. Statistical design of a national soil carbon monitoring programme for New Zealand. Manaaki Whenua Landcare Research Report LC3459 for Ministry for Primary Industry, Wellington, New Zealand.

Mehler, K., Schöning, I., Berli, M., 2014. The Importance of Rock Fragment Density for the Calculation of Soil Bulk Density and Soil Organic Carbon Stocks. Soil Science Society of America Journal 78(4), 1186-1191.

Minasny, B., McBratney, A.B., 2002. FuzME version 3.0. Australian Centre for Precision Agriculture, The University of Sydney, Australia.

Minasny, B., McBratney, A.B., Bellon-Maurel, V., Roger, J.-M., Gobrecht, A., Ferrand, L., Joalland, S., 2011. Removing the effect of soil moisture from NIR diffuse reflectance spectra for the prediction of soil organic carbon. Geoderma 167-168, 118-124.

Minasny, B., McBratney, A.B., Tranter, G., Murphy, B.W., 2008. Using soil knowledge for the evaluation of mid-infrared diffuse reflectance spectroscopy for predicting soil physical and mechanical properties. European Journal of Soil Science 59(5), 960-971.

Moreira, C.S., Brunet, D., Verneyre, L., Sá, S.M.O., Galdos, M.V., Cerri, C.C., Bernoux, M., 2009. Near infrared spectroscopy for soil bulk density assessment. European Journal of Soil Science 60(5), 785-791.

Morgan, M.G., Henrion, M., Small, M., 1990. Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. Cambridge University Press, USA.

Mudge PL, Kelliher FM, Knight TL, O'Connell D, Fraser S, Schipper LA 2017. Irrigating grazed pasture decreases soil carbon and nitrogen stocks. Global Change Biology 23(2), 945–954.

Nelson DW, LE, S., 1996. Total carbon, organic carbon, and organic matter. In: D.L. Sparks, A.L. Page, P.A. Helmke, R.H. Loeppert, P.N. Soltanpour, M.A. Tabatabai, C.T. Johnston, M.E. Sumner (Eds.), Methods of soil analysis. Part 3: Chemical methods. Soil Science Society of America, Inc., American Society of Agronomy Inc, Madison, WI, pp. 961-1010.

Neyman J. 1934. On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. Journal of the Royal Statistical Society. 97, 558-606.

Ng, W., Minasny, B., Montazerolghaem, M., Padarian,J., Ferguson, R., Bailey, S., McBratney, A.B., 2019. Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra. Geoderma 352, 251-267.

Odeh, I.O.A., McBratney, A.B., Chittleborough, D.J., 1992. Soil Pattern Recognition with Fuzzy-c-means: Application to Classification and Soil-Landform Interrelationships. Soil Science Society of America Journal 56(2), NP-NP.

Oliver, M.A., Webster, R., 2014. A tutorial guide to geostatistics: Computing and modelling variograms and kriging. Catena 113, 56-69.

Parfitt, R.L., Stevenson, B.A., Ross, C., Fraser, S., 2014. Changes in pH, bicarbonate-extractable-P, carbon and nitrogen in soils under pasture over 7 to 27 years, New Zealand Journal of Agricultural Research 57(3), 216-227.

Poeplau, C., Vos, C., Don, A., 2017. Soil organic carbon stocks are systematically overestimated by misuse of the parameters bulk density and rock fragment content. SOIL 3(1), 61-66.

R Core Team. 2020. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Ramirez-Lopez, L., Behrens, T., Schmidt, K., Rossel, R.A.V., Demattê, J.A.M., Scholten, T., 2013. Distance and similarity-search metrics for use with soil vis–NIR spectra. Geoderma 199, 43-53.

Robertson, B.L., Brown, J.A., McDonald, T.L., Jaksons. P., 2013. BAS: Balanced Acceptance Sampling of Natural Resources, Biometrics 69, 776-784.

Rodríguez-Lado, L., Rial, M., Taboada, T., Cortizas, A.M., 2015. A Pedotransfer Function to Map Soil Bulk Density from Limited Data. Procedia Environmental Sciences 27, 45-48.

Roudier, P., Hedley, C., Lobsey, C., Viscarra Rossel, R., Leroux, C., 2017. Evaluation of two methods to eliminate the effect of water from soil Vis-NIR spectra for predictions of organic carbon. Geoderma 296, 98-107.

Roudier, P., Hedley, C.B., Ross, C. 2015. Prediction of volumetric soil organic carbon from field-moist intact soil cores. European Journal of Soil Science 66(4), 651-660.

Saby, N., Bellany, P., Morvan, X., Arrouays, D., Jones, R., Verheijen, F., Kibblewhite, M., Verdoodt, A., Üveges, J., Fruedenschub, A., Simota, C., 2008. Will European soil-monitoring networks be able to detect changes in topsoil organic carbon content? Global Change Biology 14(10), 2432-2442.

Schipper, L.A., Parfitt, R.L., Fraser, S., Littler, R.A., Baisden, W.T., Ross C.W., 2014. Soil order and grazing management effects on changes in soil C and N in New Zealand pastures. Agriculture Ecosystems & Environment 184, 67-75.

Searle, R., 2014. The Australian site data collation to support the GlobalSoilMap. In: D. Arrouays, N. McKenzie, J. Hempel, A.R. de Forges, A.B. McBratney (Eds.), GlobalSoilMap: Basis of the global spatial soil information system. CRC Press, The Netherlands, pp. 127-132.

Shi, Z., Ji, W., Viscarra Rossel, R.A., Chen, S., Zhou, Y., 2015. Prediction of soil organic matter using a spatially constrained local partial least squares regression and the Chinese vis–NIR spectral library. European Journal of Soil Science 66(4), 679-687.

Singh, K., Murphy, B.W., Marchant, B.P., 2013. Towards cost-effective estimation of soil carbon stocks at the field scale. Soil Research 50(8), 672-684.

Smith, P., Soussana, J.F., Angers, D., Schipper, L., Chenu, C., Rasse, D.P., Batjes, N.H., van Egmond, F., McNeill, S., Kuhnert, M., Arias-Navarro, C., 2020. How to measure, report and verify soil carbon change to realize the potential of soil carbon sequestration for atmospheric greenhouse gas removal. Global Change Biology 26(1), 219-241.

Soriano-Disla, J.M., Janik, L.J., Viscarra Rossel, R.A., Macdonald, L.M., McLaughlin, M.J., 2014. The Performance of Visible, Near-, and Mid-Infrared Reflectance Spectroscopy for Prediction of Soil Physical, Chemical, and Biological Properties. Applied Spectroscopy Reviews 49(2), 139-186.

Stevens, D.L., Olsen. A.R., 2004. Spatially balanced sampling of natural resources. Journal of the American Statistical Association 99, 262-278.

Stockmann, U., Adams, M.A., Crawford, J.W., Field, D.J., Henakaarchchi, N., Jenkins, M., Minasny, B., McBratney, A.B., Courcelles, V.d.R.d., Singh, K., Wheeler, I., Abbott, L., Angers, D.A., Baldock, J., Bird, M., Brookes, P.C., Chenu, C., Jastrow, J.D., Lal, R., Lehmann, J., O'Donnell, A.G., Parton, W.J., Whitehead, D., Zimmermann, M., 2013. The knowns, known unknowns and unknowns of sequestration of soil organic carbon. Agriculture, Ecosystems & Environment 164, 80-99.

Bishop, T.F.A., Pringle, M.J., 2012. Paddock-scale soil carbon inventory and monitoring: a meta-analysis for sample size requirements. Proceedings of the 2012 Joint Australian and New Zealand Soil Science Societies Conference, Hobart, Tasmania, pp. 555-558.

Tang, Y., Jones, E., Minasny, B., 2020. Evaluating low-cost portable near infrared sensors for rapid analysis of soils from South Eastern Australia. Geoderma Regional, 20, e00240. doi:https://doi.org/10.1016/j.geodrs.2019.e00240

Tranter, G., Minasny, B., Mcbratney, A.B., Murphy, B., Mckenzie, N.J., Grundy, M., Brough, D., 2007. Building and testing conceptual and empirical models for predicting soil bulk density. Soil Use and Management 23(4), 437-443.

Valette, L., 2013. Farm-scale mapping of soil organic carbon on a hill country farm using visible near infrared spectroscopy. Thesis submitted in partial fulfilment of M.Sc. specialising in Agro-technology, Supagro, Montpellier, France, and Massey University, New Zealand, September 2013, hosted and supervised by Landcare Research NZ. 60 p.

VERRA 2018. Verified Carbon Standard. https://verra.org/project/vcs-program/

Viscarra Rossel, R.A., Brus, D.J., Lobsey, C., Shi, Z., McLachlan, G., 2016. Baseline estimates of soil organic carbon by proximal sensing: Comparing design-based, model-assisted and model-based inference. Geoderma 265, 152-163.

Viscarra Rossel, R.A., Lobsey, C., 2016. A new method for rapid measurements of gravel for soil organic carbon accounting. CSIRO Reports, Canberra.

Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., Skjemstad, J.O., 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131(1), 59-75.

von Haden, A.C., Yang, W.H., DeLucia, E.H., 2020. Soils' dirty little secret: Depth-based comparisons can be inadequate for quantifying changes in soil organic carbon and other mineral soil properties. Global Change Biology 26: 3759– 3770. https://doi.org/10.1111/gcb.15124

Walkley, A., Black, I.A., 1934. An examination of the Degtjareff method for determining soil organic matter, and a proposed modifcation of the chromic acid titration method. Soil Science 37(1), 29-38.

Walvoort, D.J.J., Brus, D.J., de Gruijter, J.J., 2010. An R package for spatial coverage sampling and random sampling from compact geographical strata by k-means. Computers & Geosciences 36(10), 1261-1267.

Wang JF, Stein A, Gao, BB, Ge Y. A review of spatial sampling. Spatial Statistics 2012;2:1-14.

Wang, Z., Yi, S., Chen, A., Zhou, M., Luk, T.S., James, A., Nogan, J., Ross, W., Joe, G., Shahsafi, A., Wang, K.X., Kats, M.A., Yu, Z., 2019. Single-shot on-chip spectral sensors based on photonic crystal slabs. Nature Communications 10(1), 1020.

Wendt, J.W., Hauser, S., 2013. An equivalent soil mass procedure for monitoring soil organic carbon in multiple soil layers. European Journal of Soil Science 64(1), 58-65.

Wheeler, I., 2014. Auditability of soil carbon at the farm scale (PhD Thesis), The University of Sydney, Sydney.

Wright T. 2017. Exact optimal sample allocation: More efficient than Neyman. Statistics and Probability Letters 129:50–57.

Appendix 1: Mathematical development of sample size requirement

In this Appendix we develop the mathematical method that can be used to estimate the required sample size from a pilot study of a farm or stratum. In this development, we do not provide a step-by-step justification of the steps, but instead make use of several fundamental relationships, as follows:

1 If X is some statistical quantity that is Normally distributed with mean μ and standard deviation σ, then we say $X \sim N(\mu, \sigma^2)$.

2 If $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$, then $X - Y \sim N(\mu_X - \mu_Y, \sigma_X^2 + \sigma_X^2 - 2\rho\sigma_X\sigma_Y)$, where ρ is the correlation between X and Y.

3 If $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$, then $X + Y \sim N(\mu_X - \mu_Y, \sigma_X^2 + \sigma_X^2 + 2\rho\sigma_X\sigma_Y)$, where ρ is the correlation between X and Y.

4 If $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$ and the correlation between X and Y is zero then $X - Y \sim N(\mu_X - \mu_Y, \sigma_X^2 + \sigma_X^2)$ and $X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_X^2)$.

5 If $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$, and if $\sigma_X = \sigma_Y = \sigma$ (in other words, the standard deviations are the same) then $X - Y \sim N\left(\mu_X - \mu_Y, 2\sigma^2(1-\rho)\right)$ and $X + Y \sim N\left(\mu_X - \mu_Y, 2\sigma^2(1+\rho)\right)$.

The details of the above relationships can be found in standard texts (e.g. Bain and Englehardt 1991).

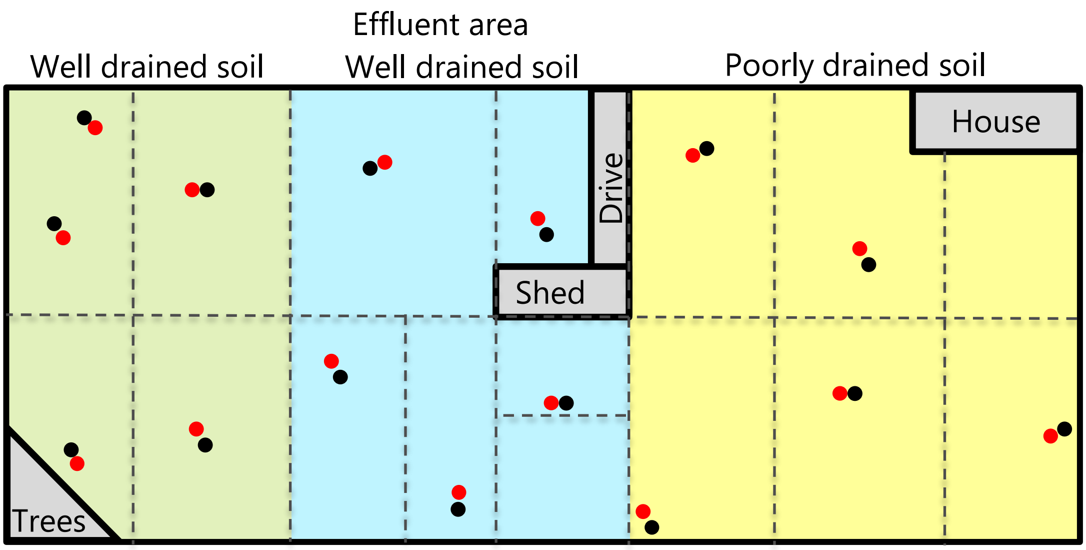

Figure 16. Example layout of a pilot study for a farm with five sample pairs in a stratum. The second element of each pair (red dot) is located a distance of 5 m from the first, in a random quarter-circle direction from North. The grey areas represent a house, drive,shed and forested area which are excluded from the carbon estimation area.

Assume that there are N_P pilot point pairs. Each pair of points is spaced a fixed small distance apart, with a random orientation between them. The point pair, or at least the centre of the point pair, is located randomly over the sampling domain. Figure 16 shows the point pair layout in schematic form. FAO recommends a minimum of five locations for the pilot survey (FAO 2019).

Assume that SOCS is Normally distributed with mean μ and variance σ_i^2, where σ_i is the standard deviation of independent SOCS samples. That is, the SOCS is distributed as $N(\mu, \sigma_i^2)$. Assume that we carry out the pilot study field work and then calculate the SOCS difference between each of the point pairs (i.e. the difference in SOCS between locations a small distance apart), with the measurements taken at the same point in time. The difference will be distributed $N(0, \sigma_s^2)$, if the standard deviation of the difference is σ_s. The difference between the two SOCS measurements taken at the same time should have a mean of zero, since the points are Normally distributed.

If a second set of SOCS measurements is taken some years later, the SOCS at each location will have changed up or down during that period. The mean change in SOCS over time is assumed to be fixed (in other words, it is spatially invariant), with a standard deviation of σ_t. Generally, σ_t is difficult to measure, and its value must be estimated from observed change in SOCS from different sites. A figure of 0.6 Mg.ha^{-1} is suitable. Since the spatial and temporal standard deviation components σ_s and σ_t are independent, the standard deviation of the difference in SOCS between the first and second measurement dates $\sigma_{\text{diff}} = \sqrt{\sigma_s^2 + \sigma_t^2}$.

Now we consider the estimation of the required minimum sample size from measurements of σ_s, and specification of the minimum change in the SOCS (∇SOCS), a significance level (Type-I error or α), and target power ($1 - \text{Type-II error rate}$ or β). The appropriate power calculation is a paired-sample two-sided t-test. The sample size estimation can be carried out in standard software, such as the *power.t.test* function in the standard R library[5]. Alternatively, the calculations may be carried out in a statistical power calculator, such as *G*Power*[6] (Faul et al. 2007, Faul et al. 2009). The calculations typically estimate the power size in terms of the effect size d:

$$d = \frac{\nabla\text{SOCS}}{\sigma_{\text{diff}}}.$$

[5] An appropriate reference is https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/power.t.test.

[6] See https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower.html.