Sampling soil organic carbon to detect change over time

Adrian Chappell, Jeff Baldock, Raphael Viscarra Rossel

June 18, 2013

Report to Grains Research and Development Corporation and Australian Department of Environment

Australian Government

Department of the Environment

GRDC

Grains Research & Development Corporation

Using the Commonwealth Coat of Arms

The terms of use for the Coat of Arms are available from the *It's an Honour* website.

IMPORTANT NOTICE – PLEASE READ

While reasonable efforts have been made to ensure that the contents of this publication are factually correct, the Commonwealth does not accept responsibility for the accuracy or completeness of the contents, and expressly disclaims liability for any loss or damage whether due to negligence or otherwise however caused that may be occasioned directly or indirectly through the use of, or reliance on, the contents of this publication.

Material in this publication is made available on the understanding that the Commonwealth is not providing professional advice and should not be taken to indicate the Commonwealth's commitment to a particular course of action. Different solutions and outcomes may apply in individual circumstances.

Before relying on any material contained in this publication, readers should obtain appropriate professional advice suitable to their particular circumstances.

References to various websites, publications and organisations, including commercial organisations and/or particular products in this publication are included for the information of the reader only and should not in any way be construed as an endorsement of any website, publication, organisation or product by the Commonwealth. Conversely, the fact that a particular website, publication, organisation or product is not mentioned in this publication should not be taken as any indication of the Commonwealth's opinion of that website, publication, organisation or product. Neither the Department of Environment nor the Commonwealth can accept any responsibility for the content of any material that may be encountered on the websites or in the publications referred to in this publication.

Contents

Figures

Tables

Acknowledgments

---

[1] The Climate Change functions of the former Department of Climate Change and Energy Efficiency are now administered by the Department of Environment.

Executive summary

Scope and purpose of sampling soil organic carbon to detect change over time

Measuring an increase in soil organic carbon stocks (SOC) over a carbon estimation area (CEA), be it an entire field or a part of a field, and over a period of time is a prerequisite to demonstrating that SOC has been stored in the CEA. Our remit was to describe a generic monitoring design that could be widely applied to detect temporal changes in SOC across a CEA ($\hat{\bar{d}}_{2,1}$) with no prior knowledge of the spatial or temporal variance of SOC within the CEA. In this situation the measurements represent the net balance between the storage or loss of SOC imposed by management practice and the loss or addition due to soil erosion and its enrichment / depletion ($\hat{\bar{d}}_{2,1} + E_c P$). Only by measuring the impact of soil erosion, can the true impact of management practice on SOC be quantified. Where net removal of SOC by soil erosion exceeds net storage due to management, no measured gain in SOC would be measured even when management practices proven to enhance SOC are implemented. Quantifying SOC movement due to erosion will therefore avoid mistaken assessments of the performance of management practice. It will also encourage adoption of management practices that store SOC to be consistent with approaches to soil conservation.

Measuring SOC change across a CEA requires collection of sufficient soil samples across the CEA to ensure that individual temporal measurements of SOC and any temporal change are representative of the entire CEA and that the variability about the measured values is as small as possible. If too few samples are taken, it is unlikely that the measured values of SOC will be representative of the CEA and the variance will be large. If the variance in SOC measured at the first sampling time is large then it will be difficult after the second sampling to demonstrate that SOC have changed over the CEA (i.e. there will be very little confidence in any measured SOC change).

The initial stages of setting up and implementing a sampling design are challenging because there is typically little information available on the variability SOC within the CEA. It is therefore difficult to establish what type of sampling design to use and how many samples to take in each sampling round over time. To provide guidance on setting up SOC monitoring in an CEA we have produced a technical document in which we have conducted a judicious review of the literature and defined the statistical justifications for recommendations of SOC monitoring (Chappell *et al.* 2013 (CPL13). The aim here is to provide a non-technical summary of CPL13 and outline our recommended approach to designing a soil sampling and monitoring program to detect changes in SOC over time within a CEA.

There are three distinct phases to the monitoring of SOC:

1. The setup of the sampling design and the assumptions / decisions necessary for implementation to reduce the magnitude of a detectable change in SOC over time at an acceptable cost.

2. Response and potential adaptation of the sampling design after measurements of SOC are made in the first round of sampling (i.e., more or less SOC variance than expected?).

3. Determination of the actual detectable change in SOC by comparing SOC measurements from the first round of sampling with the second round of sampling (e.g., 5 years later) using a means difference test.

We have restricted our consideration of sampling designs and their use over time to the most basic and easily adopted approaches. There are sound statistical bases for their use in reducing variation in space and time and we outline these bases below. Even with these restricted basic sampling designs they have the potential to become sophisticated if adapted over time to account for a mismatch between expected and measured variations in SOC. The expected variation in SOC is formed from a series of decisions that may be used to estimate the minimum detectable change (MDC). Once samples are obtained and measurements made the variance for the specific CEA and the associated MDC can be revised and compared to the initial

estimate. The expectation of variation in SOC within a CEA both spatially and through time requires that a series of decisions be made including:

1. How much does it cost for a soil professional to get to your CEA, sample your field, prepare and analyse the collected soil samples, and calculate SOC? How much variance or error is there in the measurement of SOC?

2. How much confidence do you want to have on the outcomes of the monitoring e.g., do you want to be 95% confident in the results from your sampling design or can you tolerate 80% confidence? The confidence you choose will affect the size of the SOC change that can be detected over time (and therefore the number of credits awarded) and determine the number of samples required to detect the SOC change (and therefore the sampling and measurement costs).

3. How large is your CEA? Although SOC stored in a large CEA will return a greater amount of SOC credits than a small CEA, the change in SOC must be detectable. If a large CEA comprises areas which are highly variable, perhaps because areas with different land use histories or erosion histories have been combined, it will be more difficult to reduce the variance associated with measured changes in SOC and therefore the ability to detect SOC change. In this situation, it may be better to reduce the CEA to a smaller (or several small areas), relatively homogeneous (in space and time) area, where the detection of SOC change is more likely. .

4. What is the spatial variability of SOC in your CEA? It is difficult to know at the outset how spatially variable SOC is in your CEA, but estimates can be made based on existing data from experimental sites. If a CEA is highly variable in space, more samples will be required to detect SOC change.

5. What is the temporal variability of SOC in your CEA? It is difficult to know at the outset how SOC changes over time in your CEA, but estimates can be made based on existing data from experimental sites. If a CEA is highly variable over time, more samples will be required to detect the impact of an applied management practice on SOC change.

6. What is the target minimum detectable change in SOC in your CEA? This is the minimum SOC change that you would like to be able to detect when the mean SOC is measured after the first sampling round and then again at the second sampling round (see Figure 1).

Based on assumptions of the spatial and temporal variation in the CEA, the likely measurement error associated with quantifying SOC and expectations about the confidence of the results, we can estimate the variance of the mean change in SOC.

The final assumption of this methodology relates to the variability of SOC with depth. SOC typically declines with increasing soil depth. However, there are instances where this is not the case. The variation of SOC across a CEA is most likely to be largest in the soil layer which has the largest SOC content (depending on changes in bulk density). This tends to produce less spatial variation in SOC at depth than near the surface. On this basis it appears that fewer samples could be used to represent the variability of SOC at depth. However, the benefit of a sampling design that reduces the number of samples for the largest depths is outweighed by the need to visit a greater number of locations to collect samples near the surface. Consequently, we recognise that there is little point in producing a sampling design for separate layers because soil is typically collected for all depths at each sampling site. We assume that the surface SOC (0-10 cm) is the most variable layer and that a soil sample collection protocol designed for this layer will be adequate for collecting soil at depth where SOC is less variable.

Rationale and adaptation of the proposed sampling design

If the intention of monitoring SOC is to detect significant change and attribute that change to a particular land use / management practice then samples must be obtained over the entire CEA (i.e. the area upon which the management practice is being applied). In this situation we are interested in the mean change and the variability in that mean change determines whether the change can be detected. Consequently, we

need to sample adequately across the CEA to represent the different amounts of SOC at different locations (i.e., reduce the variance in SOC). A simple random sampling design may result in some samples being located close to each other (clustered locations) which for a given available resource would be at the expense of sampling other locations in the CEA. A stratified simple random design overcomes this clustering by dividing the CEA into a number of equal area strata and then ensuring that at least one sample is obtained from each stratum.

Since we are interested in obtaining a soil sample that can be analysed to deliver an SOC representative of the entire CEA at the time of sample collection, soil acquired from each collection point in the CEA can be combined to provide a single representative composite sample. Measurement of SOC is then only required for the composite sample, which can reduce analytical costs. Although the mean SOC within the CEA can be established with this single composite sample, an estimate of variance is required to calculate detection significance and cannot be derived from the single sample. Our recommended solution is to create more than one composite sample by collecting multiple aliquots of soil from each stratum in the CEA and combining the aliquots such that each composite contains one aliquot of soil from each CEA stratum. The variance between composite samples can then be used to quantify the confidence in measured SOC across the CEA.

After a period of time (e.g., five years) has elapsed, the second round of sampling will collect new samples for SOC measurement to compare with the results of the first round of sampling. The design of the second round of sampling should be consistent with that used in the first round of sampling. The design we recommend is simple random stratified sampling with the aliquots of soil being collected in close proximity to the location used in the first sample collection. This approach minimises spatial variation and thereby enhances the potential of detecting temporal changes in SOC within the CEA. Following the recommended simple stratified random sampling design with compositing, means that we need to know how many strata and how many composites we should take and to balance those requirements against the cost. At the outset we have little or no information on the variability so we have used published studies to quantify potential variation in space and time. An alternative, at least for spatial variation, is for the first round of sampling to form a reconnaissance survey. In that survey we would need to obtain between 100 and 150 samples which would each need to be measured for SOC. The sampling and measurements costs would be large and the outcome may be to match approximately the assumed characteristics of the spatial variation we found from the literature. Consequently, although the use of the reconnaissance survey is in principal the ideal approach it is likely to be poor value for money in most situations. The exceptions will be those CEAs on farms which are not represented by the published studies. In these situations the expectations about variation derived from previously published data may not match the SOC variance measured after the first round of sampling. However, the variance obtained will be a true reflection of that present in the CEA

Main variables and decisions

The variance in the mean change of SOC is influenced by the spatial and temporal variation and the number of samples (strata and composites). We used published descriptions of SOC spatial variation (variograms) to characterise large, medium and small spatial variation of SOC. We used data from Australian long-term experimental sites to characterise SOC temporal variation (autocorrelation coefficients).

Using the selected variograms and the autocorrelation coefficients we estimated the SOC mean change total variance for a range of strata (1-10) and composites (1-10). The difference in the variograms and their mean values of the semi-variances for the strata causes more variation than differences in the range of autocorrelation coefficients. In other words, there is more variation in space than there is in time based on previously published data. This enables us to simplify the decision-making process and consider the conservative, most temporally variable situation.

Cost structure for monitoring change in carbon stocks

The selection of a target minimum detectable change is necessary to isolate the appropriate sampling cost which in turn is related to the number of strata and the number of composites. We assume the following costs: fieldwork fixed costs=$2000, time to obtain one soil core=0.33 hours, fieldwork costs=$120/hour and laboratory SOC measurement cost=$30/composite. The total cost (excluding fieldwork fixed costs) is calculated for a range of composites of samples with differing numbers of strata. A particular cost band can be used to indicate approximately how many samples (strata and composites) should be used.

Implementation

Based on the decisions made previously we are now able to calculate the minimum detectable change (MDC) of SOC over time based on simple random stratified sampling obtained from a CEA with the same number of samples collected at two points in time (t_1 and t_2). For example, we may have estimated that we can afford to collect soil at 27 locations within the CEA for each sampling event occurring at t_1 and t_2 ($N=N_1+N_2=54$) with a mean change variance of 20 (t SOC ha^{-1})^2 ($\hat{V} = 20$) approximately constant over time. The SOC is to be compared using a one-sided 5% significance test (i.e., α=0.05). We wish to be 95% sure of detecting a difference in mean SOC between t_1 and t_2 (i.e., β=0.05). Inserting these values into the equation for MDC:

$$\hat{\hat{d}}_{2,1} + E_c P = \left(X_{1-\alpha} + X_{1-\beta}\right)\left(\frac{2\hat{V}}{N}\right)^{0.5},$$

gives

$$\hat{\hat{d}}_{2,1} + E_c P = (1.734 + 1.734)\left(\frac{2(20)}{54}\right)^{0.5} = 2.98 \text{ t SOC ha}^{-1}.$$

The previous decisions placed this estimate in the most variable of situations. In this case we would expect the minimum detectable change (MDC) over a five year period to be 2.98 t SOC ha^{-1}. However, $E_c P$ is assumed to be zero which implies neglect of the influence of carbon erosion / deposition. In this situation the estimate will be uncertain and depend on the magnitude and sign of the carbon erosion / deposition. If even a first approximation to this term is provided, the uncertainty of the MDC of SOC will be reduced. In the case of carbon erosion equating to -1 t SOC ha^{-1} the MDC is increased to 3.98 t SOC ha^{-1} which means that more samples and / or less variance is required to maintain a particular MDC. In the case of e.g., carbon deposition +1 t SOC ha^{-1} the MDC is decreased to 1.98 t SOC ha^{-1} which means that fewer samples and / or more variance can be tolerated.

The decisions that were made previously about the conditions for this estimate, including the spatial and temporal variation, have a considerable impact on this estimate. The MDC will differ under different conditions and if we made different decisions about the variability that we might expect in our CEA. Using a common sampling design with difference expectations about the variance of the mean change of SOC will produce considerable differences in the target minimum detectable change (t SOC ha^{-1}).

It is evident from these results that much of what has gone before is dependent on, and therefore sensitive to, the spatial variation that we characterised to be large. This characterisation was based on available data much of which did not originate in Australia. There may be CEAs which are more variable than the situation we have portrayed as large spatial variation. In that case, the measured MDC will not match the expected or target MDC in much the same way as if a medium spatial variation was selected where large spatial variation was appropriate. There will always be a possibility of this happening until every CEA has been sampled. Consequently, in the absence of information on the spatial variation, the best approach is to take as many samples as can be afforded to gain a first approximation and to include an estimate of carbon erosion / deposition.

Bases for designing SOC stock sampling for detecting change

This document focuses on collecting soil samples from within a defined area of interest (CEA) over which the same land use or management practice is implemented. We recognise that some fields contain more than one soil type or landform and consequently the field may be subdivided and used or managed in different ways. As a result, multiple CEAs may exist within a single field or the entire field may be considered as a single CEA.

Much of the statistical definitions follow those of de Gruijter *et al.* (2006) because they provided a comprehensive overview of monitoring land resources. Their book describes the many and varied decisions that must be made in the design and implementation of a monitoring strategy and which include the fundamental question of why the sampling is being undertaken i.e., what is the purpose of monitoring? Our remit was to define a generic monitoring design that could be widely applied to detect temporal changes in SOC across a CEA with no prior knowledge of the spatial or temporal variance of SOC within the CEA.

The underlying purpose of the monitoring is to support the development of a Carbon Farming Initiative (CFI) methodology which enables farmers, land holders and their representatives etc. to measure and detect carbon in a transparent and auditable manner so that SOC sequestration by an 'additional' management practice can be rewarded. In other words, the effects of the additional practice on the sequestration of carbon in soil can be quantified and verified. Although beyond our remit, we describe how the loss and gain of SOC may be affected as much by soil redistribution (erosion and deposition) as it is by land use and management practices. Consequently, we show that unless the amount of SOC moving with any net soil redistribution is measured at the same time as SOC is measured at defined locations within the CEA, it will not be possible to attribute change in SOC to a particular management practice. We refer to work on Australian carbon erosion, soil erosion and SOC enrichment to rectify the misconception that SOC is merely redistributed within the field by soil erosion. The removal of SOC from the field is readily achieved by wind erosion and dust emission and not uncommon with water erosion.

Amongst the many decisions that must be made in monitoring we follow de Gruijter *et al.* (2006) with respect to the separation of 'design-based' and 'model-based' approaches. We concur with Allen *et al.* (2010) and Bishop and Pringle (2012) that a 'design-based' method is useful to estimate the mean SOC of a field particularly given restrictions on the costs of sampling and measurement. Although also beyond our remit, we show how sampling and monitoring design is dependent on costs and demonstrate a straightforward method to reach a compromise between the number of samples and costs of fieldwork and measurement. Unfortunately, little information is available on the cost of sampling and measurement, particularly the number of replicates used to quantify measurement error. In this report, we assumed values to demonstrate the optimisation process. However, we suggest that a more rigorous investigation of the costs is required.

Not included in the remit of the project, but which should probably be included as part of a protocol for monitoring change in SOC stock, is the soil depth of SOC and how it should be sampled and measured over depth increments. We recommend below that SOC is not mixed from other depths because of the reduction in concentration. However, retaining separate depth increments will inevitably increase measurement cost and require consideration of sampling and measurement cost for each depth increment and the difficulty in reconciling the optimisation for each depth increment. An alternative approach, which we adopted here, is to consider the implications for sampling using only the SOC at 0-10 cm assuming that this layer will provide an over-estimate of the number of samples from other layers (all other things being equal). When measured separately over discrete depth intervals (e.g., 0-10 cm, 10-30 cm, 30-60 cm), SOC stocks decrease approximately exponentially with increasing depth. The variation of SOC over space and time is also different for each depth layer. For example, Singh *et al.* (2013) provided SOC stock variograms for different depths for a single field (Figure 1).
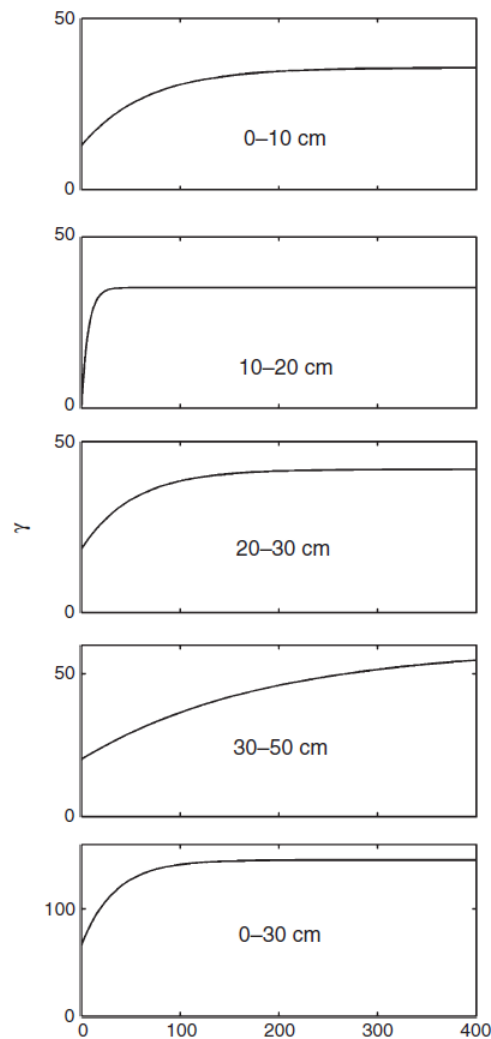
Figure 1 Semivariance of SOC stocks (t SOC ha^{-1})^2 for lag separation distance (m) for different depths of a cropping field in central-western New South Wales on the lower plains of the Macquarie River (Reproduced from Singh et al, 2013; Figure 4)

Singh *et al.* (2013) found that the deepest layer (30-50 cm) had the most variation with the largest spatial structure. Variation appeared to reduce towards the surface but the smallest spatial structure was found in the layer 10-20 cm. Notably, the sill variance is similar for variograms of each layer. In contrast, the sill variance of the aggregated soil depth 0-30 cm is approximately twice that of the individual soil depth layers. This indicates that if the aggregated (0-30 cm) SOC stock variogram is used to design a sampling scheme it will not represent adequately the most variable soil layer and will considerably over-estimate the number of samples necessary to achieve a particular goal. The benefit of a sampling design that reduces the number of samples for the largest depths is outweighed by the need to visit a greater number of locations to collect samples near the surface. Consequently, we recognise that there is little point in producing a sampling design for separate layers because soil is typically collected for all depths at each sampling site. In the work that follows we assume that the surface SOC stock (0-10 cm) is the most variable layer and that a sampling design for this layer will be adequate for collecting soil at depth where SOC is less variable. Inevitably, there will be situations (e.g. Figure 1) where the most variable layer is not 0-10 cm. However, we have little prior information about where in the soil profile SOC stock is most variable.

Monitoring SOC change to verify the effects of land use or management practices

Soil monitoring networks (SMNs) are being considered in many countries (Morvan *et al.*, 2008) to better understand carbon balances in terrestrial ecosystems and to verify the effects of land use or management practices on SOC change. The fundamental requirement of SMNs were defined by Saby *et al.* (2008) as

> "…continuous or repeated observation, measurement and evaluation of soil and / or related environmental or technical data for defined purposes, according to prearranged schedules in space and time, using standardised methods for data collection and analysis".

SOC may remain unchanged, increase or decrease over time in response to land use and management practices. It is essential to assess whether these changes are detectable by soil monitoring taking into account the uncertainties caused by spatial heterogeneity, temporal variation, sampling methods and analytical errors (Saby *et al.*, 2008). Furthermore, differences may also be due to the definition of SOC sequestration. These vagaries must be eradicated to unequivocally demonstrate that SOC sequestration has occurred by ensuring a temporal increase is documented relative to pre-treatment SOC and linked to a net depletion of atmospheric CO_2 (Olson, 2013). SOC sequestration was defined by Olson (2010) as a

> "Process of transferring CO_2 from the atmosphere into the soil…which are stored or retained as part of the soil organic matter…The sequestrated SOC process should increase the net SOC storage during and at the end of a study to above the previous pre-treatment baseline levels and results in a net reduction in the CO_2 levels in the atmosphere."

Olson (2013) suggested that SOC sequestration should consider only C acquired directly from the atmosphere and from inside the monitored CEA. In particular, natural or human-induced erosion and deposition of soil and the associated carbon (E_c) it contains must be excluded otherwise changes to SOC stocks may be attributed falsely to defined land use / management practices.

A useful way of considering the problem is to make explicit the net change in the mean estimate of SOC ($\hat{\bar{z}}$) between two time periods t_1 and t_2 that is being tested. Following Woodward (1992) we define the null hypothesis (H_0) as:

$$H_0: \hat{\bar{z}}(t_1) = \hat{\bar{z}}(t_2),$$

$$H_1: \hat{\bar{z}}(t_1) = \hat{\bar{z}}(t_2) + \theta \ (\theta \neq 0).$$

The alternative hypothesis H_1 is the adjustment due to $\theta = \hat{\bar{d}}_{2,1} + \Delta E_c P$ which between sampling periods t_1 and t_2 is the net results of losses and gains of SOC by all processes ($\hat{\bar{d}}_{2,1}$) in particular soil erosion (E_c) and enrichment / depletion (P) in the intervening time. The contribution of soil erosion and deposition to this process has been largely neglected. However, it is axiomatic that without the measurement of net soil redistribution and its enrichment / depletion of SOC stock, uncertainty will be introduced in one of two cases:

1. net increase in SOC stock may be due (at least partially) to soil deposition (e.g., accumulation of SOC-enriched dust) and performance of the management practice will be over-estimated;
2. net decrease in SOC stock may be due (at least partially) to soil erosion and will reduce SOC storage and cause management practices to appear falsely ineffective i.e., performance of the management practice will be under-estimated.

Common **misconceptions** used to justify and perpetuate the exclusion of soil erosion and deposition by wind and water from C cycling and C accounting models include:

1. soil erosion in Australia is no longer a threat to food production and the environment

2. soil erosion need not be measured because it is accounted for by land use and management practices
3. soil erosion merely redistributes SOC within a field,

Overcoming soil loss remains one of the Australian Research Council National Research Priorities for an environmentally sustainable Australia. Awareness of the impact of soil erosion on soil and land degradation has increased over the last 30-40 years and as a consequence, soil conservation measures have been widely adopted. Recently reductions in soil erosion have been estimated based on field measurements and linked to improvements in land management (Chappell *et al.*, 2012). However, these average regional decreases in soil erosion hide considerable variability indicating that on some farms rates of erosion may not have been reduced.

Soil erosion remains one of the largest causes of soil degradation across the world (Lal, 2001) and Australian agriculture is no exception from it now and in the future when a changed climate may exacerbate the problem. Although soil erosion may be related to land use and management practices it may also occur regardless of these factors and therefore must be accounted for at the monitored CEA.

The third misconception has perhaps delayed the inclusion of soil erosion in carbon cycling models and carbon accounting models (e.g., NCAS). In contrast to this misconception, wind erosion and dust emission readily remove considerable quantities of soil and SOC from Australian fields in agricultural and rangeland regions (Chappell *et al.*, 2012; Webb *et al.*, 2013). The material is transported rapidly vast distances and either deposited in other regions or removed from the terrestrial ecosystem to the ocean. It is not uncommon for water erosion to remove soil on sloping land (Loughran *et al.* 2004) and for the material to be deposited on adjacent fields or transported greater distances by streams and rivers and subsequently deposited on floodplains or in the ocean. Furthermore, large amounts of soil erosion are not needed to make a significant difference to the SOC change over time because eroded material is typically considerably enriched (relative to the soil) with SOC (Webb *et al.*, 2012).

Regardless of the amount of soil erosion and because soil erosion is highly variable in space and time, soil erosion should be monitored or measured retrospectively (cf. Chappell *et al.*, 2012) at the same time that SOC is measured. Otherwise, it will not be possible to unequivocally demonstrate that measured changes in SOC stocks can be attributed to land use and management practices alone. In SOC simulation models, change in SOC stocks over time (*t*) are expressed typically as the difference between inputs of carbon (*I*) and losses due to decomposition of existing stocks with rate modifiers applied to an intrinsic decay constant (*k*)

$$\frac{\partial C}{\partial t} = I - (1 - h) \times abkC \quad , \qquad (1)$$

where *h* is the humification quotient, the degree to which the soil matrix protects SOC from mineralisation, *a* is the temperature rate modifier, often an exponential function of temperature, and *b* is the soil water rate modifier. At present, losses or gains due to soil erosion amount to an unrecognised error term which should be included in Eq 1. A first approximation to soil redistribution was included in the SOC model by adding a zero-order loss / gain term (Sanderman and Chappell, 2013),

$$\frac{\partial C}{\partial t} = I - (1 - h) \times abkC - E_c P \quad . \qquad (2)$$

The mass of redistributed carbon ($E_C > 0$ for net erosion and $E_C < 0$ for net deposition) is considered the fractional change in soil mass times the SOC concentration. However, SOC is concentrated at the surface relative to the bulk of the soil and soil redistribution at the soil surface may produce a relative enrichment (*P*) of SOC in eroded or deposited soil (Gregorich *et al.* 1998). Unless these missing components are measured at the same time as SOC is measured, change in SOC will be due to the combination of land use and management practices and net soil redistribution. Sanderman and Chappell (2013) showed that modest unrecognised amounts of soil redistribution (10–20 t ha^{-1} yr^{-1} from experimental sites around

Australia) produced uncertainties in sequestration rates of similar magnitude to measured sequestration rates (ca. 0.3 to 1.0 t CO_2 ha^{-1} yr^{-1}) for many management practices recommended for building SOC stocks (Sanderman *et al.*, 2011).

This relatively straightforward approach demonstrates how and why soil erosion should be included in carbon cycling and carbon accounting models. However, there is another side to this issue which is less straightforward to tackle but which has the potential to have an even greater impact on SOC sequestration. Erosion of the topsoil changes the composition and structure of the soil at the eroded location. Deposition of soil elsewhere also modifies the soil micro-environment. Consequently, the hydrological characteristics and the temperature etc. will change and alter the SOC decomposition rates. Therefore, excluding from C cycling models these lateral fluxes will ignore change due either to direct or indirect effects.

A straightforward and well-established method to estimate net soil redistribution at a given location has been demonstrated using the ^{137}Cs technique (Kachanoski and de Jong, 1984) and used in south-eastern Australia to show a reduction in soil erosion likely associated with soil conservation measures (Chappell *et al.*, 2012).

> To verify that land use or management practices have changed SOC, we strongly recommend that SOC change by soil redistribution is measured at the same time as SOC content is measured.

Statistical rationale for monitoring SOC change

1.1   Introduction

In Australia, there are two national soil programs which address the components of monitoring but which have very different purposes. The Soil Carbon Research Program (SCaRP) has established the amount of SOC stored in soils under various agricultural practices (Sanderman *et al.* 2011). SCaRP intended to provide an assessment of current SOC stocks and the effect of land use history. The proposed National Soil Condition Monitoring Program (NSCMP) is intended to monitor environmental change through time using repeated sampling of SOC and pH on an approximately 5 year return basis and thus establish the ongoing impact of land use and management practices (Grealish *et al.* 2011). State programs are also being conducted in Tasmania (Soil Condition Evaluation and Monitoring - SCEAM), New South Wales (Monitoring, Evaluation and Reporting Program - MER) and Western Australia (Soil monitoring network for Western Australian wheatbelt). Despite differences in purposes between these programs they are all based on McKenzie *et al.* (2002) and hence are similar in approach, particularly with soil sample collection (25 m square), analysis, data collection and the soil variables measured (Grealish *et al.*, 2011).

Soil monitoring programs face important challenges when attempting to detect significant change and identify the driving forces responsible for change (Goidts *et al.*, 2009). These challenges are caused primarily by the apparent paradox of minimising sample area to maximise the likelihood of detecting temporal change whilst maximising sample area to minimise spatial variance. In particular for SOC, its typically large spatial and small temporal variability relative to its content may prevent the detection of change. For example, repeated sampling over time of a single 25 m square area (plot), common to Australian soil monitoring programs, will ensure difference is detected at that plot, but it is unlikely to ensure that change is attributed to land use and management practice conducted across the field (Chappell and Viscarra Rossel, 2013). This is because SOC change detected for a small portion of the field may not represent adequately the SOC change which has occurred across the majority of the field. Many soil samples may be obtained from across the field but SOC measurement costs of each sample make it expensive to ensure that they represent adequately the area of interest (CEA). A straightforward solution of sampling to represent adequately the SOC across the CEA, is to create a composite (or bulk) of a given number of soil samples hereafter called aliquots (*n*).

Composite sampling in this situation is the aggregation and mixing of soil aliquots (samples of soil from the field) for subsequent measurement of a composite sample representative of SOC within the CEA. Since only the composite soil is measured the number and hence cost of analyses is reduced. This approach assumes, following de Gruijter *et al.* (2006), that (i) the property of interest, in this case SOC, is directly measured in the aliquots or is defined as a linear transformation of the measured property; (ii) after mixing no physical, chemical or biological interactions between the aliquots take place that influence the measurement. Although compositing reduces laboratory costs, it may introduce two interrelated sources of variance: that due to imperfect mixing of the sampled aliquots and that due to subsampling the mixed composite. However, the potential increase in variance could be counteracted by analysing multiple replicates from the composite (Brus *et al.*, 1999).

1.2   Spatial sampling

Simple random sampling may be adopted across a CEA prior to creating a composite. This is done by determining the minimum and maximum coordinates of the area, generating two pseudo-random coordinates from a uniform distribution between those extremes and then accept a sample location if both values that form the coordinate fall within the area. The spatial mean of the CEA ($\bar{z}$) for a variable (z) is estimated by:

$$\hat{\bar{z}}_{Si} = \frac{1}{n}\sum_{i=1}^{n} z_i, \tag{3}$$

where n is the sample size and z_i is the value at sampling location i and Si is used to denote simple random sampling designs. This estimator is p-unbiased meaning that repeated sampling, measurement and calculation would find on average the true value for $\bar{z}$. The unbiased condition remains if the errors are purely random, i.e., zero on average. The **sampling variance** of the estimated mean $V(\hat{\bar{z}})$ is estimated by:

$$\hat{V}(\hat{\bar{z}}_{Si}) = \frac{1}{n(n-1)}\sum_{i=1}^{n}(z_i - \hat{\bar{z}}_{Si})^2, \tag{4}$$

and the standard error is estimated by $\sqrt{\hat{V}(\hat{\bar{z}}_{Si})}$.

A single composite does not itself convey information about variability in that compositing approach. The approach should be repeated a number of times to provide sampling variance. If m composites are formed each from a simple random sample of size n, the spatial mean is estimated by:

$$\hat{\bar{z}}_c = \frac{1}{m}\sum_{i=1}^{m} z_{cj}, \tag{5}$$

where z_{cj} is the value measured on composite j and the subscript c indicates that this estimator applies to composite random samples. The sampling variance of $\hat{\bar{z}}_c$ is estimated by:

$$\hat{V}(\hat{\bar{z}}_c) = \frac{1}{m(m-1)}\sum_{j=1}^{m}(z_j - \hat{\bar{z}}_c)^2, \tag{6}$$

and the standard error is estimated by $\sqrt{\hat{V}(\hat{\bar{z}}_c)}$.

The **spatial variance** of z between locations in the area $S^2(z)$ is estimated by:

$$\widehat{S^2}(z) = \frac{1}{(n-1)}\sum_{i=1}^{n}(z_i - \hat{\bar{z}}_{Si})^2, \tag{7}$$

The variance between locations in the area can be estimated in the case of m composite simple random samples by:

$$\widehat{S^2}(z) = \frac{n}{(m-1)}\sum_{j=1}^{m}(z_j - \hat{\bar{z}}_c)^2, \tag{8}$$

With simple random sampling the sampling variance is usually larger than with most other types of design at the same costs for two possible reasons (Brus *et al.*, 1999; de Gruijter *et al.*, 2006):

1. spatial coverage by the sample may be poor,
2. visiting sampling locations that are irregularly distributed may be more time-consuming for logistical reasons and higher per-sample costs results in a smaller sample size.

With stratified simple random sampling the CEA may be divided into equal area strata and simple random sampling is applied to locate soil collection locations within each stratum. The strata should be approximately uniform in area to maximise the benefits. The main reason for stratification is to improve sampling efficiency compared to simple random sampling. However, inappropriate stratification or sub-optimal allocation of sample size could reduce efficiency. de Gruijter *et al.* (2006) suggested that this could occur if the stratum means differ little or if the sample sizes are strongly disproportional to the surface areas of the strata.

The strata may be defined in terms of one or more ancillary variables that are known everywhere in the CEA and that are known or expected to be correlated with the SOC. The strata can use a pre-existing stratification or be defined based on *a priori* classification e.g., soil type map based on the relationships

between the ancillary variables and SOC or if the ancillary variables are quantitative such as obtained by remote/proximal sensing (yield), then the strata can be calculated by classification. If no suitable ancillary variables are available for stratification, one may consider stratification on the basis of spatial coordinates. Optimal efficiency is achieved by making use of available prior information (e.g. stratification using ancillary variables or compact geographical stratification). This leads to a decision-making process which enables users to maximise their sampling efficiency (Figure 2).
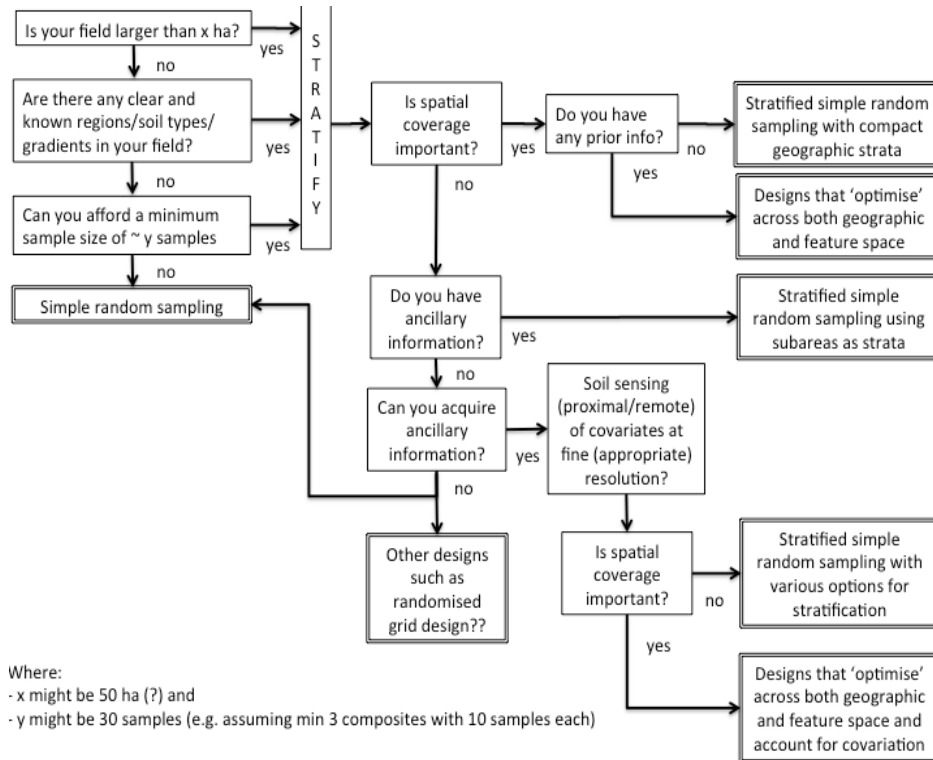
Figure 2 Decision tree for maximising sampling efficiency based on prior information

An outline of the methods of stratification by ancillary variables and compact geographical stratification can be found in de Gruijter *et al.* (2006). The latter approach is particularly useful where the CEA is irregularly shaped and it may be difficult to determine strata with equal area. Walvoort *et al.* (2010) provide software in R (R Development Core Team, 2010) to perform the compact geographical stratification.

A regular grid may be used to stratify the CEA prior to sampling (Figure 3). In this case, the strata have equal area and therefore equal volumes of soil material can be collected at the locations. A general approach is to determine the stratification and then composite the samples weighted by the area in which samples are bulked. In practice, stratification is formed from strata that are as homogeneous as possible. Two or preferably more aliquots may be placed within each stratum and these aliquots can form composites with their counterparts in the remaining strata (Figure 3). Once the composites are formed, well mixed and sub-sampled they can be measured and used to estimate the variance of the composite samples. The mean of the CEA for each composite is estimated by:

$$\hat{\bar{z}}_{St} = \sum_{h=1}^{H} a_h \hat{\bar{z}}_h, \tag{9}$$
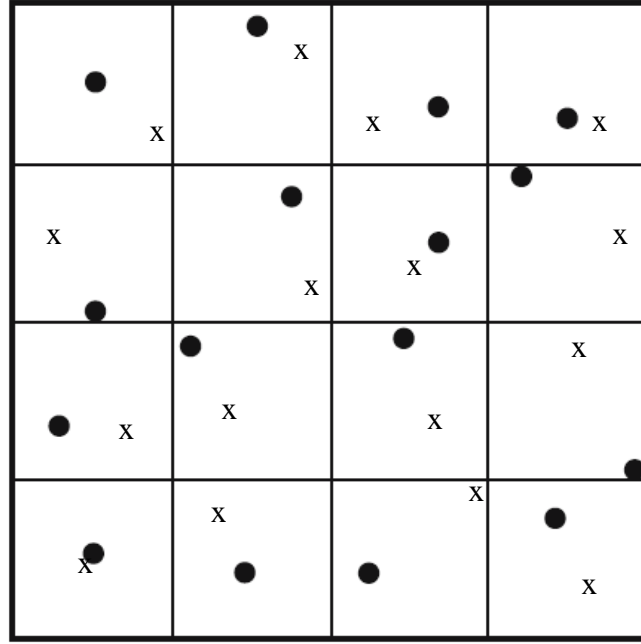
Figure 3 Example of a simple random sample stratified using a regular grid. The strata all have the same area and in each stratum two locations are selected by simple random sampling. Two composite samples can be formed by grouping all samples selected in the first draw (solid circles) and those selected in the second draw (crosses) (Modified from de Gruijter _et al._, 2006)

where H is the number of strata, a_h is the relative area of stratum h, $\hat{\bar{z}}_h$ is the sample mean and St indicates the stratified simple random sampling design. The **sampling variance** of $\hat{\bar{z}}_{St}$ can be estimated by

$$\hat{V}(\hat{\bar{z}}_{St}) = \sum_{h=1}^{H} a_h^2 \hat{V}(\hat{\bar{z}}_h), \tag{10}$$

where $\hat{V}(\hat{\bar{z}}_h)$ is the estimated variance of $\hat{\bar{z}}_h$ and is given by

$$\hat{V}(\hat{\bar{z}}_h) = \frac{1}{n_h(n_h-1)} \sum_{i=1}^{n_h} (z_{hi} - \hat{\bar{z}}_h)^2, \tag{11}$$

and n_h is the sample size in stratum h. The standard error of the estimated mean is estimated by $\sqrt{\hat{V}(\hat{\bar{z}}_{St})}$.
An unbiased estimator of the **spatial variance** $S^2(z)$ is:

$$\widehat{S^2} = \widehat{\overline{z^2}}_{St} - (\hat{\bar{z}}_{St})^2 + \hat{V}(\hat{\bar{z}}_{St}), \tag{12}$$

where $\widehat{\overline{z^2}}_{St}$ denotes the estimated mean of the target variable squared (z^2), obtained in the same way as $\hat{\bar{z}}_{St}$ (Eq. 9) but using squared values (de Gruijter _et al._, 2006).

1.3   Space-time sampling

The theory for sampling in time is similar to that for sampling in space, although the practicalities will differ. Hence, we do not describe monitoring here and continue with a summary of the theory for space-time

sampling. The reader is referred to the book by de Gruijter *et al.* (2006) for the theory and the practicalities of monitoring.

The efficiency of a sampling pattern for monitoring is partly determined by the distribution of the sampling units in space-time. An important aspect of the sampling pattern for monitoring is whether at all sampling times the same locations are revisited or whether this restriction is relaxed and a proportion of the sampling locations are replaced by new locations. De Gruijter *et al.* (2006) described four basic types of sampling pattern: static, synchronous, static-synchronous and rotational patterns (Figure 4).
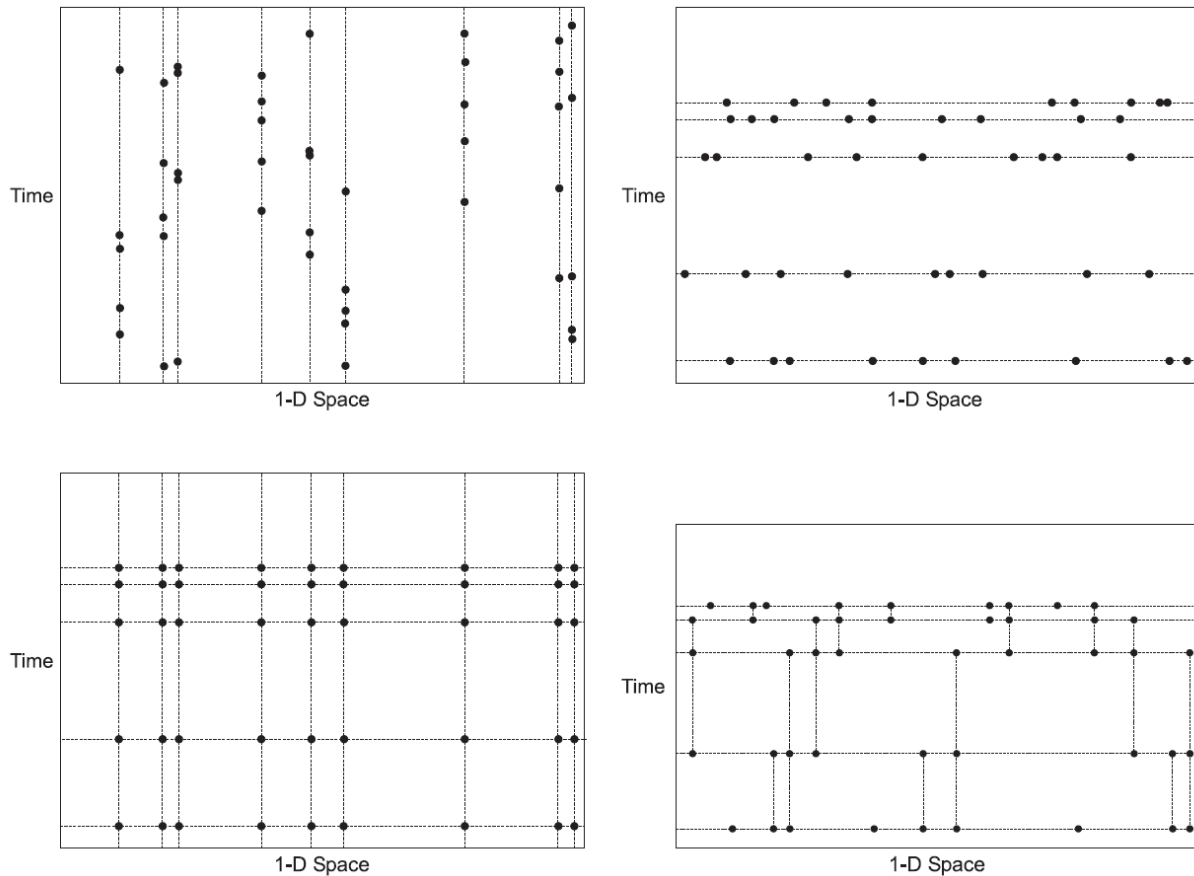
Figure 4 Four basic types of sampling pattern for monitoring: static (upper-left), synchronous (upper-right), static-synchronous (lower-left) and rotational (lower-right). Each vertical line corresponds to repeated samples in space and each horizontal line corresponds to repeated samples in time. (Reproduced from Brus and Noij, 2008; Figure 3)

Static sampling occurs at a fixed set of locations and may or may not follow the same pattern in time. Synchronous sampling refers to a different set of samples for each sampling time which may or may not have the same spatial patterns. If they are the same and they coincide spatially then the pattern is static-synchronous. This is essentially a combination of static and synchronous sampling patterns (also known as a pure panel).

De Gruijter *et al.* (2006) summarise the options available for the estimation of the change in the spatial mean (Table 15.1 of their book) and suggest that synchronous and rotational patterns are suitable and a static-synchronous pattern is highly suited to the detection of change in the spatial mean. In a static-synchronous pattern the mean difference of estimated means $\hat{\bar{z}}(t_1)$ and $\hat{\bar{z}}(t_2)$ between events t_1 and t_2, $\hat{\bar{d}}_{2,1}$ is estimated by

$$\hat{\bar{d}}_{2,1} = \hat{\bar{z}}(t_2) - \hat{\bar{z}}(t_1). \tag{13}$$

The locations are fixed and all of the locations are revisited which implies that in estimating the sampling variance of the change, a possible temporal correlation between the estimated means $\hat{\bar{z}}(t_1)$ and $\hat{\bar{z}}(t_2)$ must be taken into account. The true sampling variance equals

$$V(\hat{\bar{d}}_{2,1}) = V(\hat{\bar{z}}(t_2)) + V(\hat{\bar{z}}(t_1)) - 2\rho(\hat{\bar{z}}(t_2), \hat{\bar{z}}(t_2)). \tag{14}$$

where ρ is the temporal correlation between the two estimated means. As it increases the sampling variance of change gets smaller. The correlation can be maximised when the sampling locations between events coincide. With destructive soil sampling, distances from the original locations should be kept as small as is practically possible given the original disturbance at the location. Lark (2009) notes that this technically adds another term to the variance of the mean difference. This locational error is not included here.

A synchronous design (as opposed to a static synchronous design) may also be used to estimate the change in the spatial mean of SOC. This approach assumes that the samples taken at different times are mutually independent and the estimated means between t sampling events $\hat{\bar{z}}(t_1)$ and $\hat{\bar{z}}(t_2)$ are uncorrelated. The sampling variance of $\hat{\bar{d}}_{2,1}$ equals

$$V\left(\hat{\bar{d}}_{2,1}\right) = V(\hat{\bar{z}}(t_2)) + V(\hat{\bar{z}}(t_1)), \tag{15}$$

which can be simply estimated by:

$$\hat{V}(\hat{\bar{d}}_{2,1}) = \hat{V}(\hat{\bar{z}}(t_2)) + \hat{V}(\hat{\bar{z}}(t_1)). \tag{16}$$

Since there is no covariance-term, synchronous designs are in general less efficient than static-synchronous designs (de Gruijter *et al.*, 2006). Bishop and Pringle (2012) advocated a synchronous design to estimate the change in the spatial mean of SOC. They expected their results to be less precise estimates of change and hence a conservative estimate of the optimal number of samples to detect change (Lark, 2009). Their main concern in the adoption of this pattern was to reduce the opportunity to cheat the carbon accounting system using prior knowledge of the locations in the field to be revisited (static-synchronous pattern) for example by adding organic inputs to the known sampling sites. This approach may also avoid sampling soil disturbed by previous sampling and the potential for a changed soil micro-environment. However, following this logic any part of the field effected by soil erosion and deposition may also have a changed soil micro-environment.

The sampling variance of the difference $V(\hat{\bar{d}}_{2,1})$ can be substituted for the spatial variance of the target variable and used to estimate e.g., the required sample size. However, the ability to adapt the static-synchronous monitoring design is restricted by the initially fixed sampling locations. Only the time period to revisit those sites could be changed in response to the first round of sampling. Otherwise, a synchronous (or rotational) pattern would be formed. The implications are that if the first round of sampling has performed as well as, or better than, expected the static-synchronous pattern should be retained because it provides better precision. If the first round of sampling has performed worse, or in any case if there is a need to increase the number of samples in the second round, then a synchronous (or rotational) pattern can be adopted.

The sampling design can be adapted over time if the measured SOC variance differs from the expected SOC variance. Three scenarios and adaptive approaches are described below (Figure 5). In Scenario 1, the expected SOC variance is found to be larger than the measured SOC variance. In this situation it would probably not be worthwhile reducing samples because this may reduce the ability to detect change. For example, removing samples may result in not returning to some strata and thereby poorly representing the variation across the CEA. Despite the expected SOC variance being larger than the measured SOC variance, we suggest that the first round sampling design be retained and used in the second round to maintain the ability to detect SOC change.

In the second scenario the variance after the first round is larger than expected based on the decisions used to allocate strata and composites. We decide to adapt the sampling design to reduce the variance and can either include additional strata or as shown in Figure 5 include another composite. This means that we should return to the original sample locations of round one and obtain an additional sample in each strata (shown in red) which is subsequently used to form an additional composite.

Scenario 1: Expected SOC variance ≥ measured SOC variance after first sampling round
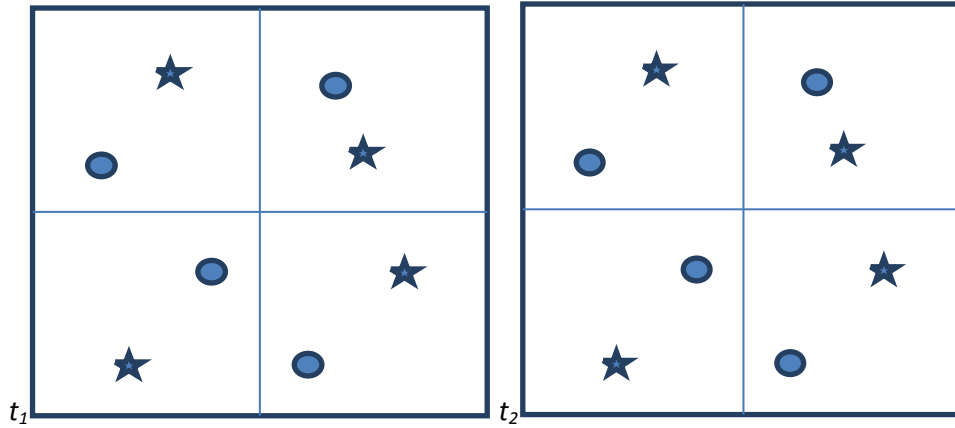
Scenario 2: Expected SOC variance < measured SOC variance after first sampling round
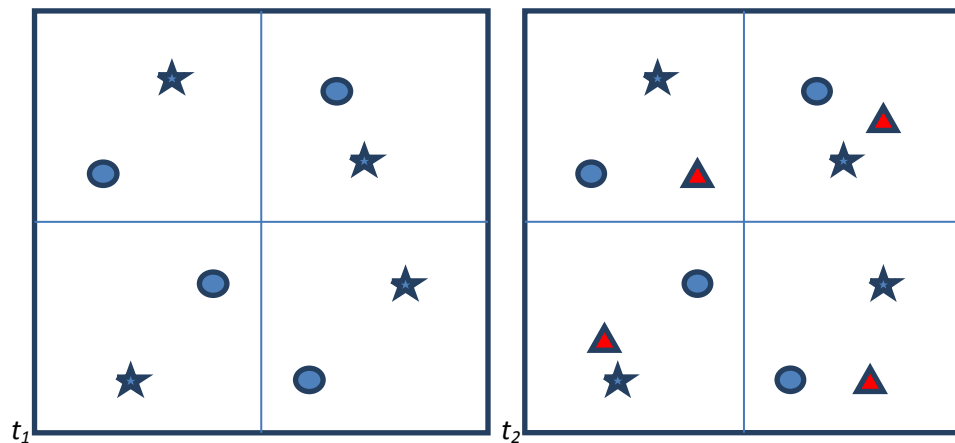
Scenario 3: Expected SOC variance << measured SOC variance after first sampling round
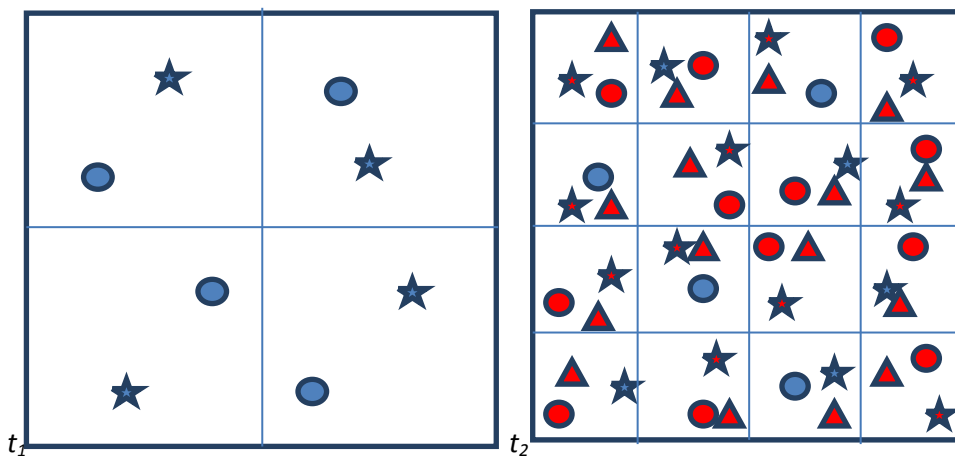
Figure 5 Sampling design and its adaptation (t_2) based on the difference between expected SOC variance and measured SOC variance after the first round of sampling (t_1)

In the third scenario the variance after the first round is much larger than expected. To achieve our target minimum detectable change we must adapt the sampling design to reduce the variance and should probably increase the number of strata and the number of composites. We should revisit the existing sample locations and ensure that the additional samples are included in their respective strata (shown in red).

These adaptation strategies should in general be avoided because they require considerable effort to merge within an existing strategy. Instead of adapting a strategy it is preferable to be conservative at the outset and use as many strata and composites as can be afforded. Furthermore, it is preferable to evaluate the performance of this conservative sampling design after the second round of sampling has been finalised and the SOC variance determined; this is the first time that variation in time and space will have been sampled.

> To reduce measurement costs and maximise the precision of the sampling variance for monitoring SOC, we recommend:
>
> 1. For the initial sampling round stratified simple random sampling with locations initially randomly allocated within each stratum and soil bulked across all strata to form multiple composite samples where each composite sample contains an aliquot of soil taken from each stratum.
> 2. If the first round sampling variance is adequate, plan to revisit the same locations (static-synchronous design). If the first round sampling variance is inadequate adopt a synchronous (or rotational) design and either increase the number of samples (strata)

Quality measure and constraints for monitoring SOC change

1.4 Design-based optimisation of sample sizes

At the outset of a sampling design, sample size may be optimised by minimising the variance for a given maximum allowable cost or to minimise the cost for a given maximum allowable variance. To select the optimal sample size of n and composites H a simple cost (C) model was introduced by Brus and Noij (2008):

$$C = C_0 + C_{fld} + C_{lab}, \qquad (17)$$

$$C_{fld} = 2 \times H \times n \times t \times c_{fld},$$

$$C_{lab} = 2 \times H \times c_{lab},$$

where C_0 are the fixed costs which are independent of the number of composites (H) and aliquots (n) i.e., costs of preparing the fieldwork, travel costs, equipment costs, etc., t is the time in hours needed to sample one aliquot, c_{fld} are the costs of fieldwork per hour and c_{lab} are the costs of the laboratory measurement per composite. The optimal combination of n and H is independent of C_0, so this may be separated from the processing. The costs C_{fld} and C_{lab} are doubled for the situation of two sampling events over time. The C_{lab} cost is that associated with the receipt, preparation and measurement of soil organic carbon concentration, soil bulk density and the calculation of SOC.

In the absence of definitive information for these parameters, we estimated costs as follows: C_0=$2000, t=0.33 hours to sample one aliquot, c_{fld}=$120/hour and c_{lab}=$30/composite. Figure 6 shows the total cost (C) for a range of sample composites from different numbers of strata. A disproportionate increase in cost with number of composites samples is evident.

In this initial sampling situation estimates for the costs of sampling requires prior estimates of the standard deviations in the area or the strata / composites etc. For example, in the case of simple random sampling the sample size needed to estimate a mean such that, with a specified large probability (1-α), the relative error $|(\hat{\bar{z}} - \bar{z})/\bar{z}|$ is smaller than a particular limit r, can be calculated (following de Gruijter et al., 2006 p. 84) by:

$$n = \left(\frac{u_{1-\alpha/2} . \check{S}(z)}{r \check{\bar{z}}} \right), \qquad (18)$$

where $u_{1-\alpha/2}$ is the $(1 - \alpha/2)$ quantile of the standard normal distribution, $\check{S}(z)$ is a prior estimate of the standard deviation of z in the area and $\check{\bar{z}}$ is a prior estimate of the mean. In the equation, $\check{S}(z)/\check{\bar{z}}$ is a prior estimate of the coefficient of variation in the area which could be obtained from previous sampling in the same area. In the case of stratified simple random sampling de Gruijter et al., (2006; p.94-95) demonstrate that the total sample size needed to keep the variance below a maximum value V_{max}, assuming that the cost per location is equal for the strata is:

$$n = \frac{1}{V_{max}} \left(\sum_{h=1}^{H} a_h \check{S}_h \right)^2. \qquad (19)$$
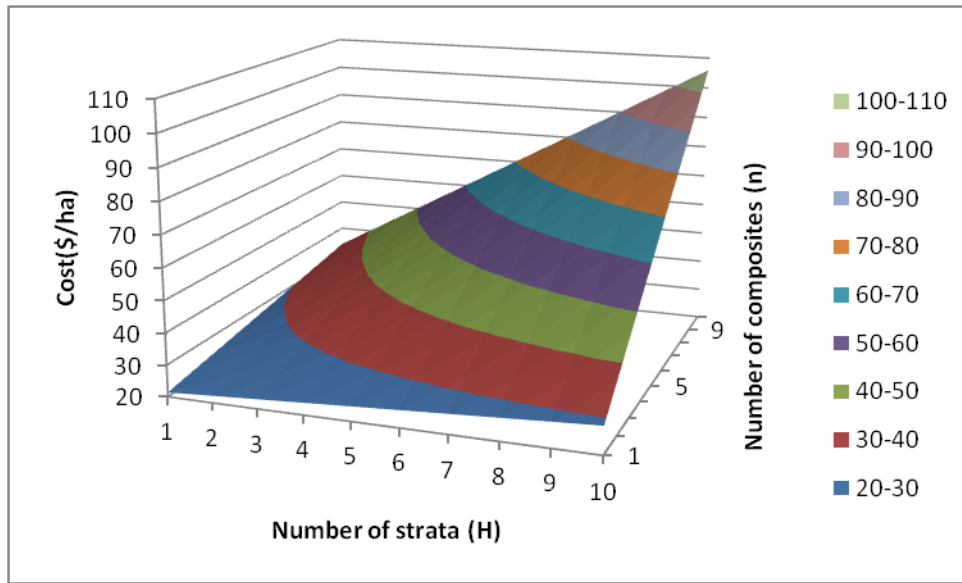
Figure 6 The total cost (including fixed overhead cost C_0=$2000, fieldwork cost c_{fld}=$120/hour sampling time t=0.33 hour and laboratory cost c_{lab}=$30/composite) associated with a sampling design involving stratification and composited soil for SOC measurement

1.5 Model-based optimisation of sample sizes

If prior information on the spatial variability is available in the form of a variogram, Domburg *et al.* (1994) demonstrated how it may be used to estimate the sampling variance of the sample mean of an area. De Gruijter *et al.* (2006) describe the special prediction equations that apply to basic sampling strategies including simple random sampling:

$$E_\xi\{V_p(\hat{\bar{z}}_{SI})\} = \frac{1}{n}\bar{\gamma},$$ (20)

and stratified simple random sampling:

$$E_\xi\{V_p(\hat{\bar{z}}_{St})\} = \sum_{h=1}^{H}\frac{\bar{\gamma}_h}{n_h},$$ (21)

where $\bar{\gamma}$ and $\bar{\gamma}_h$ are the mean semi-variance between two random points in an area or in stratum h, respectively; $E_\xi(.)$ is the statistical expectation over realisations from the model ξ underlying the chosen variogram, V_p is the variance over realisations from the design p (the usual sampling variance in the design-based approach). The $\bar{\gamma}$ is calculated by numerical integration or by Monte Carlo simulation, repeatedly selecting a pair of random locations, calculating its semi-variance and averaging (e.g., Bishop and Pringle, 2012). Notably, in the case where strata have equal area and are as compact as possible, Brus & Noij (2008) suggest that the mean semi-variances within strata are approximately equal for the strata. They approximated the mean semi-variance using the mean semi-variance within a square of the same area.

1.6 Hypothesis testing

The baseline condition is referred to as the null-hypothesis H_0 and is important to consider carefully because different decisions may occur due to the baseline condition for the same dataset. In our case, a farmer or landowner initiates the investigation of SOC and must show that the target quantity $\hat{\bar{d}}_{2,1}$, the

mean difference in estimated means of SOC, is greater than zero and statistically significant. De Guijter *et al.*, (2006 Ch2) provide useful discussion on hypothesis testing to account for uncertainties in estimation. Woodward (1992) also provides a useful way of considering the problem:

$$H_0: \hat{\bar{z}}(t_1) = \hat{\bar{z}}(t_2),$$

$$H_1: \hat{\bar{z}}(t_1) = \hat{\bar{z}}(t_2) + \theta \ (\theta \neq 0),$$

The alternative hypothesis H_1 is the adjustment due to $\theta = \hat{\bar{d}}_{2,1} + \Delta E_c P$ which between sampling periods t_1 and t_2 is the net results of losses and gains of SOC by all processes in particular soil erosion (Ec) and enrichment / depletion (P) in the intervening time. The uncertainty due to reaching an incorrect conclusion is the minimum detectable change (MDC) which is related to the probability of the errors on the conclusion. In general, the smaller the MDC, the larger the required sample size for a given probability of false acceptance error (de Gruijter *et al.*, 2006).

Our $H_0: \hat{\bar{d}}_{2,1} + \Delta E_c P \leq 0$ that is the average difference in SOC has stayed the same or may have decreased over time (adverse effect). The alternative hypothesis $H_1: \hat{\bar{d}}_{2,1} + \Delta E_c P > 0$ is that the average difference in SOC has increased over time and is considered a 'positive' effect. In statistical hypothesis testing two types of errors may be made. We may reject H_0 and conclude that there is a positive effect when in reality there is no effect or even an adverse effect (false rejection; type-I error). We assign a probability denoted α to this type of error and decide on an acceptable value depending on the implications of making a false rejection i.e., if consequences of a false rejection are serious then we might select e.g., 1%. Alternatively, we may accept H_0 and conclude that there is no effect or an adverse effect, when in reality there is a positive effect (false acceptance; type-II error, β). The probability $1 - \beta$ is referred to as the power of the test and is used as a quality measure. First the critical value is calculated for the mean beyond which H_0 is rejected. The power is the probability that one correctly concludes that there is a positive effect, that $\hat{\bar{d}}_{2,1} > 0$. The power of the test depends on $\hat{\bar{d}}_{2,1}$ i.e., the greater $\hat{\bar{d}}_{2,1}$, the larger the power.

Based on Woodward (1992), the one-tailed test (for change with direction) statistic is commonly based on the t-test:

$$(X_{1-\alpha} + X_{1-\beta})^2 = \frac{\hat{\bar{d}}_{2,1}^{\ 2} + \Delta E_c P}{\frac{V(\hat{\bar{z}}(t_1))}{N_1} + \frac{V(\hat{\bar{z}}(t_2))}{N_2}} . \tag{22}$$

where X is a standard normal distribution. To find a sample size we might assume that $N_1 = N_2 = N$ and specify the difference between the means attributable to management practice and the loss/gain due to SOC erosion / deposition, $\hat{\bar{d}}_{2,1} + E_c P$, to detect with a power $1 - \beta$ given that a significance test of size α is used. We must also have a reasonable estimate of the variance $\hat{V}$

$$N = \frac{(X_{1-\alpha} + X_{1-\beta})^2 \hat{V}(\hat{\bar{d}}_{2,1})}{(\hat{\bar{d}}_{2,1} + \Delta E_c P)^2}, \tag{23}$$

or might assume that it remains constant over time:

$$N = \frac{(X_{1-\alpha} + X_{1-\beta})^2 2\hat{V}}{(\hat{\bar{d}}_{2,1} + \Delta E_c P)^2}, \tag{24}$$

If there are unequal sample sizes we might assume that $N_2 = r N_1$ where r (their ratio) is already known:

$$N_2 = \frac{(X_{1-\alpha} + X_{1-\beta})^2 \left[\frac{\hat{V}(\hat{\bar{z}}(t_2))}{r} + \hat{V}(\hat{\bar{z}}(t_1)) \right]}{(\hat{\bar{d}}_{2,1} + \Delta E_c P)^2} . \tag{25}$$

In general there is a reduction in benefit when $r > 3$. The formula (Eq. 22) may be inverted to give an expression for $\hat{\bar{d}}_{2,1}$, that is the difference between means which it is possible to detect with the specified power (and size) of test or more usefully, the smallest difference detectable with at least the given power. In the case of a one-sided test, where $\hat{V}$ is assumed constant over time, this gives:

$$\hat{\bar{d}}_{2,1} = \left(X_{1-\alpha} + X_{1-\beta}\right)\left(\frac{2\hat{V}}{N}\right)^{0.5} - \Delta E_c P, \tag{26}$$

and which when normalised by dividing by $\sqrt{\hat{V}}$ provides the minimum detectable effect size:

$$\Delta = \frac{\hat{\bar{d}}_{2,1} + \Delta E_c P}{\sqrt{\hat{V}}} = \left(X_{1-\alpha} + X_{1-\beta}\right)\left(\frac{2}{N}\right)^{0.5}, \tag{27}$$

Similarly, a value for the power β, given N and $\hat{\bar{d}}_{2,1}$ is obtained from:

$$X_{1-\beta} = \frac{\hat{\bar{d}}_{2,1} + \Delta E_c P}{\sqrt{\hat{V}}}\left(\frac{N}{2}\right)^{0.5} - X_{1-\alpha}. \tag{28}$$

To illustrate the application of these equations we create a relevant and plausible scenario for the measurement of SOC across a field at two separate times. We assume that simple random sampling (not stratified) obtained from the field, the same number of samples from t_1 as from t_2. The SOC is to be compared using a one-sided 5% significance test (i.e., α=0.05). We wish to be 95% sure of detecting an average difference between t_1 and t_2 of 3 t SOC ha with SOC redistribution (negative value = erosion) of -1 t SOC ha^{-1} (i.e., β=0.05, $\hat{\bar{d}}_{2,1} = 3$ and $E_c P$=-1). The sample variance is expected to be approximately constant over time at 20 (t SOC ha^{-1})^2($\hat{V} = 20$). Using equation 24, we can estimate the required sample size at t_2 and the value in brackets is the outcome of neglecting SOC erosion ($\Delta E_c P = 0$):

$$N_1 = N_2 = \frac{(1.734 + 1.734)^2[2(20)]}{(3-1)^2} = 60.14 \ (26.73),$$

and the total would be $N=N_1+N_2$=120.27 (53.45). Note that careful consideration of the rounding is needed when implementing these sampling designs.

Assuming we can afford to obtain from the field only N=54 samples from both periods t_1 and t_2 (r=1 i.e., $N_1=N_2$=27). The SOC is to be compared using a one-sided 5% significance test (i.e., α=0.05). We wish to be 95% sure of detecting a difference in mean SOC between t_1 and t_2 (i.e., β=0.05). The sample variance is expected to be approximately constant over time at 20 (t SOC ha^{-1})^2($\hat{V} = 20$). Using equation 26, the minimum detectable difference (MDC) is below and the value in brackets is the outcome neglecting SOC erosion ($E_c P$=0):

$$\hat{\bar{d}}_{2,1} = (1.734 + 1.734)\left(\frac{2(20)}{54}\right)^{0.5} + 1 = 3.98 \text{ t SOC ha}^{-1} \ (2.98 \text{ t SOC ha}^{-1})$$

How confident can we be of detecting an average change between t_1 and t_2 of 4 t SOC ha^{-1} ($\hat{\bar{d}}_{2,1} = 4$, $E_c P$=-1) when using N=54 samples from each period (r=1 i.e., $N_1=N_2$=27)? The SOC is to be compared using a one-sided 5% significance test (i.e., α=0.05). The sample variance is expected to be approximately constant over time at 20 (t SOC ha^{-1})^2($\hat{V} = 20$). Using equation 28, the power of the test is below and the value in brackets is that is the outcome neglecting SOC erosion ($E_c P$=0)

$$X_{1-\beta} = \frac{4-1}{\sqrt{20}}\left(\frac{54}{3}\right)^{0.5} - 1.734 = 1.752 \text{ or } 1\text{-}0.05\text{=}95\% \ (2.9134 \text{ or } \approx 1\text{-}0.009\text{=}99.1\%),$$

Since we know that these values change with the power and size of the test, their variation can be considered by plotting MDC against required sample size for a range of power using equation 27 (7). For example, the normalised MDC requires additional samples as we increase our confidence from 80% to 99% (from left to right). Similarly, for a given sample size the MDC increases across the same range of confidence. However, if we neglect the SOC redistribution as described in the applications above there is small decrease in the total number of samples and MDC required.

(a)

(b)
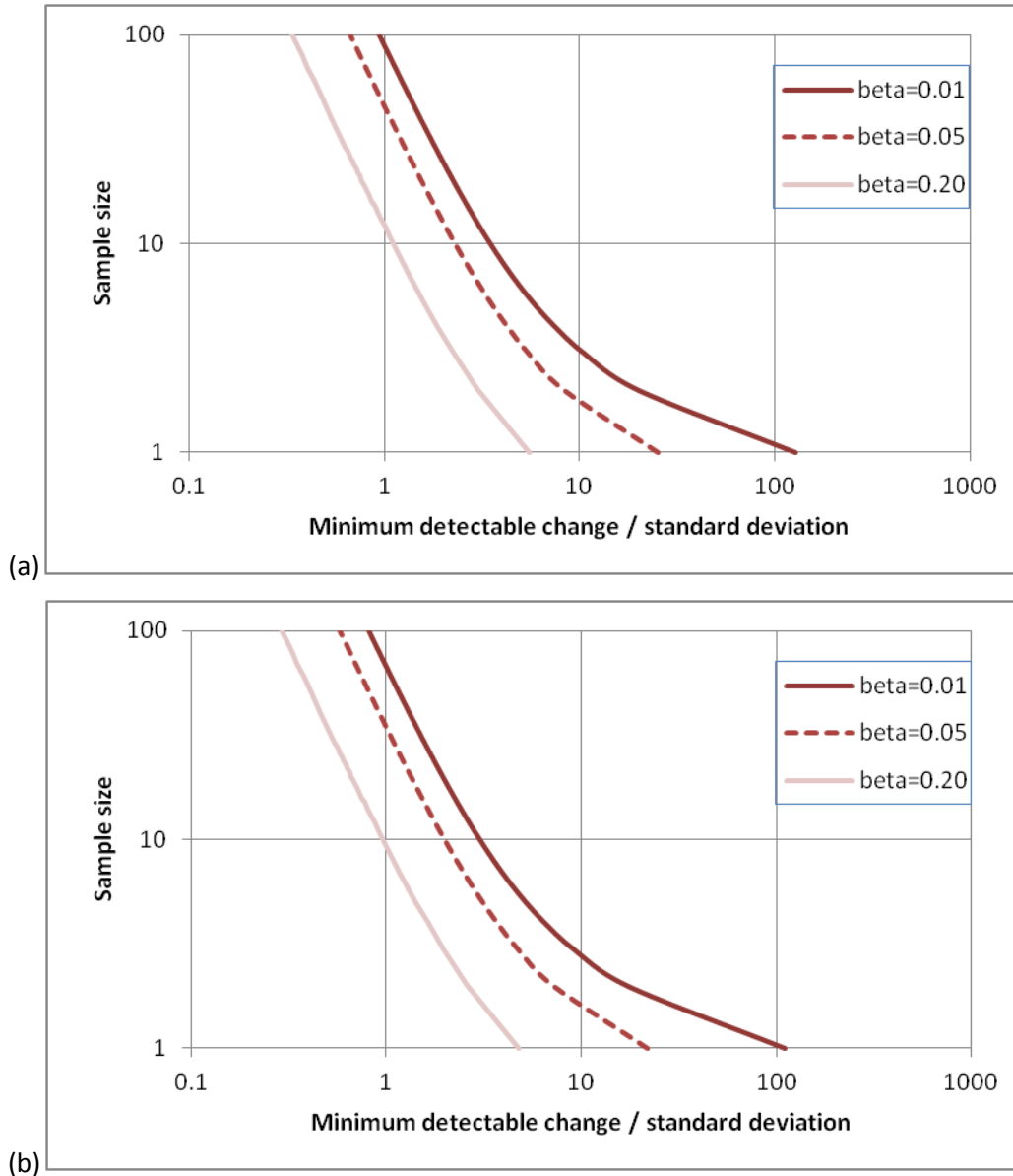
Figure 7 Required sample size against normalized minimum detectable change with β and $\alpha = 0.05$ for (a) SOC change (4 t SOC ha[-1]) and SOC erosion (1 t SOC ha[-1]) and (b) only SOC change. Both axes are plotted using a log-scale.

The application of the above equations using relevant and plausible scenarios for the measurement of SOC demonstrates that neglecting SOC erosion provides estimates which are overly optimistic. This outcome is reasonable since it implies that all of the SOC change over time is inappropriately attributed to management practice.

Statistical model for monitoring SOC change

We assume for simplicity that stratification is formed using a regular grid placed over the CEA. In this situation, the grid readily forms cells or strata that are equal in area and consequently define the number of samples (*n*) used to form the composite or aliquot. Brus and Noij (2008) assumed equal sampling variances in the first and second sampling round:

$$V(\hat{\bar{d}}_{2,1} + \Delta E_c P) = (2 - 2\rho)V(\hat{\bar{z}}(t_1)) = (2 - 2\rho_k)V(\hat{\bar{z}}(t_2)), \qquad (29)$$

and used a temporal correlation coefficient ρ_k at lag k to moderate the variance. They showed how the sampling variance may be estimated from stratified simple random sampling using composites:

$$V_{St}(\hat{\bar{z}}) = \sum_{h=1}^{H} w_h^2 \frac{\bar{\gamma}_h}{n_h} = \frac{1}{H^2} \sum_{h=1}^{H} w_h^2 \frac{\bar{\gamma}_h}{n_h}, \qquad (30)$$

where *H* is the number of strata (the number of aliquots per composite), w_h is the relative area of stratum *h*, $\bar{\gamma}_h$ is the mean semi-variance within stratum *h* and n_h is the number of sample points per stratum (number of composites). Since the strata have equal area and in this case are assumed to be as compact as possible, Brus and Noij (2008) suggest that the mean semi-variances within strata are approximately equal for the strata. They approximated the mean semi-variance using the mean semi-variance within a square of the same area to provide

$$V_{St}(\hat{\bar{d}}_{2,1} + \Delta E_c P) = (2 - 2\rho_k)\frac{1}{H n_h} \bar{\gamma}_h, \qquad (31)$$

In addition to sampling variance, the total variance $V_{tot}(\hat{\bar{d}}_{2,1})$ includes the measurement variance ($V_{lab}(\hat{\bar{d}}_{2,1})$) or error):

$$V_{tot}(\hat{\bar{d}}_{2,1}) = V_{St}(\hat{\bar{d}}_{2,1} + \Delta E_c P) + V_{lab}(\hat{\bar{d}}_{2,1} + \Delta E_c P). \qquad (32)$$

A one-sided 5% significance test (i.e., α=0.05) and a confidence of 95% in detecting a difference in mean soil organic carbon (SOC) between t_1 and t_2 (i.e., β=0.05) with a variance of 20 (t SOC ha^{-1})^2 is associated with a sampling design involving stratification and composited soil (Figure 8). The results show that only when there are very few strata and / or few composites of soil samples does the minimum detectable change increase appreciably. Since we know that composite samples are expensive, these results indicate that there should be at least as many strata as there are composites that may be afforded.

Evidently, with increasing number of composites there is a diminishing rate of increase in fieldwork cost and an approximately linear increase in measurement costs. Under these cost and sampling constraints there appears to be an optimal solution to sampling which requires approximately three composites which would come from around nine or ten strata (Figure 8). We changed the cost of fieldwork per hour to AUD 80 and the results suggested that around three or four composites should be used with four strata. Inevitably, there is considerable uncertainty about the optimal sample size particularly about the two main parameters (ρ and $\bar{\gamma}_h$) used in the estimation of variance of the mean change between two sampling periods. To appreciate the variability in these parameters we consider in the following section extant data and its use in guiding sampling design.
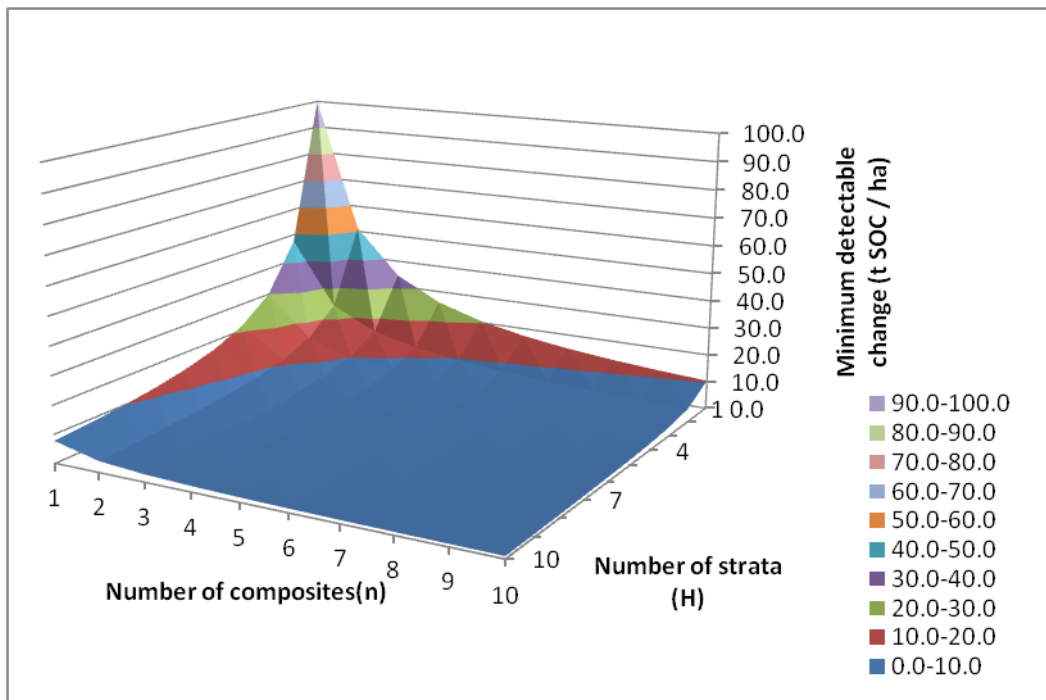
Figure 8 The minimum detectable change (one-sided 5% significance test (i.e., α=0.05) and 95% confidence in detecting a change in mean soil organic carbon (SOC) between t_1 and t_2 i.e., β=0.05) and a variance $\widehat{V} = 20$ (t SOC ha^{-1})^2 (for mean value $\bar{\gamma}_h$ of an exponential variogram with sill=30 (t SOC ha^{-1})^2 and effective range=0.15 km and autocorrelation ρ_k=0.2) associated with a sampling design involving stratification and composited soil for SOC measurement

Following Brus and Noij (2008) we set aside the fixed cost fieldwork because the optimal number of strata and composites is independent of it. The total, fieldwork and measurement cost of sampling under the constraints used to produce Figure 8 (assuming the same costs as above) is shown in Figure 9 for a contour of $\hat{\bar{d}}_{2,1} = 3.7$ t SOC ha^{-1}.
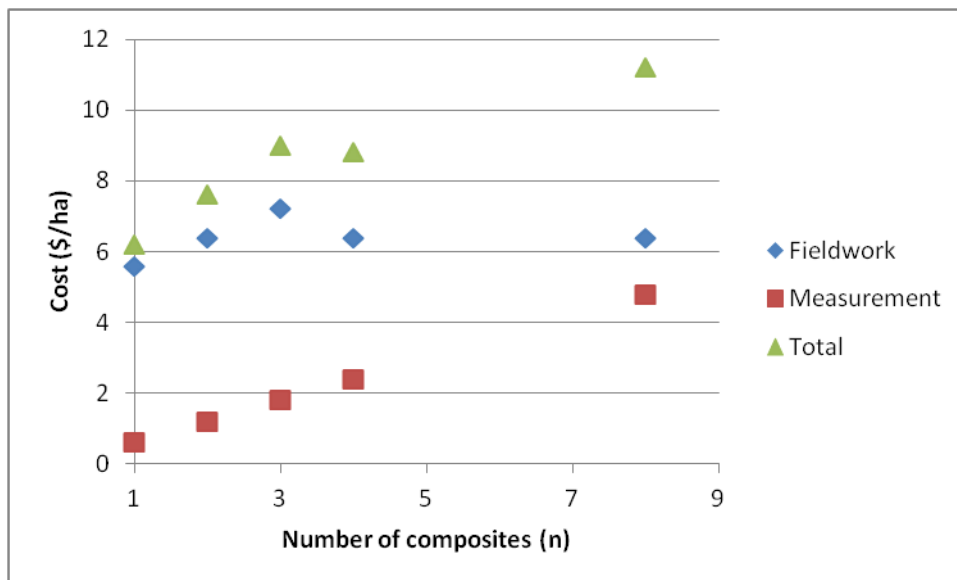
Figure 9 The cost of a stratified simple random sampling design using the constraints applied to Figure 8 for a minimum detectable difference $\hat{\bar{d}}_{2,1} = 3.7$ t SOC ha^{-1} in mean soil organic carbon (SOC) between t_1 and t_2

Using available data and its variability to guide initial sampling design

To implement the recommended simple stratified random sampling design with compositing, we need to know how many strata and how many composites to take and to balance those requirements against the cost. At the outset we have little or no information on the variability in the CEA so we have used published studies from Australian and international experiments to characterise variation in SOC over space and time. The initial sampling design can be based on the expected variation of SOC over space and time in the CEA (drawn from published studies). An alternative to using published studies to estimate spatial variation in SOC is to undertake a reconnaissance survey in the first sampling round. In that survey we would need to obtain 100 to 150 samples which would each need to be measured individually for SOC. The sampling and measurements costs would be large and the outcome may be to match approximately the assumed characteristics of the spatial variation we found from the literature. Consequently, although the use of the reconnaissance survey is in principal the ideal approach, it is likely to be poor value for money in many situations.

1.7   Spatial correlation

We are interested in using variograms to describe the variation of SOC that occurs in fields across Australia and the variation of SOC over large areas or regions of Australia. For example, we might be interested in a particular agricultural region influenced by soil type where a particular land use or management practice is common.

A preliminary description of the spatial dependence of soil organic carbon (SOC variograms) from several field studies in different countries was provided by McBratney and Pringle (1999). Bishop and Pringle (2012) augmented that database with more recent studies and presented 18 variograms in total (Figure 10). They found many more variograms published than these provided but used criteria of being peer-reviewed, covering areas 1–400 ha and 'raw' rather than de-trended or transformed data. Note that the units of semivariance are (soil organic carbon concentration expressed as a percentage of total soil mass)^2.

We used the %SOC variograms collated by Bishop and Pringle (2012) to identify characteristic variograms (Figure 11) and their model parameters (Table 1). The conversion of variograms (%SOC^2 to [t SOC ha^{-1}]^2) assumed an average soil bulk density of 1.3 Mg m^3 (obtained from SCaRP). We selected a range of variograms to represent large variability with a small spatial structure, medium variability with a medium spatial structure and small variability with a large spatial structure.

These variograms are based on measurements made over a very small support which for our purposes can be regarded as a point and thereby producing a point or punctual variogram. However, our requirement is to know the mean value for the stratum (Eq. 31) and potentially even use a different value for each stratum. The variogram can be related directly to the dispersion variance and the latter can be estimated from the former and thereby provide a representative estimate of e.g., the field within-stratum variance. The dispersion variance is an average squared difference, which explicitly accounts for individual sample support and the mean. For a finite region R that is divided into field B, which are in turn further subdivided into approximately equal sized square areas with length b for stratified sampling, the total dispersion variance $s^2(b \in R)$ is partitioned as:

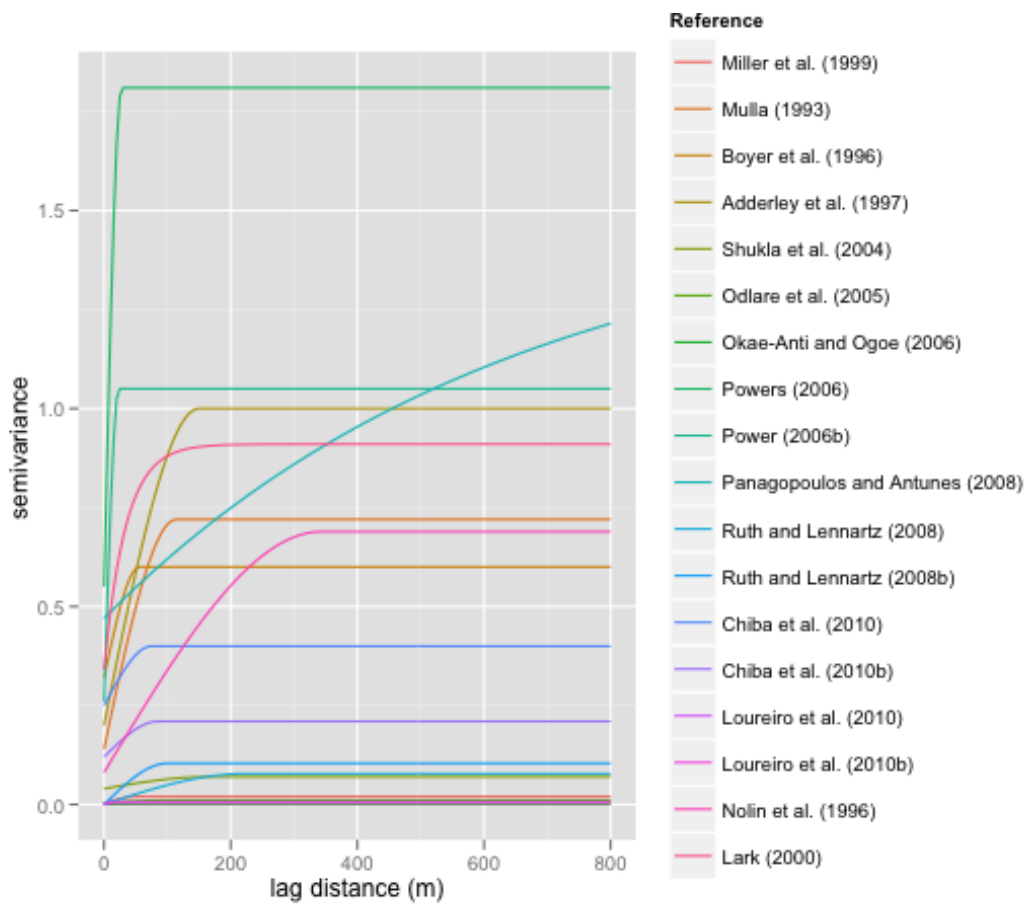$$s^2(b \in R) = s^2(b \in B) + s^2(B \in R). \qquad (33)$$

Figure 10 Published soil organic carbon variograms (% soil organic carbon concentration[2]; reproduced from Bishop and Pringle, 2012)
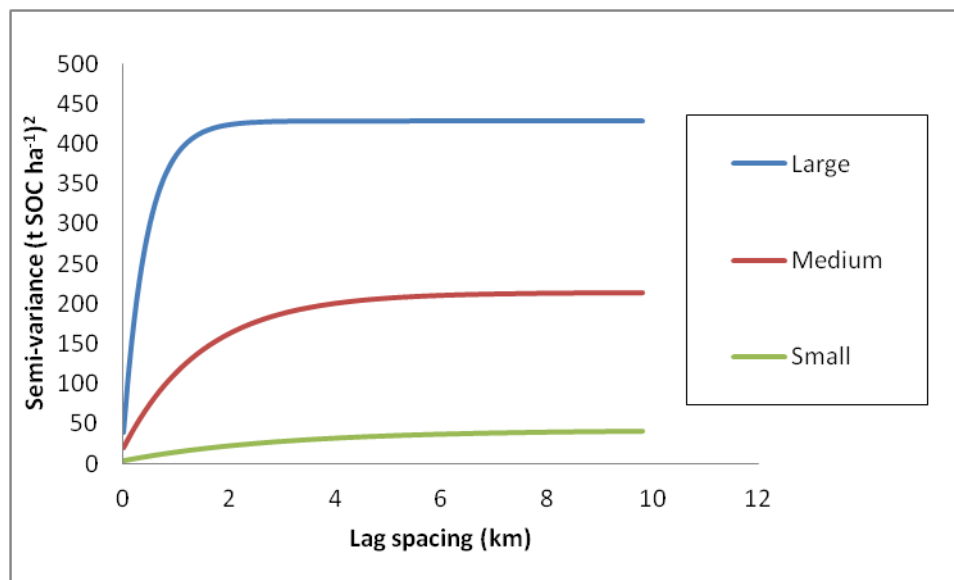

Figure 11 Characteristic variograms (t SOC ha[-1])[2] used in subsequent analyses. The parameter values of these models are shown in Table 1.

Table 1 Parameters of point variograms used to estimate the mean values $\bar{\gamma}$ of point exponential variograms $\gamma(h)$ for different sized strata within a field of 1 sq km (100 ha)

Model	Sill (c; %SOC^2)	Nugget (c_0 0.1xc)	Sill (c; $t\,SOC\,ha^{-1})^2$	Distance parameter (a; km)	Mean values $\bar{\gamma}$ of the point exponential variogram $\gamma(h)$			
Strata size					1 km	0.5 km	0.2 km	0.1 km
Large	1	39	390	0.15*	273.57	193.09	113.74	79.10
Medium	0.5	19.5	195	0.5*	71.92	48.51	31.92	25.80
Small	0.1	3.9	39	1*	9.75	6.98	5.17	4.54

*The exponential model is asymptotic there is an effective range $a' = 3a$

This is known as Krige's Relation (Journel and Huijbregts, 1978). The dispersion variance partitions can be calculated from mean values $\bar{\gamma}$ of the point variogram $\gamma(h)$. This is because the variogram is a function of the sample support:

$$\gamma_b(h) = \bar{\gamma}(b, b_h) - \bar{\gamma}(b, b), \qquad (34)$$

If $|h|$ is large relative to the distances across the support then $\bar{\gamma}(b, b_h)$ is approximately the point semivariance at lag h and

$$\gamma_b(h) = \gamma(h) - \bar{\gamma}(b, b), \qquad (35)$$

such that when $|h| \gg \sqrt{area\ of\ b}$ the regularized variogram is derived from the point support by subtracting the dispersion variance of the support which depends only on the length of the support of the regularization (Journel and Huijbregts, 1978; p. 90). The procedure in which the variogram from one support is obtained from that of a smaller support is known as regularization. It is particularly useful when bulking samples because the supports can be very large (e.g., portions of, or entire, fields) and if a variogram is known for a small support then the regularized variogram for the bulked support can be deduced (Webster and Oliver, 2007; p. 64).

There are two approaches to calculating the mean values $\bar{\gamma}$ of the point variogram $\gamma(h)$: numerically using a computer or by analytical solutions in successive stages (Journal and Huijbregts, 1978). The former is used by Bishop and Pringle (2012) and Singh *et al.* (2013). The numerical approach is dependent on the sample support which is largely unknown particularly if compact geographical areas are used. However, we used this approach and assumed that the support was square. We provide an example for the characteristic 'large' variogram displayed in Figure 11 and Table 1. Figure 12 demonstrates how equation 35 may be reconciled using the approximation for regularization of the exponential model and hence the mean value $\bar{\gamma}$ estimated for use in $V_{St}(\hat{\bar{d}}_{2,1})$ (Eq. 31). It shows the exponential point (or punctual) variogram for a 1 sq km (100 ha) field (with an effective range of 0.15 km) and the theoretical regularized variogram (Eq. 35) for several block sizes. As the block size increases from a point, the sill diminishes, the nugget variance disappears and the approach of the variogram near the origin becomes concave upwards (not evident at this display scale).

The mean values $\bar{\gamma}$ of each point variogram $\gamma(h)$ were calculated using the numerical (approach outlined above) for a range of block sizes which represented the subdivision of a field into smaller strata (Figure 13). Thus, we are able to estimate the mean values $\bar{\gamma}$ of a range of point variograms $\gamma(h)$ and to consider its impact on the sampling variance of the difference $V_{St}(\hat{\bar{d}}_{2,1} + \Delta E_c P)$ (Eq. 31).
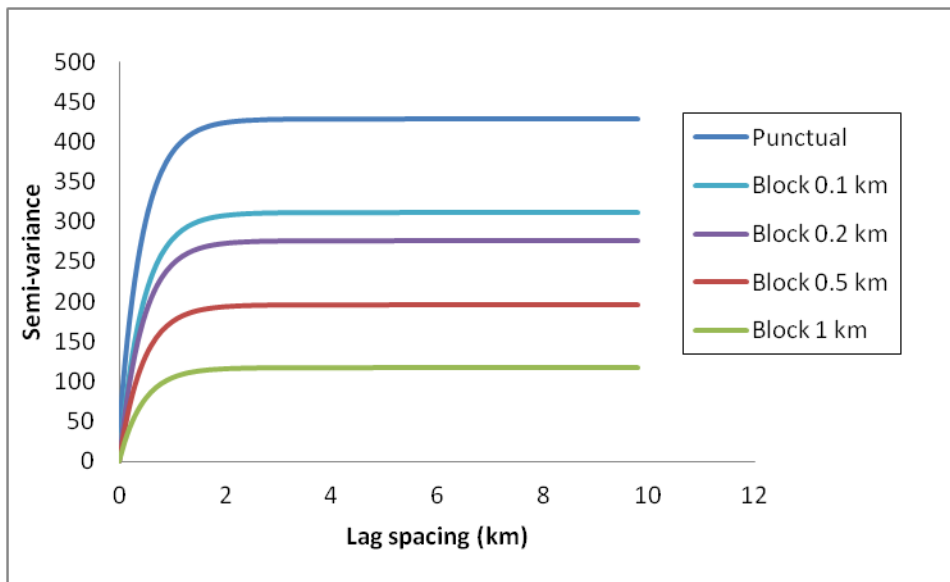
Figure 12 Punctual (or point) exponential variogram with a nugget effect and its regularized equivalents for several block sizes
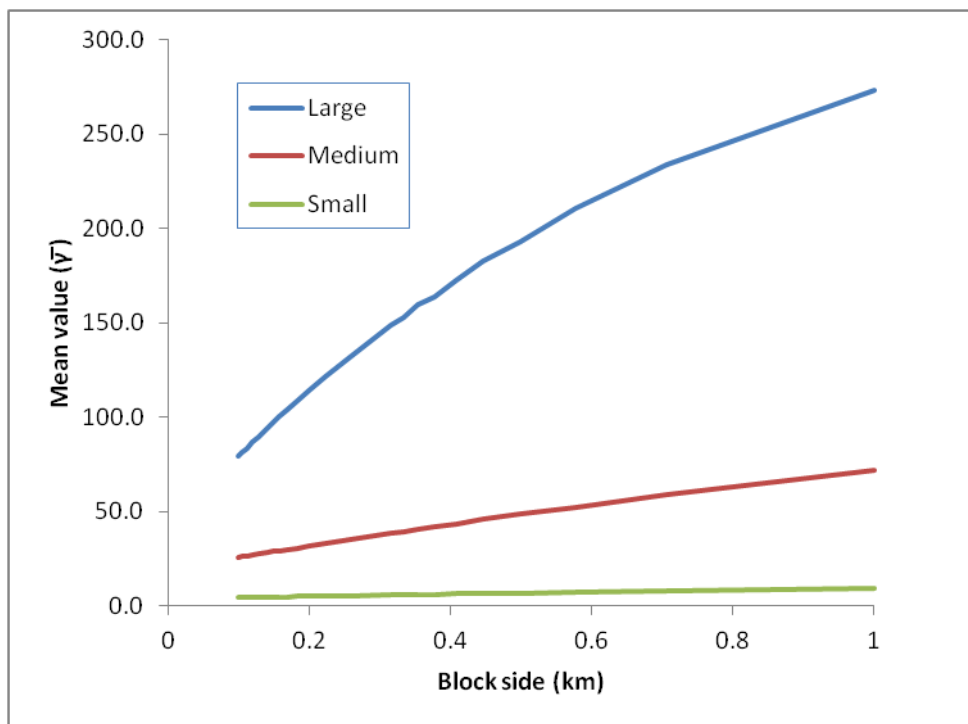

Figure 13 Mean values $\bar{\gamma}$ of the point variogram $\gamma(h)$ were calculated for a range of block sizes which subdivided a field (1 sq km) into smaller strata

1.8   Temporal correlation

The population autocorrelation coefficient at lag k, ρ_k, is the ratio of the covariance to the variance:

$$\rho_k = \frac{cov(x_{t,}x_{t-k})}{var(x_t)}.$$ (36)

This yields a set of values for ρ_k which can be plotted against the temporal lag to show how the pattern of correlation varies. This correlogram will display values close to 0 when the time series at that lag interval is random. If there is a short-term correlation the values will start close to +1 and decrease to 0 when the length or range of correlation is exceeded. The data are most easily examined if reported at regular intervals. Unfortunately, measurements of SOC are rarely measured regularly. However, models of carbon cycling fitted to those observations provide regular intervals of values which closely approximate the variation of SOC over time. We propose to calculate ρ_k on these model output data as a first approximation.

One of the most reliable and often used sources of data in Australia on the temporal variation of SOC are from experimental sites. These types of data have been used to calibrate the FullCAM (RothC) carbon cycling model (e.g., Skjemstad *et al.*, 2004). In particular, two long-term (>10 years) trials are available: Tarlee (Mediterranean, 34.28 S 138.77 E) and Brigalow (semi-arid, subtropical, 24.83 S 149.73 E). The Mediterranean and equi-seasonal rainfall distributions experienced at the two sites represent the rainfall distribution of a significant fraction of the Australian agricultural region. Potential exceptions include WA which has a more Mediterranean distribution and northern Australia which is more episodic. Two additional sites (Wagga Wagga and Warra) have been used by Luo *et al.* (2011) to model long-term soil carbon dynamics and sequestration potential in semi-arid agro-ecosystems. The data for the Warra site Queensland, Australia (26.93°S, 150.92°E) were obtained from Dalal *et al.* (1995). The site had a subtropical climate and has been used for cereal production since 1935. The data for the site at Wagga Wagga Agricultural Institute, New South Wales, Australia (35.11°S, 147.37°E) were obtained from staff at the institute. The site has a temperate climate with uniform rainfall distribution. Figure 14 shows the measured data and the modelled predictions of SOC for the selected experimental sites.

We used the model estimates of mainly the surface layer (0-10 cm) to calculate the autocorrelation coefficient for several lag time intervals. Note that the surface layer was not available for the Tarlee site so we used the 0-30 cm layer. Figure 15 shows an example of the values for Total SOC (t SOC ha^{-1}) at the Brigalow experimental site. Table 2 summarises the variation in ρ_k for each of the experimental sites and treatments. Our results show some periodicity for Warra which are probably artefacts of variation in the modelling. These results are excluded from further analyses and are not discussed further. The results for the remaining experimental sites show, as expected, a decline in the correlation related to time and the rate of that decline is dependent on the experimental site and presumably the land use / management practice. However, for the purposes of subsequent analysis here we chose to use only the five year interval for the three experimental sites (excluding Warra). We used the average (Table 2 in red) where there were more than one autocorrelation values for each site i.e., Brigalow $\rho_k = +0.121$, Tarlee $\rho_k = +0.279$ and Wagga Wagga $\rho_k = +0.388$.
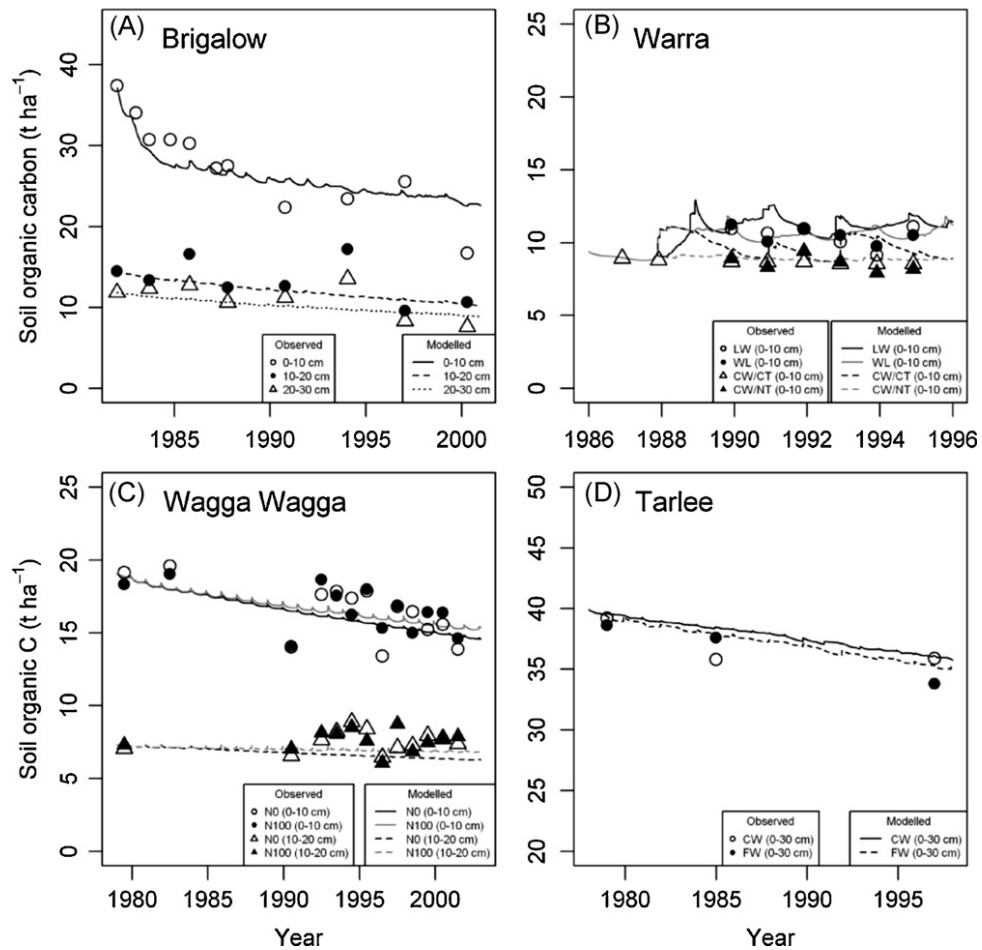
Figure 14 Modelled and observed soil (total) organic carbon (t SOC ha^{-1}) in different soil layers under different agricultural practices in Australia (reproduced from Luo *et al.*, 2011)
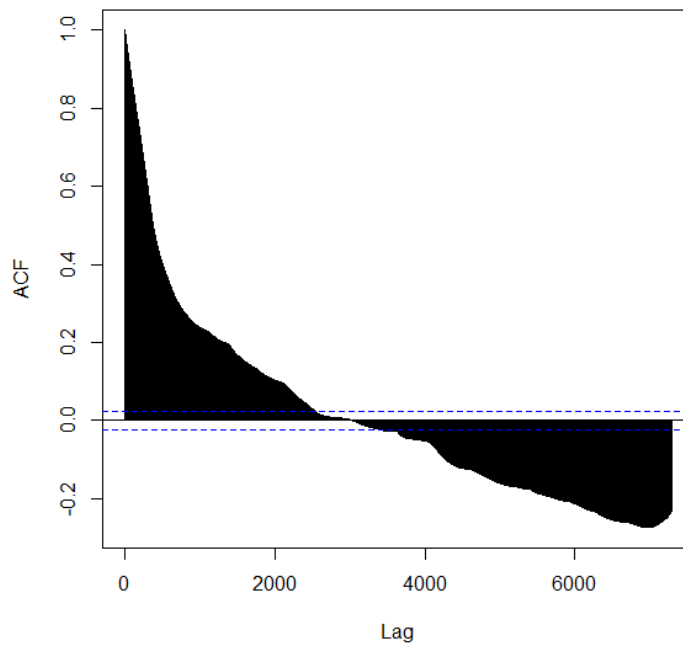

Figure 15 Autocorrelation function (ACF) for lag time interval (days) of soil organic carbon (t SOC ha^{-1}) for Sorghum-Wheat at the Brigalow experimental site

Table 2 Variation in the autocorrelation coefficient for several lag time intervals of modelled soil organic carbon at experimental sites across Australia

Experiment site and treatment	ρ_{3year}	ρ_{5year}	ρ_{7year}
Brigalow (TotalC 0-10 cm)			
Sorghum-Wheat	+0.229	+0.121	+0.021
Warra (TotalC 0-10 cm)			
Continuous Wheat, Conventional tillage	-0.116	+0.461	-0.263
Continuous Wheat, no-till	-0.008	-0.085	-0.074
Lucerne-Wheat	-0.044	-0.032	-0.021
Wheat-Lucerne	+0.115	+0.153	+0.161
Wagga Wagga (TotalC 0-10 cm)			
Continuous Wheat 000N	+0.622	+0.389	+0.168
Continuous Wheat 100N	+0.628	+0.387	+0.155
Tarlee (TotalC 0-30 cm)			
Continuous Wheat	+0.550	+0.284	+0.045
Fallow Wheat	+0.538	+0.274	+0.027

1.9   Sampling costs

We proposed values for fieldwork and measurement costs based on our experience in Australia (Table 3). In any case, these values are tentative first approximations.

Table 3 Values proposed for the cost parameters for sampling and measurement

Cost parameters	Singh *et al.* (2013)	Proposed for Australian SOC
Fieldwork fixed cost	NA	AUD 2000
Time (t per hour) to sample one aliquot	0.125	0.33
Fieldwork costs (c_{fld} cost per hour)	AUD 132	AUD 120
Laboratory costs (c_{lab} cost per sample)	AUD 20	AUD 30

Uncertainty in outcomes of monitoring designs

For use with our statistical model there are numerous different values that could be used and which influence the number of strata and composites to provide a solution to stratified simple random sampling of SOC stocks in fields across Australia. In this limited analysis we are unable to consider the full extent of this variation. Instead, to indicate likely solutions we have selected values based on the available data. This has provided characteristic values which may be used in various combinations with the other variables.

To demonstrate this variability and its impact on the decision-making process for stratified simple random sampling and its associated cost we have made some simplifying assumptions. We have essentially fixed values for some of the variables:

a. the cost structure is as used above,
b. a field size of 1 km^2
c. a one-sided 5% significance test (i.e., α=0.05)
d. 95% confidence in detecting a difference in mean soil organic carbon (SOC) stock between t_1 and t_2 (i.e., β=0.05).

This leaves the following variables for which we require values:

- the variance of the mean change which is influenced by the number of strata and the number of composites
- the mean value of the semi-variance which is also influenced by the number and size of the strata and
- the autocorrelation in the mean change over time.

Using the selected variograms (Table 1), and each of the autocorrelation coefficients (Table 2) we have estimated the SOC mean change total variance (Eq. 32) for a reduced range of strata (1-10) and composites (1-10). The results are plotted against cost for the same combination of strata and composites (Figure 16). Note that the cost is for two sampling rounds of a given combination of strata and composites. The difference in the variograms and their mean values of the semi-variances for the strata causes more variation (within a plot) than differences in the range of temporal autocorrelation coefficients (between plots). In other words, there is more variation in space than there is in time. This enables us to simplify the decision-making process.

Focusing on one of the plots in Figure 16, shows that as minimum detectable change (MDC) increases the cost decreases for a given variogram. For a given MDC there is variation in the cost which is a function of the number of strata and number of composites that may be used. Notably, that variation extends over the space that is covered by each of the variograms i.e., there is an envelope for each set of results and some overlap between the results.

Once a particular MDC has been chosen for a particular variogram the range of costs must be reconciled to a practical stratified random sampling design. This can be done by considering the composition of those costs in terms of the number of strata and the number of composites (Figure 17). For example, taking the most conservative situation: an autocorrelation coefficient of +0.121 (Figure 16a) and the largest variation for an MDC of 4 t SOC ha^{-1} provides a cost range from \$16-\$24 ha^{-1} (which depends on the number of strata and composites). In Figure 17 the cost band of \$20 ha^{-1} reveals that we can use a range of numbers of strata and composites. Following the shape of the contour indicates that as we reduce the strata we must increase the number of composites. We know that the total cost of sampling involving fieldwork and laboratory costs (excluding the fixed fieldwork cost) increases more rapidly with an increase in composites than with strata. So minimising the number of composites for our \$20 ha^{-1} cost contour band to 2 suggests that we need to use around 9-10 strata.
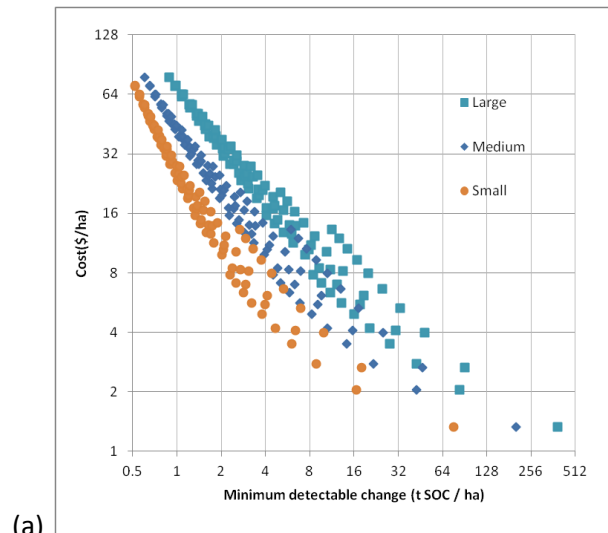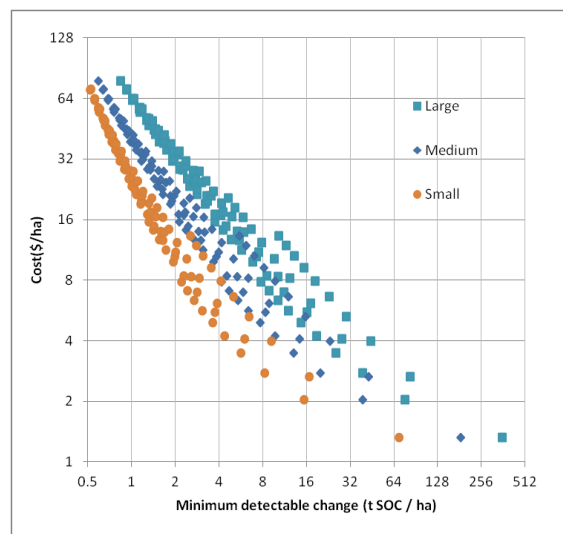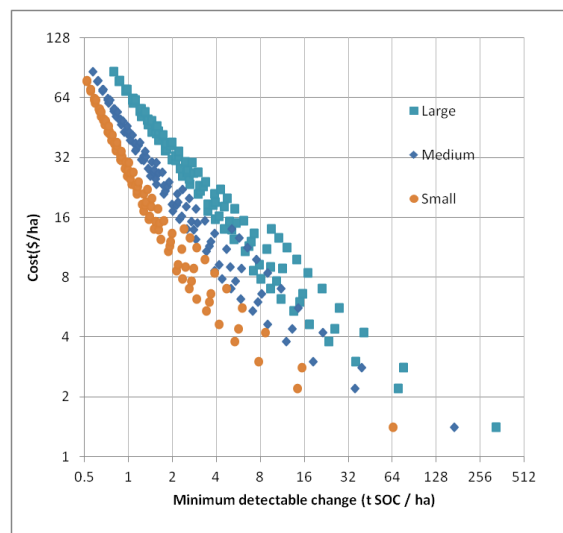
Figure 16 Variability in stratified simple random sampling using three variograms to produce mean semi-variances for a range of strata and composites. The autocorrelation coefficients $\rho_k = +0.121$ (a), $+0.279$ (b) and $+0.388$ (c)
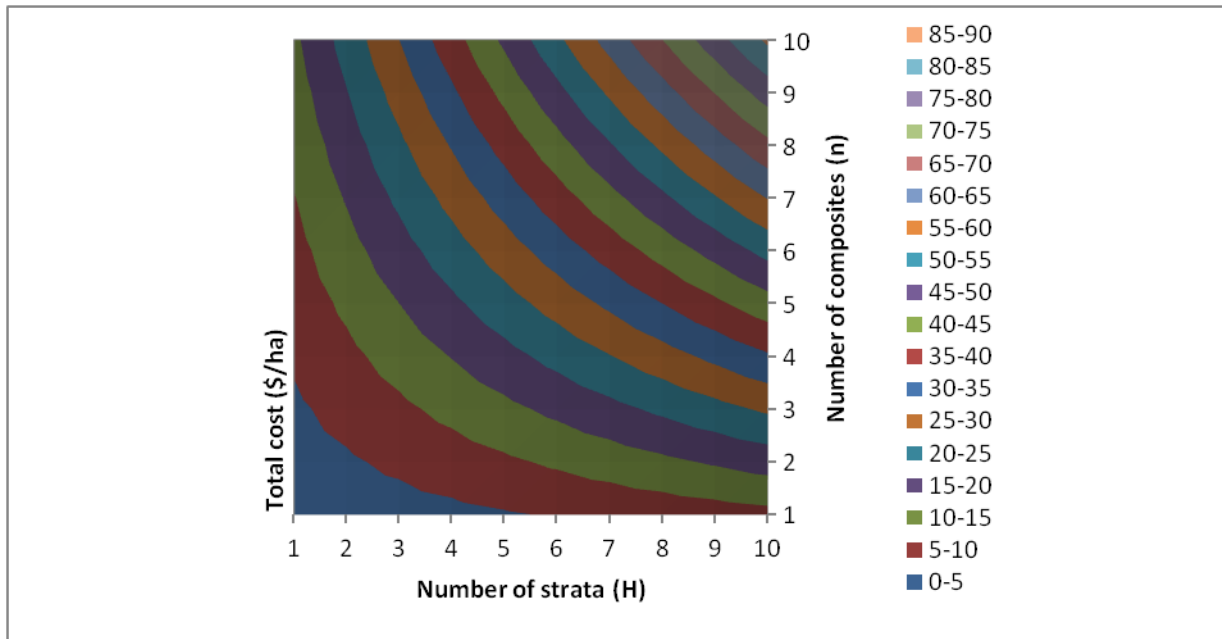
Figure 17 The total cost per ha (excluding fixed cost) associated with a sampling design (two rounds) involving stratification and composited soil for SOC measurement

Although perhaps 10 strata could be used it may be difficult to implement across a field. The regular grid advocated to provide a basic stratification of a field, makes the stratification dependent on readily divisible numbers. In this situation we would probably opt for 10 strata and 2 composites if the field was rectangular or 9 strata and 2 composites (if the field was square) and tolerate a reduced MDC.

Once these decisions have been made it is then possible to repeat the statistical tests in section 4.3. Unlike the results of that previous section we are now able to make estimates based on simple random stratified sampling obtained from a CEA with the same number of samples from t_1 as from t_2. We have estimated that we can afford to obtain from the field N=18 samples (i.e., 9 strata and 2 composites) from each period t_1 and t_2 (r=1 i.e., $N=N_1+N_2$=36) with a mean change variance of 8 (t SOC ha^{-1})^2($\hat{V}=8$) approximately constant over time. The SOC is to be compared using a one-sided 5% significance test (i.e., α=0.05). We wish to be 95% sure of detecting a difference in mean soil organic carbon (SOC) between t_1 and t_2 (i.e., β=0.05). Assuming that there is no SOC redistribution and under these given conditions, the minimum detectable difference (MDC) is

$$\hat{\hat{d}}_{2,1} = (1.734 + 1.734)(\tfrac{2(8)}{36})^{0.5} - 0 = 2.31 \text{ t SOC ha}^{-1}.$$

However, it is not clear from these results how sensitive the MDC is to the choice of mean change in variance of SOC. This is not straightforward to assess because as is evident from the above calculation the MDC is a function of both variance and number of samples (composites x strata). In turn the variance has been estimated using the number of samples (composites x strata) and assumptions about the variability of SOC over space and time. Consequently, to appreciate the impact of uncertainty in each of the variables we need to consider multiple interactions simultaneously which is likely best achieved using Monte Carlo simulations in a sensitivity analysis.

Without the time to conduct a sensitivity analysis we provide a partial solution which demonstrates albeit in a limited way the impact of chosen values / assumptions on minimum detectable change in SOC. The example described above, which estimates an expected minimum detectable change (MDC) of 2.31 t SOC ha^{-1} over 5 years, is based on the most variable of situations. The decisions that were made about the

conditions for this estimate, including the spatial and temporal variation, have a considerable impact on the MDC.

Table 4 illustrates how the MDC would differ if we made different decisions about the variability that we might expect in our CEA. Costs are not considered variable and are fixed at those values used previously. Statistical levels of significance are also not considered to be variable and are fixed at those values used in the previous scenario. Given the small variance associated with temporal variation we also fixed that to the most variable situation ($\rho_k = +0.121$).

Table 4 Influence of decisions and chosen values on the calculation of the minimum detectable change in soil organic carbon stocks over 5 years (assuming no SOC redistribution)

Samples (strata x composites)	Cost ($/ha)			Target minimum detectable change (t SOC ha^{-1})		
	Field	Laboratory	Total	Large spatial variation	Medium spatial variation	Small spatial variation
4 x 2	5.76	1.2	6.96	12.27	6.35	2.96
6 x 2	8.64	1.2	9.84	7.46	3.96	2.03
8 x 2	11.52	1.2	12.72	5.32	2.89	1.60
4 x 3	8.64	1.8	10.44	7.88	4.14	2.07
6 x 3	12.96	1.8	14.76	4.88	2.65	1.48
8 x 3	17.28	1.8	19.08	3.52	1.97	1.19
4 x 4	11.52	2.4	13.92	5.82	3.11	1.64
6 x 4	17.28	2.4	19.68	3.65	2.03	1.20
8 x 4	23.04	2.4	25.44	2.64	1.52	0.98

The results of this table should be considered carefully. They should not be interpreted to mean that assuming the spatial variance is small will enable a small minimum detectable change (MDC). In contrast, these results show that the MDC is influenced by the spatial variation in the CEA. If that variation is under-estimated then so too will the variance mean change SOC and MDC be under-estimated and the measured values will not match the expected or target values. Notwithstanding the importance of taking a conservative approach (large spatial variation), Table 4 shows that a small MDC over a 5 year time period can be achieved by using only a moderate number of strata and few composites. Notably in the most spatially variable situation, when using only two composites and any fewer than 8 strata very quickly increases the MDC to > 5 t SOC ha^{-1}.

It is evident from these results that much of what has gone before is dependent on, and therefore sensitive to, the spatial variation that we characterised to be large. This characterisation was based on available data much of which did not originate in Australia. There may be CEAs which are more variable than the situation we have portrayed as large spatial variation. In that case, the measured MDC will not match the expected or target MDC in much the same way as if a medium spatial variation was selected where large spatial variation was appropriate. There will always be a possibility of this happening until every CEA has been sampled. Consequently, in the absence of information on the spatial variation, the best approach is to take as many samples as can be afforded to gain a first approximation.

Summary and conclusions

The flowchart below (Figure 18) is a summary of the process (decisions and information) required to initiate a stratified simple random sampling design and identify the number of strata and composites.
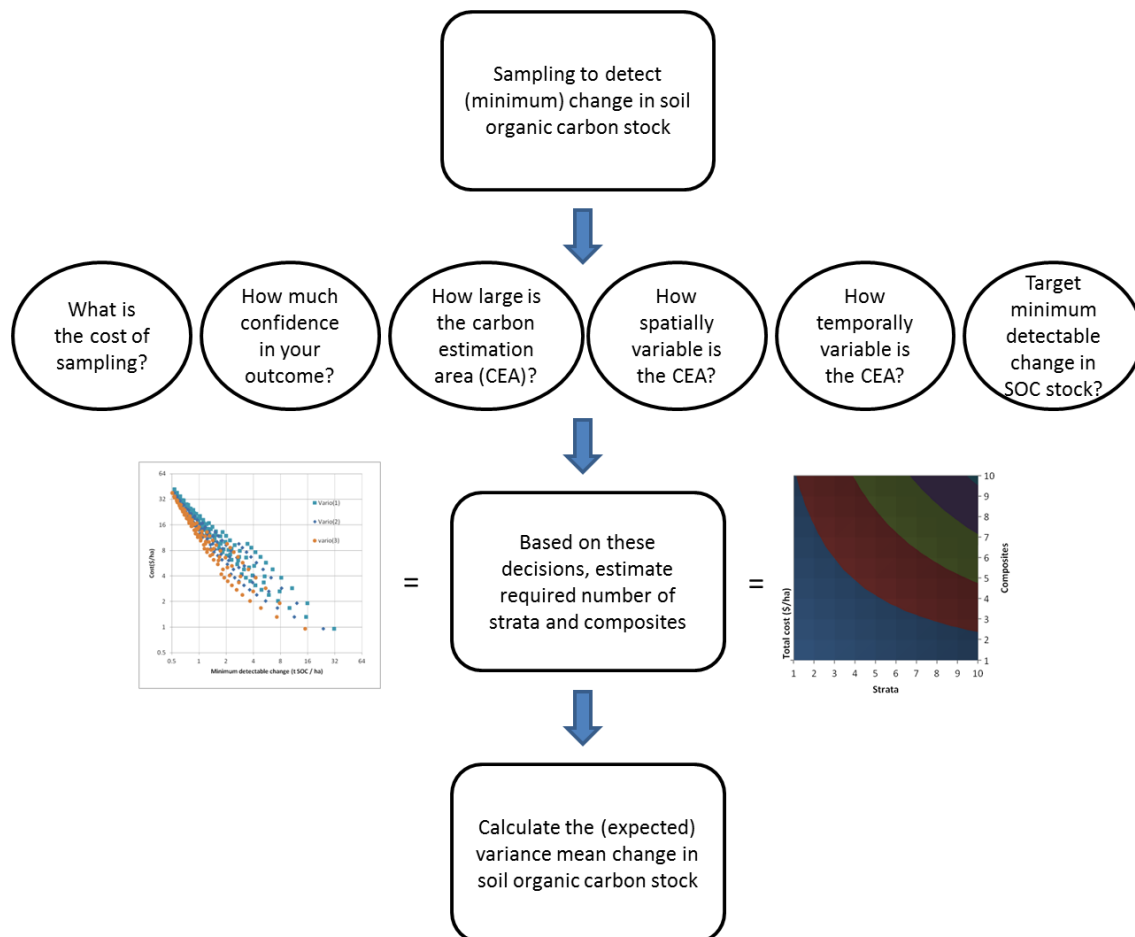

Figure 18 Flowchart showing the decisions and information necessary to determine the number of strata and composites to calculate the minimum detectable change in soil organic carbon stocks

Once the sampling has been undertaken and the measurements of soil organic carbon stocks have been made, the variance of the composites can be compared to the expected variance of the CEA (Figure 19). Depending on the difference between the measured and expected variance the sampling design may be adapted to further reduce the variance after the second round of sampling.

For statistical testing of hypotheses on the temporal change of a global mean, that is the mean of an area of interest as a whole, a design based method is most appropriate. This implies that sampling locations are selected by probability sampling. Stratified simple random sampling is recommended to ensure that samples are spread evenly across the area of interest (CEA). This approach requires that each stratum covers an equal area or where unequal areas are used the samples are weighted accordingly. The proposed sampling design makes use of the temporal correlation. This information is maximised if the distances between sample locations at the first and second sampling time are as close as possible. This leads to a static-synchronous pattern of sampling. Composite values from each of the stratum considerably reduce measurement cost. Repeated composites ensure that information on the sampling variance is provided and do not increase prohibitively the total sampling cost.

Examination of data, available from the literature including studies in Australia, provided values to represent variation of SOC in space and time. Using these data suggested that the variation of SOC in time is much smaller than that over space. This finding leads to a preliminary conclusion that information about the variogram and its mean value for a particular stratum size is of considerable importance to the sampling. Unfortunately, this type of information is not available for individual fields. Data on the variogram parameters could be collected in a reconnaissance survey. However, this is an expensive exercise particularly when focused on a single field because more than one hundred samples and measurements of SOC would be required. A reconnaissance survey of this type would be more cost effective across the farm and would provide a valuable first approximation to the spatial variation within a field. Notably, sampling of this type across a farm may cross two or more soil types in which SOC may be more or less variable. This situation should be treated with appropriate sampling and mapping approaches.
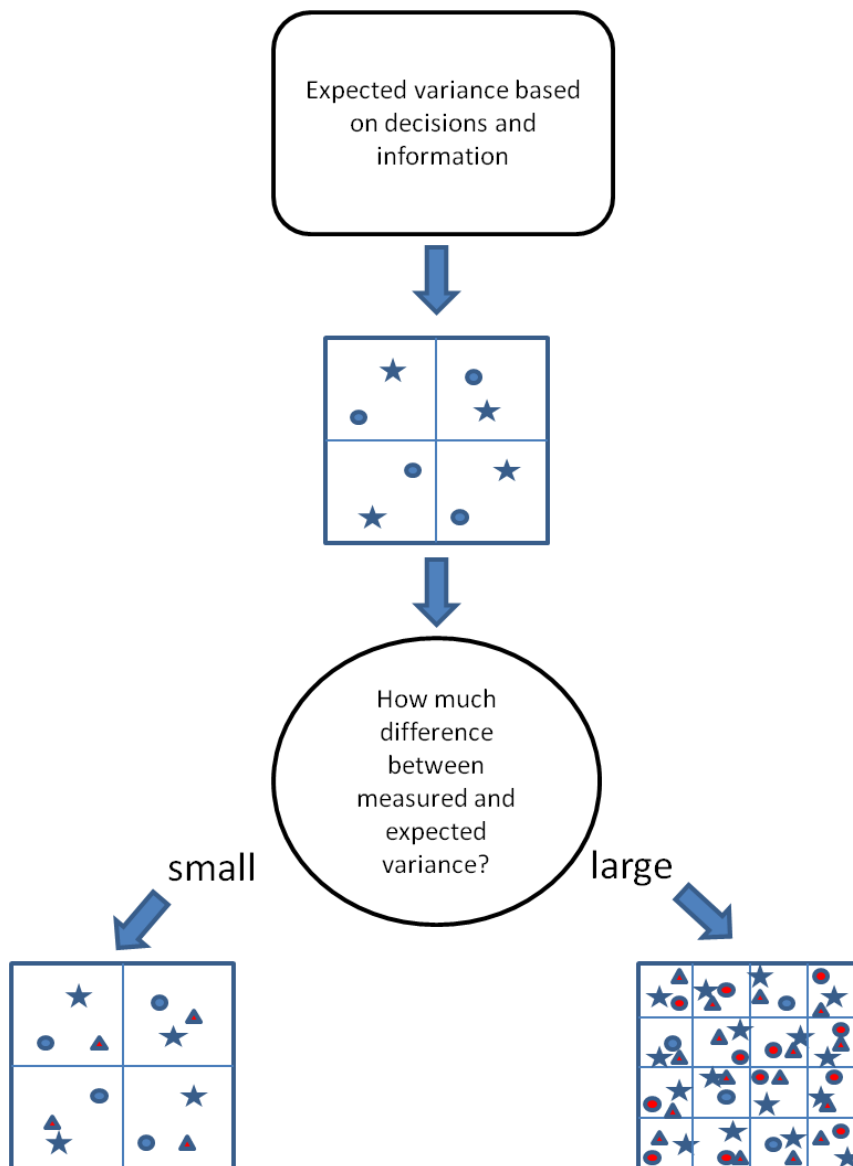

Figure 19 Flowchart showing how the stratified simple random sampling decision may be adapted based on the comparison of measured and expected soil organic carbon stock variances

In the absence of information on SOC variability, particularly on its spatial variability, it is recommended at the outset of sampling to be conservative and design a sampling strategy for areas of interest exhibiting large spatial variability. This will lead to collecting more samples than perhaps is necessary and bearing a larger cost than is necessary but will ensure that the desired detectable change is achieved. In short, choosing 'large' variation will estimate mean values of the semi-variances which are larger than the other variograms and will most likely produce a minimum detectable change smaller than that expected. Any attempt to inappropriately reduce cost via a reduction in either number of strata or number of composites, resulting in an inadequate number of samples, will not provide the expected detectable change.

There are two sources of information that may be used to inform the most likely influential parameters of the model. A carbon cycling model can be used to predict change in carbon over time for a specific location. It is not essential that the predictions are absolutely correct, only that they provide a consistent estimate of carbon and its decomposition. Notably, the direct effect of soil redistribution to add and subtract SOC is not included in most carbon cycling models. Also not included in C cycling models is the indirect effect of soil redistribution to change the soil micro-environment (moisture, temperature, roughness etc.) and influence carbon decomposition. However, the potential remains for such modelled estimates to provide a first approximation of the autocorrelation function and the value which should be used in the sampling design. Similarly, it is likely that a temporal series of yield maps spanning at least 5 years, or maps of some other soil property related to SOC that has a fine resolution at the field scale, could be used to approximate the variogram and hence estimate the mean value and components of variance for defining an appropriate number of strata and composites for a CEA.

Inevitably, in work of this nature there is a need for a sensitivity analysis to identify the most influential parameters of the sampling design and also to determine approximate ranges for values of those parameters. Additional funding could be used to develop a numerically optimised solution to how many strata and how many composites should be used to achieve a minimum detectable change. For example, this could be done relatively quickly by fitting a polynomial to the cost surface and another polynomial to the MDC surface and finding their intersection for an MDC less than a given value and thereby identify the possible combinations of strata and composites to achieve the target.

Glossary

^{137}Cs technique – A technique for the estimation by wind, water and tillage of the net (ca. 30-40 years) outcome of soil erosion and deposition based on the measurement of background levels of radioactive material fixed strongly to the soil which was released in atmospheric nuclear weapons testing.

Aliquot – An exactly defined (depth, surface area etc.) sample of the soil which in this case is also associated with an exactly defined location in a field.

Ancillary variables – also known as secondary variables, which are related often directly by some physical process but which may also have an indirect relationship to the primary variable which in this case is soil organic carbon e.g., crop yield may be related to soil organic carbon.

Autocorrelation coefficient – a measure of the similarity between observations as a function of the time separation between them and used to identify patterns or change over time.

Covariance – a statistical description of how two random variables behave in relation to each other. The covariance is associated with the variables' correlation.

Compact geographical stratification – an approach to classification which uses the coordinates of a fine regular grid placed within an irregular area of interest to define polygons which have equal area and which are as compact as possible.

Composite – a sample of soil which represents soil from many other samples and is typically used to provide a single representative sample from a field or area of interest

Correlation – the quantified similarity between two or more properties in either space (spatial correlation) or time (temporal correlation). The correlation can be positive or negative indicating a proportional increase in one variable as the other variable increases and *vice versa*, respectively. Correlation in either space or time can be generalised as measurements made close together in either space or time are typically more similar than those made at larger space or time intervals. This is commonly referred to as spatial or temporal dependence.

Correlogram – a plot of autocorrelation coefficients for different time intervals; it is typically plotted with the lag time interval on the x-axis and the autocorrelation coefficients on the y-axis.

Destructive soil sampling – removing a core of soil from the ground leaves a void which will naturally become in-filled over time. However, samples of soil which require a repeat measurement should avoid such disturbed locations and return as close as possible to the original location.

Exponential variogram – a exponential model which is fitted typically using least-squares and a best-fit statistic to the experimental variogram to determine the values of the model parameters.

Monte Carlo simulation – computational algorithm that relies on random sampling to obtain numerical results, often used to model phenomena with significant uncertainty in inputs.

Nugget variance – the often small amount of semi-variance that remains unexplained in a semi-variogram and which is due to the combination of measurement and location error and spatial variation which is smaller than sampling resolution

Point or punctual variogram – the semi-variogram is a function of the sample support and consequently when samples are measured over such small support (soil cores) that they may be considered to be a point then the variogram represents the spatial variation for point samples; sometimes referred to as punctual variogram

Regularization – The variogram is a function of the sample support. The larger the support is the more variation each measurement encompasses and the less there is in the intervening space to appear in the variogram. It is commonly used to refer to the change in variogram structure with sample support.

Rotational sampling pattern – a compromise between static sampling and synchronous sampling in the sense that the locations of the previous sampling time are partially replaced by new ones.

Sample support – traditionally defined as the exact specification of a sample in each dimension and important because it influences the structure of the variogram

Sampling variance – refers to variation in a particular statistic (e.g., the mean) calculated of a sample after being repeated many times.

Semi-variance – a calculation of the variation due to difference in values at different separation distances in space. The variance is evidently halved to avoid duplication with a given direction of a particular distance.

Sensitivity analysis – Often conducted using Monte Carlo simulation to determine those variables which have the most influence on the outcomes of the results particularly when the variables have considerable input uncertainty

Sill variance - the value of the semi-variance where it reaches a plateau in the variogram and indicates the maximum amount of variation which is approximately that of the variance.

Spatial correlation – see correlation.

Spatial variance – the amount of variation typically about the mean that is found from samples at different locations in space.

Standard normal distribution – the Normal distribution with a mean of 0 and a standard deviation of 1

Static sampling pattern - Static sampling occurs at a fixed set of locations and may or may not follow the same pattern in time.

Static-synchronous sampling pattern – Synchronous sampling refers to a different set of samples for each sampling time which may or may not have the same spatial patterns. If they are the same and they coincide spatially then the pattern is static-synchronous.

Synchronous sampling pattern - Synchronous sampling refers to a different set of samples for each sampling time which may or may not have the same spatial patterns.

Temporal correlation – see correlation.

Variogram – or the experimental variogram, which is subsequently fitted with a model, characterises the spatial variation. It is constructed by plotting the separation distance on the x-axis and the matching semi-variance on the y-axis.

References

Allen D.E., Pringle M.J., Page K.L. & Dalal R.C. 2010. A review of sampling designs for the measurement of soil organic carbon in Australian grazing lands. *The Rangeland Journal*, 32: 227-246.

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 Soil Science Conference, Hobart, Tasmania.

Brus, D.J & Noij, I.G.A.M. 2008. Designing sampling schemes for effect monitoring of nutrient leaching from agricultural soils. EJSS doi: 10.1111/j.1365-2389.2007.00996.x

Brus, D.J., Spätjens, L.E.E.M., and de Gruijter, J.J. 1999. A sampling scheme for estimating the mean extractable phosphorus concentration of fields for environmental regulation. *Geoderma*, 89:129–148.

Chappell A., Viscarra Rossel R.A. 2013. The importance of sampling support for explaining change in soil organic carbon. *Geoderma*, 193-194: 323-325.

Chappell A, Sanderman J, Thomas M, Read A, Leslie C. 2012 The dynamics of soil redistribution and the implications for soil organic carbon accounting in agricultural south-eastern Australia. *Global Change Biology*, 18, 2081–2088.

De Gruijter, J.J., Brus, D.J., Bierkens, M.F.P. & Knotters, M., 2006. Sampling for Natural Resource Monitoring. Springer, Berlin, Heidelberg, New York.

Dalal, R.C., Strong, W.M., Weston, E.J., Cooper, J.E., Lehane, K.J., King, A.J., Chicken, C.J., 1995. Sustaining productivity of a Vertisol at Warra, Queensland, with fertilisers, no-tillage, or legumes. 1. Organic matter status. Aust. J. Exp. Agric. 35, 903–913.

Domburg P, de Gruijter JJ, Brus DJ 1994 A structured approach to designing soil sampling schemes with prediction of sampling error from variograms. *Geoderma* **62**, 151–164. doi:10.1016/0016-7061(94)90033-7

Goidts, E., van Wesemael, B., Van Oost, K., 2009. Driving forces of soil organic carbon evolution at the landscape and regional scale using data from stratified soil monitoring. Global Change Biology 15, 2981-3000.

Grealish, G. Clifford, D., Wilson, P and Ringrose-Voase A 2011. National Soil Condition Monitoring Program for soil pH and soil carbon. Objectives, Design, Protocols, Governance and Reporting. CSIRO Publication, Canberra.

Gregorich EG, Greer KJ, Anderson DW, Liang BC 1998. Carbon distribution and losses: erosion and deposition effects. Soil and Tillage Research, 47, 291–302.

Journel, A.G., Huijbregts, ChJ., 1978. *Mining Geostatistics*. Academic Press, London,

Kachanoski RG, de Jong E 1984. Predicting the temporal relationship between soil cesium-137 and erosion rate. Journal of Environmental Quality, 13, 301–304.

Lal R. 2001. Soil degradation by erosion. *Land Degradation and Development* 12: 519-539.

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, 748-756.

Loughran, R. J., G. L. Elliott, D. J. McFarlane, and B. L. Campbell 2004. A survey of soil erosion in Australia using caesium‐137, Aust. Geogr. Stud., 42(2), 221‑233, doi:10.1111/j.1467-8470.2004.00261.x.

Luo Z., Wang E., Sun O.J., Smith C., Probert M.E. 2011. Modelling long-term soil carbon dynamics and sequestration potential in semi-arid agro-ecosystems. *Agricultural and Forest Meteorology*, 151: 1529-1544.

McBratney, A.B., Pringle, M.J., 1999. Estimating average and proportional variograms of soil properties and their potential use in precision agriculture. Precision Agriculture 1, 125-152.

McKenzie NJ, Henderson B, McDonald WA 2002. Monitoring soil change: principles and practices for Australian conditions. CSIRO Land and Water Technical Report 18/02, Canberra.

Morvan, X., Saby, N.P.A., Arrouays, D., Le Bas, C., Jones, R.J.A., Verheijen, F.G.A., Bellamy, P.H., Stephens, M., Kibblewhite, M.G., 2008. Soil monitoring in Europe: a review of existing systems and requirements for harmonisation. Science of the Total Environment 391(1): 1–12.

Olson, K.R. 2013. Soil organic carbon sequestration, storage, retention and loss in U.S. croplands: Issues paper for protocol development. *Geoderma*, (195-196): 201-206.

Olson, K.R., 2010. Impacts of tillage, slope and erosion on soil organic carbon retention. *Soil Science* 175, 562–567.

R Development Core Team, 2010. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0 / http://www.R-project.org.

Saby, N.P.A., Bellamy, P.H., Morvan, X., Arrouays, D., Jones, R.J.A., Verheijen, F.G.A., Kibblewhite, M.G., Verdoodt, A., Berenyiuveges, J., Freudenschu, A., Simota, C., 2008. Will European soil-monitoring networks be able to detect changes in topsoil organic carbon content? *Global Change Biology* 14, 2432-2442.

Sanderman J., Chappell A. 2013. Uncertainty in soil carbon accounting due to unrecognised soil erosion. Global Change Biology, doi: 10.1111/gcb.12030

Sanderman J, Baldock J, Hawke B, Macdonald L, Massis-Puccini A, Szarvas S 2011. National Soil Carbon Research Programme: field and laboratory methodologies. CSIRO Publication, Canberra.

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

Skjemstad, J.O., Spouncer, L.R., Cowie, B., Swift, R.S., 2004. Calibration of the Rothamsted organic carbon turnover model (RothC ver. 26.3), using measurable soil organic carbon pools. *Aust. J. Soil Res*. 42, 79–88.

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

Webb N, Chappell A, Strong C *et al.* 2012. The significance of carbon-enriched dust for global carbon accounting. Global Change Biology, doi:10.1111/j.1365-2486.2012.02780.x.

Webster R, Oliver MA 2007. *Geostatistics for environmental scientists*. (Wiley: Chichester, UK)

Woodward, M. 1992. Formulae for sample size, power and minimum detectable relative risk in medical studies. *Journal of the Royal Statistical Society*, Series D (The Statistician), 41(2):185-196.