Geoderma Regional 13 (2018) 1–14

Contents lists available at ScienceDirect

Geoderma Regional

journal homepage: www.elsevier.com/locate/geodrs

Auditing on-farm soil carbon stocks using downscaled national mapping products: Examples from Australia and New Zealand

Brendan Malone[a,*], Carolyn Hedley[b], Pierre Roudier[b,c], Budiman Minasny[a], Edward Jones[a], Alex McBratney[a]

[a] Sydney Institute of Agriculture, The University of Sydney, Eveleigh, NSW 2015, Australia
[b] Landcare Research, Private Bag 11052, Manawatu Mail Centre, Palmerston North 4442, New Zealand
[c] Te Pūnaha Matatini, A New Zealand Centre of Research Excellence, Private Bag 92019, Auckland 1142, New Zealand

ABSTRACT

A method has been tested to audit soil carbon stocks at the farm-scale. This is needed for emission trading schemes and for tracking sustainability of the soil resource through time. The method begins with baseline maps derived from national models which are then disaggregated with fine scaled environmental data via statistical modeling; before optimally stratifying to guide soil sampling positions. Field sampling provides a statistically valid estimate of soil carbon stocks, and the method is repeated through time to monitor any changes in stocks. Case studies are presented from Australia and New Zealand.

While spatial downscaling is useful for generating soil maps relevant to the farm scale, the optimal stratification of these maps for guiding soil sampling for baseline soil carbon auditing purposes should not be recommended if the national scale mapping is thought to be unreliable. Because of unresolvable differences between spatial scales associated with bias and incorrect specification of uncertainties, results from Australia revealed stratified simple random sampling was not as efficient when compared to the less costly simple random sampling. Conversely, results from New Zealand do show that stratified simple random sampling to be more efficient than simple random sampling.

From soil sampling, clear differences in carbon stocks to 30 cm were observed when comparing the stocks measured at sites from both countries. In New Zealand, soil carbon stocks were estimated to be as high as 101 t/ha for the top 30 cm of soil. For the Australian sites which were all situated in the Hunter Valley region of NSW, the highest measured soil carbon stocks were 25 t/ha for the top 30 cm. The largest soil carbon variance was observed at the New Zealand hill country farm, where the landscape consists of alluvial terraces, complex broken valley sides and a summit plateau mantled with volcanic ash. In Australia, the presence of subsoil pedogenic carbonate (marl) contributed to high variance estimates relative to the other sites.

1. Introduction

Soil is the largest reservoir of terrestrial carbon, containing more than three times that of the combined reservoirs of atmospheric carbon and vegetation biomass (Jobbágy and Jackson, 2000). Soil carbon is a dynamic variable, and can potentially influence the direction and magnitude of carbon cycle-climate feedbacks (Friedlingstein et al., 2014). With elevated levels of atmospheric CO_2 there has been considerable recent interest into carbon sequestration, whereby atmospheric CO_2 is converted into soil carbon which is long-lived. In effect, this is reversing the carbon lost from soils historically through land use change and degradation (Minasny et al., 2017; Paustian et al., 1997). It has been estimated that soils worldwide have lost almost half their initial carbon levels, and sequestration could exert a significant role in mitigating greenhouse gas emissions (Stockmann et al., 2013).

There has been much global research investigating various management practices that promote soil carbon sequestration (Sanderman and Baldock, 2010). Relatedly, there is also particular interest in the investigation of efficient methods for the auditing of soil carbon (de Gruijter et al., 2016). Examples of soil carbon monitoring programs that have been initiated throughout the world, and how they operate are described in Arrouays et al. (2014a). Some specific examples from Europe and the US are described in Saby et al. (2008) and Spencer et al. (2011) respectively. For monitoring soil carbon, first one needs to establish baseline levels (stocks), which is then followed up with successive revisits to measure and monitor changes through time.

* Corresponding author.
E-mail address: brendan.malone@sydney.edu.au (B. Malone).

https://doi.org/10.1016/j.geodrs.2018.02.002
Received 15 December 2017; Received in revised form 16 February 2018; Accepted 23 February 2018
Available online 27 February 2018

The motivation of this particular research is to test a potentially useful soil carbon auditing protocol that is applicable to the farm or agricultural landholding spatial domain. The auditing protocol largely builds upon the one proposed by de Gruijter et al. (2016) where their objective variable was the estimation of whole farm soil carbon stocks (SCS) with sufficient statistical confidence. The new development in this context being the incorporation of disaggregated national and regional scale digital soil carbon mapping.

For background context, using a direct measurement approach, de Gruijter et al. (2016) established that design-based sampling is the most appropriate approach for auditing soil carbon. With design-based sampling, the implication is that the sampling locations are selected by probability sampling and that the inference or estimation is based on the sampling design used to select the sampling locations. The benefit of this is that the subsequent inference is unbiased and free of any modeling assumptions. The simplest method in probability sampling is by Simple Random Sampling (SRS) where within a target area (a farm etc.), a fixed number of locations where a sample is to be collected, are selected at random and independently of each other. One common approach towards improving the sampling efficiency of SRS is via Stratified Simple Random Sampling (StSRS), where the target area is divided into smaller sub-areas (strata) followed by SRS within each stratum. On the proviso that the stratifying variables are appropriate, the gains in efficiency of StSRS include a lower expected sampling variance, such that the within strata sampling variances will be quite different from each other, but collectively lower than that attained by SRS. The outcome of this is that for the same sampling variance, a fewer number of samples is required for StSRS compared to SRS.

de Gruijter et al. (2016) describe various ways to stratify a farm for the intention of soil carbon auditing. These may be scaled according to what is known about the spatial variation of soil carbon on the farm. For example, if nothing is known of soil carbon variability, compact geographical stratification may be used (Brus et al., 1999). This just means stratification of the grid coordinates within the mapping domain, and has the useful property of achieving spatial coverage of samples within the sampling domain. The stratification may be improved if there is some ancillary information available that describes the environmental spatial variation on the farm. For example, maps of elevation (and subsequent derivatives), yield data, and any other data that may have been collected via proximal soil sensing such as electrical conductivity or gamma radiometrics as a few common types. With a multivariate suite of ancillary data, stratification would entail a k-means classification procedure. Stratification is best achieved however if a digital soil map of soil carbon is available for use as a univariate source of information. It is argued that a digital soil map of on-farm soil carbon variation embodies the quantitative relationships between covariates (ancillary information) and measured soil carbon together (de Gruijter et al., 2016). With the univariate source of information, stratification of a soil carbon map involves locating stratum boundaries along the variable distribution.

The caveat to this type of stratification (using the soil carbon map as a stratifying variable) is that there is an assumption the digital map is error free. Obviously, this is a false assumption as it is known that varied sources of uncertainty are propagated through to spatial predictions of phenomena including soil variables. Consequently, it has become expected that in digital soil mapping literature, produced maps also need to be accompanied by associated maps of quantified uncertainty. Using digital soil map uncertainties, de Gruijter et al. (2015) proposed an alternative stratification method whereby they are explicitly incorporated. The stratification method, called *Ospats*, uses a raster of predicted values and associated error variances for deriving the sampling strata. *Ospats*, by taking into account the prediction variations is able to produce stratifications that represent transitions between the 'knowing nothing about soil carbon variation' (high mapping uncertainty) and 'knowing a lot about soil carbon variation' (low mapping uncertainty) situations. The theoretical and mathematical explanation of *Ospats* is described in detail in de Gruijter et al. (2015) with some details of this also described in the materials and methods section of this paper.

Ospats is embedded into a soil carbon auditing protocol (detailed in de Gruijter et al., 2016) that describes the steps to follow for first establishing baseline SCS (first round of mapping and sampling) through to deriving the sample design for follow-up revisits in order to quantify whether change has occurred. A useful feature of the protocol is that information collected during the establishment or first round of sampling are incorporated into the design of the next round of sampling and subsequent revisits. This mechanism usefully allows for the continual improvement of sample design efficiency and the realisation of more accurate digital soil mapping, and ultimately more precise estimates of on-farm SCS. A criticism of this type of space-time sampling design is that the stratification at sampling time 1 differs from that of time 2, and as a consequence the sampling locations also change, i.e. the sampling locations are not fixed. This can lead to less precise estimates of the change in carbon stocks, as the approach does not exploit temporal correlation. This is indeed an issue, but from a practical perspective for the auditing of soil carbon, it is not reasonable to sample in the same place twice. For SCS auditing, it is necessary to lessen the opportunity for unethical practices such as site tampering. Also estimations of SCS are destructive. If one returns to a position one metre away there could be an effect of short range variability, for example from tree roots or manure additions (e.g. see Goidts et al., 2009 and Hedley et al., 2012). Using *Ospats* and the associated stratification and sampling approach seem appropriate trade-offs in this regard.

For universality of application however, *Ospats* and the associated auditing protocol needs relevant farm scale digital mapping of SCS and associated prediction uncertainties. In reality, with the exception of some farms, most agricultural landholdings will not likely have this type of mapping easily on hand.

To circumvent this issue, it was proposed that relevant digital mapping could be acquired via the disaggregation or downscaling of national or regional scale mapping (Malone et al., 2017). Projects such as the GlobalSoilMap (Sanchez et al., 2009; Arrouays et al., 2014b), have created the situation that acquisition of digital soil information (with associated uncertainties) is at least a possibility across the spatial extent of a farm. In reality though, while such large spatial extent digital soil mapping products are invaluable for aiding the decision making process at the spatial scales they were intended for, they are not particularly relevant for considering issues at the farm management scale, which is typically finer. Spatial downscaling can be invaluable in this regard because the original mapping can be linked with fine scale environmental data via statistical modeling to derive similarly fine scale predictions of the target variable. The implicit assumption here is that the covariate information is strongly related to the target variable, which is being derived at the fine scale resolution.

To be tested in this study, is an addendum to the soil carbon auditing protocol established by de Gruijter et al. (2016). This is the coupling together of spatial downscaling (for deriving the digital soil maps with associated uncertainty) with the optimized stratification and sampling (*Ospats*). This allows the protocol to be generalized and potentially used where relevant farm scale soil carbon mapping is not presently available.

With examples from field studies conducted in both Australia and New Zealand, this paper reports on efforts to establish baseline SCS at selected sites in those countries. The materials and methods detail some of the underlying processes involved in spatial downscaling and optimizing stratification and sample selection. We report the stocks for the selected sites that were investigated. Importantly this paper discusses practical issues of implementing the auditing protocol that is proposed, and subsequently details some possible recommendations that require practitioner decisions to be made.

2. Materials and methods

The structure of describing the methods related to the generalized soil carbon auditing protocol is staged with the following problem definition:

What is required to estimate total soil organic carbon stocks (to a specified soil depth) with an associated variance estimate for a given landholding or farm where there is little to no existing soil carbon information?

In statistical terms, the farm or landholding is the "target universe" (de Gruijter et al., 2006). In addition to the space dimension that the target universe entails, there is also a time dimension. Auditing needs to occur at a particular repeated time interval to establish a change in the measure. In this research we do not estimate whether SCS have changed or not. Rather, we describe methods for the establishment of baseline on-farm SCS. In this study, the target variable for auditing SCS is reported in tonnes per ha to a specified depth. For both Australia and New Zealand the specified depth is 0–0.3 m. For different projects the target variable may differ. For example the target depth could be different, or the auditor may use cumulative mass of soil rather than fixed depth sampling. There might be other variants too, but whatever the target variable, it is important to specify it clearly at the outset of the auditing, and honor it through time.

From the initial problem definition, the sequence of methods is as follows and illustrated in Fig. 1.

1. The acquisition of national scale soil carbon mapping.
2. Spatial downscaling for generating relevant farm scale soil carbon mapping with associated uncertainty quantification.
3. Running and optimization of *Ospats* for generating strata and sample location selection.
4. Fieldwork component involving visiting the sites and specified sample locations to collect soil for analysis of carbon stocks.
5. Statistical inference for quantifying on-farm SCS.

3. Site descriptions

3.1. Australia

Sites selected in Australia were centered upon the Lower Hunter Valley region in New South Wales (32.83°S 151.35°E). Situated about 140 km north of Sydney, this region covers an extent of approximately 220 km^2 (Fig. 2A). Further general information about this region is described in Malone et al. (2014). There are varied land uses throughout the region, but it is best known as a viticultural hub. Four vineyards were selected from this area for this study. Each vineyard in general terms is quite distinguished from the others in terms of prevalent management practices and soil types. The participating Australian landholdings are shown in Fig. 2B:

1. AusCarb_S1: Relatively newly established vineyard (last 10–15 years) of multiple wine grape cultivars from previous land use of native vegetation and pasture for grazing. The site has approximately 28 ha of managed vineyard which is situated predominantly on weathered in situ soils from underlying mudstone parent materials. Chromic Luvisols and Dystric Nitosols (WRB) are the predominant soil types at this site. Between vine row management of soils involves little to no tillage with maintenance of perennial grasses.
2. AusCarb_S2: Well established vineyard exclusively dedicated to growing Semillon wine grapes. The site has approximately 7 ha of managed vineyard. Soils are predominately Dystric Regosols (WRB) composed of fine sandy to clay loam sediments that have been alluvially derived. Soils are lightly tilled every second row during September (around bud burst) and post-harvest to improve soil aeration and water capture. Continuous cover cropping is maintained in rows that are not tilled.
3. AusCarb_S3: Well established vineyard and open pasture landholding of approximately 94 ha. Soils, particularly where vines are

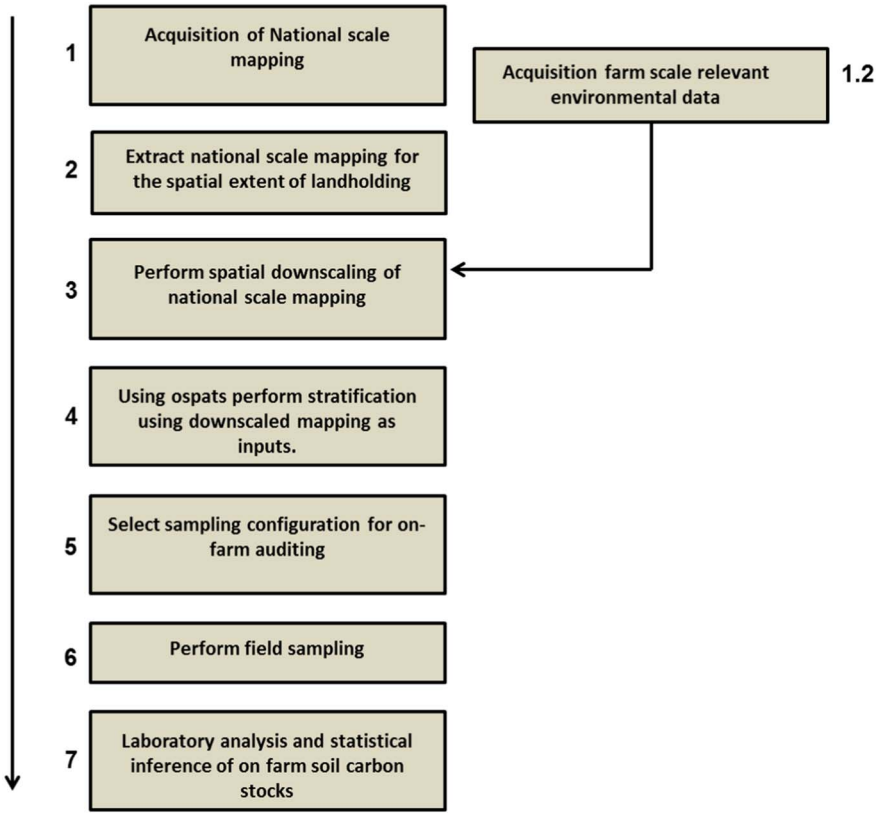

Fig. 1. Workflow for establishing baseline on-farm soil carbon stocks using a combination of downscaled national scale mapping and optimal stratification.

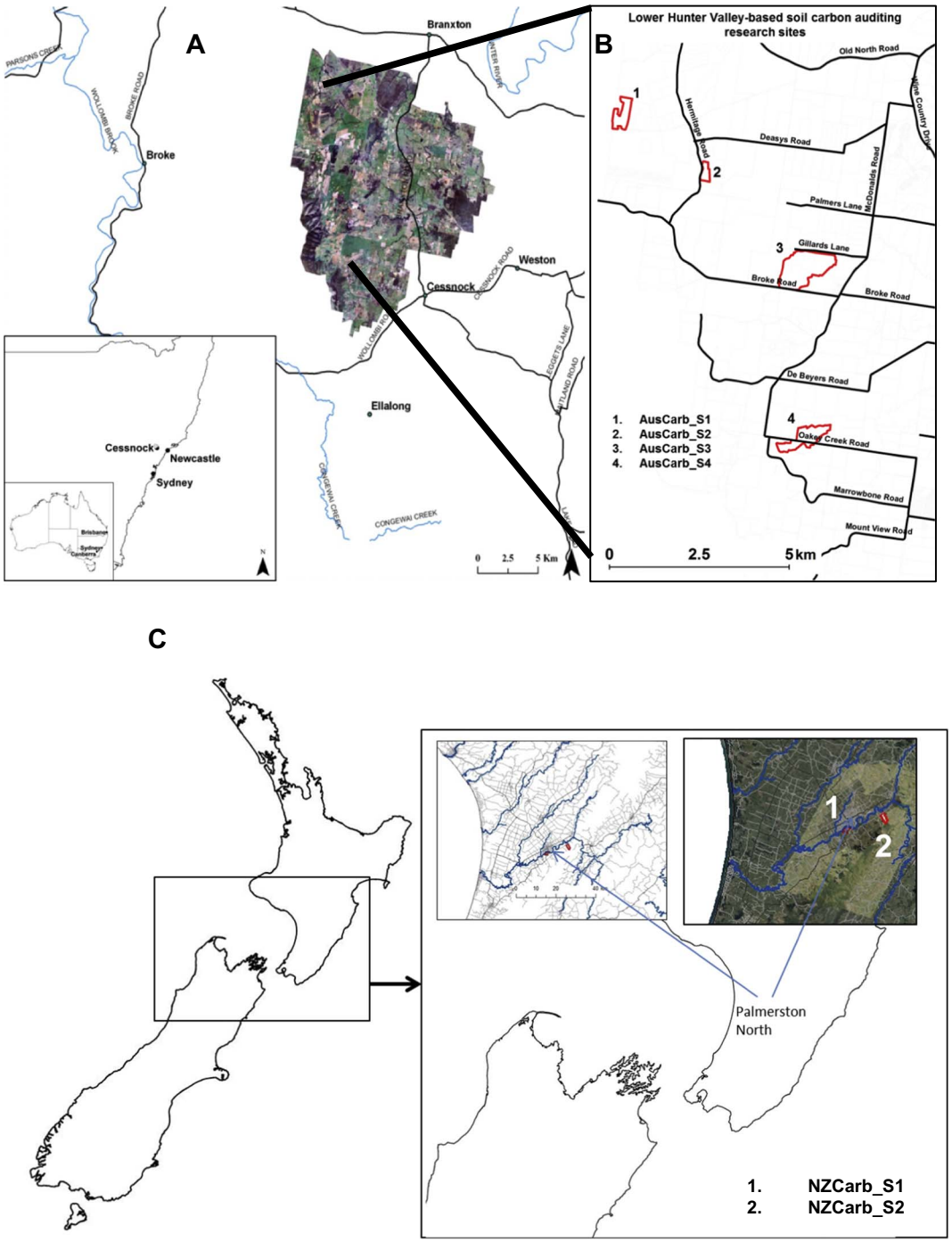

Fig. 2. Locality maps of study sites in both Australia and New Zealand. The Lower Hunter Valley location in relation to New South Wales and Australia (1A). The Australian study sites (1B). New Zealand study sites (1C).

established, often contain varying amounts of marl (loose, earthy deposits of calcium carbonate material). Because of the presence of marl in this environment, soils generally have neutral pH soils or higher and classify out as Calcic Luvisols (WRB). Other soils are predominantly either in situ weathered mudstones or a colluvium material derived from these same parent rocks and are usually Dystric Nitosols (WRB).

4. AusCarb_S4: Well established vineyard that has approximately 42 ha of managed wine grapes. The vineyard supports multiple wine grape cultivars upon soils that are quite variable. Soils can range from uniform medium clay soils derived from mudstones (Dystric Nitosols) to light textured (some places quite gravelly) alluvial soils (Dystric Regosols). The vineyard has a long history of continuous cultivation which occurs twice annually.

3.2. New Zealand

Sites selected in New Zealand are centred in the Manawatū region, in the lower North Island (Fig. 2C). The predominant pastoral farming of this region extends from flat alluvial river plains to rolling and steep hill country. Two pastoral farms were selected to represent these contrasting landforms:

1. NZCarb_S1 is located at Massey University, in Palmerston North,

adjacent to the Manawatū River (S40.374493° E175.61498°; 34 m ASL). It is a 160 ha farm, managed as a low input, sustainable pasture-based dairy farm with a once-a-day milking, spring calving system. Ryegrass (*Lolium perenne L.*) and white clover (*Trifolium repens L.*) pasture predominates and some fodder crops are also grown. Soils are predominantly fluvial recent soils (Dystric Fluvisols, WRB), forming in greywacke alluvium; drainage characteristics vary from well drained to poorly drained, depending on texture fineness of deposition layers and topographic position with respect to the river. Average annual rainfall is approximately 1000 mm, mean winter temperature is 8.5 °C and mean summer temperature is 17.9 °C.

2. NZCarb_S2 is a 218 ha subcatchment of a 476 ha sheep and beef farm owned by Massey University, 15 km NW of Palmerston North (S40.336969°, E175.733440°; rising from 60 m to 340 m ASL). It is managed as a productive sheep and beef farm, where stock are rotationally grazed, with ewes set stocked two weeks prior to lambing. The farm is divided into a flat land block (99 ha) and hill block (119 ha). Average annual rainfall is 1100 mm on the flat land, with a summer dry period. Pastures are predominantly a perennial ryegrass and white clover mix. Soils on the flat land are derived from windblown loess, with compacted subsoils causing slow drainage and waterlogging in winter time. These are Perch-gley Pallic soil (Stagnic Planosols). On the hill slopes Firm Brown and Orthic Brown soil (Fragic Cambisols and Cambisols) occur and on the summit Allophanic Brown soils (Andic Cambisols) (Hewitt, 2010; FAO, 1998). The hill block is rolling to very steep hill country, with the farm backing on to the slopes of the southern-most extent of the Tararua ranges. Rainfall gradually increases with elevation to reach 1300 mm of rain per year at the top of the farm where the environment is moist, relatively cool and wind-swept by prevailing westerlies and south easterlies. The pastures on the hill block are predominantly browntop, crested dogstail mixed with perennial ryegrass and white clover.

4. Data acquisition: national soil carbon model

4.1. Australia

We acquired national SCS mapping that was produced by Viscarra Rossel et al. (2014). This digital map quantifies the Australian national extent spatial variation of organic carbon stocks to 0.3 m. This digital map has a spatial resolution of 90 m and also has quantified prediction uncertainties, which are expressed as a 95% prediction interval.

4.2. New Zealand

Soil carbon data used for the New Zealand national model of SCS is the same data that is used in the New Zealand Soil Carbon Monitoring system to assess SCS changes with land use change to 0.3 m soil depth (McNeill et al., 2014). It contains the Historic Soils and Land Use and Carbon Analysis System (LUCAS) soils database. The Historic Soils database consists principally of National Soils Database data obtained for soil survey purposes, with additional data from other providers. The LUCAS soils database consists of soils data sampled from plots, principally from the natural forest land use class, although there are records from other land use classes as well. There are 1320 Historic records and 435 LUCAS records, for a total of 1755 records (McNeill et al., 2014). To develop the national SCS model, this point dataset was used with environmental data layers associated with soil carbon formation (including terrain, climate, land use and soil factors, e.g. stone content, soil order) to model SCS at a 1-km grid nominal resolution. A generalized linear regression model (GLM) was used in preference to a linear model, as the distribution of the carbon data is best described by a gamma distribution, not Gaussian, as would be required for a linear regression model. The GLM models the mean of the log of the soil carbon as a linear function of explanatory variables, and allows both the distribution of the response variable and the transformation of the explanatory variables to be separately specified (NZAGRC, 2016).

5. Acquisition of farm scale environmental data

In order to determine the high resolution spatial pattern of phenomena using a statistical downscaling approach such as *dissever*, there is a need to acquire equally high resolution covariate data to drive the downscaling. For the Australian and New Zealand sites this information was acquired by proximal soil sensing, although this was not possible at NZCarb_S2 where much of the terrain is too steep for driving across with an ATV. The proximal soil sensor surveys acquired bulk soil electrical conductivity and gamma radiometric data with associated positional and height information using differential GPS equipment. The proximal sensors used in Australia were a RSX-1 gamma radiometric detector consisting of a 4 L Sodium-Iodine crystal (Radiation Solutions Inc., Mississauga, Ontario, Canada), and a Geonics DUALEM-21S electromagnetic induction instrument (Geonics Ltd., Mississauga, ON, Canada; DUALEM Inc. www.dualem.com, Milton, ON, Canada). In New Zealand the same gamma sensor was used, and the electrical conductivity surveying was undertaken using a Geonics EM38 sensor (Geonics Ltd., Mississauga, ON, Canada). Data from these instruments were logged continuously with measurements being taken approximately every 1 m given an average speed of 5 km h^{-1}. Each logged measurement was tagged with a spatial reference coordinate using a digital GPS receiver (SMART6-L GNSS; NovAtel, Canada). In Australia, all three instruments were attached to a John Deer ATV and driven on average to a parallel line spacing of approximately 15 m. In New Zealand the surveys were conducted at different times with the sensors mounted onto an ATV (Hedley et al., 2004; Hedley et al., 2016). As proximal surveys were conducted mostly within vineyards in Australia, the 15 m line spacing was equivalent to every second vine row. In New Zealand the surveys used a swath width of 10 m.

For some theoretical background, the gamma-ray spectrometer records the amount of radioactive isotopes in the soil (top 30–50 cm) based on the principle that each gamma ray photon relates to a discrete energy window which is characteristic of the source isotope (Minty et al., 1998). Gamma radiometrics is a passive sensing technique that detects the varying amounts of naturally occurring radioisotopes of potassium (^{40}K), uranium (^{238}U-series) and thorium (^{232}Th-series) from soil as they produce high-energy gamma-rays with sufficient intensities to be picked up by the detector. Additionally, a total count gamma-ray measurement can be taken over the entire spectrum range too. Measurements were recorded in counts per second (cps). The reason for the interest in collecting this data for latter soil inference is that the radioisotopic signature of soils has previously been linked to soil physical, chemical and geochemical properties including soil texture and carbon, among others (Cook et al., 1996; Rawlins et al., 2009).

In terms of the electromagnetic induction as briefly mentioned earlier, EMI sensors measure bulk soil electrical conductivity. From these data, other soil properties of interest may be inferred using EMI data like soil moisture, pH and texture, which through covariance will also be useful for prediction of the spatial variation of soil carbon. The DUALEM-21S (Dualem, Milton, ON, Canada) sensor used in this project has dual-geometry (horizontal and vertical) receivers at separations of 1 and 2 m from the transmitter. This configuration provides four simultaneous depths of conductivity sounding or depth of exploration, that correspond to integrated conductivity measurements in mS m^{-1} for depths of 0–0.5 m, 0–1 m, 0–1.6 m and 0–3.2 m.

For each of the sites, the raw data (GPS, radiometric, and EMI data) were pre-processed then used as observational data for creating maps based on a regular 10 m grid cell resolution. We used punctual ordinary kriging to make these maps for each variable. Secondary terrain variables, slope gradient and topographic wetness index (function of slope and catchment area) were derived from the ground elevation. In

summary, for each Australian and New Zealand site we amassed the following fine scale covariate maps for use in the spatial downscaling of the national carbon map:

- Digital Elevation model: Elevation, slope gradient, terrain wetness index
- Gamma radiometric data: Total count, gamma K, Th and U
- Bulk soil electrical conductivity: 0–0.5 m, 0–1 m, 0–1.6 m and 0–3.2 m (Australia). 0–0.75 m and 0–1.5 m (New Zealand).

At NZCarb_S2, only LiDAR survey data was available, as the terrain was mostly too steep for vehicle mounted proximal soil surveying. The LiDAR data was collected by a local aerial mapping company and provided as a 3D geolocalised classified point cloud at approximately 1 pulse per m^2. The classified point cloud was used to create a digital elevation model at 10 m resolution by interpolating the elevation values of the ground points with a natural neighbor algorithm, using the "spdinterp" program of the Sorted Pulse Data (SPD) software library (Bunting et al., 2013). For the NZCarb_S2 site, the digital elevation model and derived terrain attributes were used as covariate data layers for the downscaling process.

6. Spatial downscaling of national digital soil maps

Spatial downscaling is a critical component of the proposed auditing protocol as it allows the delivery of relevant farm scale mapping to be used in the *Ospats* algorithm. However, in order to implement *Ospats*, a requirement is to have associated maps of the prediction variance. Within the context of spatial downscaling, one approach for doing this was proposed in Malone et al. (2017) whereby simulations of the national mapping are derived, that are subsequently downscaled using the *dissever* algorithm as described in Malone et al. (2012). Another approach would be with area-to-point kriging such as that shown in Orton et al. (2016). This approach is based on a statistical model, and results in a map with disaggregated predictions, as well as a map with kriging variances. From Malone et al. (2017), realisations are made in consideration of the prediction variance associated with the national mapping, such that the mean and variance of the simulated mapping will be same as the prediction and variance of the national mapping product. To numerically explore the uncertainty space, a random component (with some specified spatial correlation structure) is introduced into the simulations. This can efficiently be done using unconditioned Gaussian random fields with specified covariance parameters including nugget, partial sill and anisotropy etc., (Wood and Chan, 1994). The procedure for generating the unconditioned random fields is commonly referred to as sequential Gaussian simulation (Chiles and Delfiner, 1999). After generating a random field, we can then derive a possible realisation by using the following equation:

$$sim_{map} = \sigma x + \mu \quad (1)$$

where sim_{map} is the simulated map, σ is the map of the standard deviation of predictions from the national mapping, x is the Gaussian random field map, and μ is the national map. For each of the sites in Australia, 100 realisations were generated. For the covariance parameters of the random fields we selected the distance parameter that was found from the fitted variogram of the national map predictions from each site. The sill and nugget were set to 1 and 0 respectively. The underlying assumption with these parameters is the random field has unit variance, and the variogram has no error at 0 m separation distance.

The simulated national scale maps are then spatially downscaled. The underlying objective of *dissever* is to converge towards a solution that is mass-preserving, i.e. the mean of fine scale predictions is equivalent to the associated value of their encapsulating coarse scale pixel (of the national scale mapping). In its original conceptualization, downscaling is performed by modeling the relationship between fine resolution covariates and the coarse scaled map using a weighted generalized additive model (GAM), followed by subsequent iterations to achieve mass balance. Work by Roudier et al. (2017) generalized *dissever* through the allowance of users to select from a suite of potential models besides GAMs. In fact one way to implement *dissever* is to optimize the downscaling by comparing outcomes from different model structures then selecting the one where error is minimized. For the work carried out in Australia, the selected model type used was the quantile regression forest model (QRF) which is a generalized implementation of the random forest model from Breiman (2001). The selection of QRF was based on its superiority compared with other candidate models including Cubist models, Random Forests and GAMS (results not shown). A full description and theoretical discussion of the QRF model can be found in Meinshausen (2006). In words, the *dissever* algorithm involves (mathematical notation of the algorithm is detailed in Malone et al., 2012):

Initialization Steps:

1. Create fine grid coarse resolution map (source map) via nearest neighbor resampling. Resolution and extent of the fine gridded map is the same as that of the available predictive covariates to be used for regression modeling.
2. Regress (using a selected model type) fine gridded values against suite of available covariates.
3. Upscale via averaging the fine gridded estimates to source map resolution.
4. Estimate deviation from mass balance for each coarse grid pixel i.e. mean of fine grid values = value from associated pixel of source map.

Iteration steps:

5. Correct fine gridded estimates with deviation factor from step 4 (or step 10 if iteration number is greater than 1).
6. Regress fine gridded values against suite of available covariates.
7. Upscale via averaging the fine gridded estimates to source map resolution.
8. Check whether upscaled estimates from step 7 are changed from previous iteration (or step 3 if performing iteration 1).
9. If estimated change (from step 8) is greater than some pre-defined threshold proceed to next step, otherwise STOP. In Malone et al. (2012) an averaged absolute difference between upscaled map from present iteration and previous iteration was used. An arbitrarily selected threshold of 0.001 was used to determine if iteration should proceed or not.
10. Estimate deviation from mass balance for each coarse grid pixel i.e. mean of fine grid values = value from associated pixel of source map. Go back to Iteration step 5.

For each simulated national map, *dissever* used the farm survey mapping as predictive covariates. One hundred simulations of the national mapping was performed which ultimately created 100 downscaled maps. All 100 downscaled maps were stacked together from which the mean and variance were then calculated. Added to the variance was a measure of the downscaling error in terms of the deviation from mass balance. This deviation measure is estimated for every pixel at each iteration of the *dissever* algorithm. Once *dissever* terminates we estimate the mean square error which in this case is:

$$MSE_{down} = \frac{1}{N}\sum_{i=1}^{N}(CV_i - FV_i)^2 \quad (2)$$

where N is the number of pixels of the coarse scaled map, and CV_i is the mapped value from the coarse map (national map) at pixel i, while FV_i is the average of the downscaled predictions from the last *dissever* iteration encapsulated by the same pixel i.

The workflow for deriving downscaled outputs in New Zealand entailed the implementation of the R package version of *dissever* (Roudier et al., 2017). Random Forest models were used in this instance. In order to quantify the uncertainties with the downscaling, bootstrapping (using 500 iterations) of the final dissever model was used to derive the upper and lower predictions, along with a mean prediction for each pixel. The workflow that was conducted in New Zealand was the original way that uncertainties were characterized prior to the formal simulation approach that was proposed in Malone et al. (2017) and used for the Australian sites in this study.

7. Using *Ospats* for stratification and sample selection

With farm scaled maps and associated uncertainties generated via spatial downscaling, implementation of the *Ospats* algorithm follows. The objective is to generate an optimal stratification of the target universe given those two required inputs. Summarizing from de Gruijter et al. (2016), given an initial stratification solution (which could be a compact geographical stratification, or k-means clustering of ancillary data etc.), through an iterative reallocation, grid points are assigned and re-assigned to strata of which there are a specified number. An objective function O is then estimated from this stratification. O is the model-based prediction of the sampling variance of the estimated mean, and needs to be minimized (via iteration). Pending further rounds of re-allocation and estimation of O, iteration stops when O is minimized.

Naturally, consideration in regards to the number of strata and number of samples is necessary. de Gruijter et al. (2016) describe the approaches for optimizing those parameters. They approach the issue using a Value of Information approach (Morgan et al., 1990) by comparing the difference in data value and data cost (which is equivalent to an expected financial gain) with different sample sizes and stratification configurations. The intention is to increase the expected financial gain, taking into account the costs associated with sampling, the expected reward for sequestering carbon, plus the expected sampling variance. From eq. 21 in de Gruijter et al. (2016) the optimal sample size of a farm can be estimated as:

$$n' = \left(\frac{CP \bullet A \bullet Z_y \bullet \overline{O}}{f \bullet \sqrt{2}} \right)^{2/3} \tag{3}$$

where n' is the optimal sample number, CP is the carbon offset price, or the amount in dollar terms, the price of sequestered carbon per mass unit (for demonstrative purposes we use A\$20 t^{-1}), A is the area of the farm in hectares, Z_y equals 1.645 (90% quantile of the standard normal distribution), $\overline{O}$ is the O/N (where O is the objective function from *Ospats* and N is the number of grid points discretising the farm), and f is the cost of sampling and analysis per point of the farm — for the Australian sites the value selected was \$120 given prior experience of sampling and its costs. Z_y is set to 1.645 because we wish to be able to determine with 90% confidence whether carbon stocks between the first and second sampling campaigns are different, when we take into consideration the sampling variance. For the first round of sampling i.e. the establishment of baseline SCS (as we have not sampled yet), we assume the expected sampling variance to equal $\overline{O}$. It is upon further resampling that we take into consideration the sampling variance of the previous sampling together with that from *Ospats* to determine the optimal sample number. In practice, selecting the optimal sample number is estimated for different strata number configurations. Following this we can then estimate the optimal number of strata. Again, following on from de Gruijter et al. (2016) this is estimated by evaluating the incremental financial gain estimated for each stratification compared to that found when not using any stratification. To determine this, one needs to use *Ospats* with no stratification, and then determine the optimal sample number using the equation above. The optimal strata number is the one where the incremental financial gain is greatest. From de Gruijter et al. (2016) the incremental financial gain is estimated as:

$$G_{inc}(H) = CP \bullet A \bullet Z_y \sqrt{2} \left(\overline{O}_1 \sqrt{\frac{1}{n_1'}} - \overline{O}_H \sqrt{\frac{1}{n_H'}} \right) + f(n_1' - n_H') \tag{4}$$

where $\overline{O}_H$ is the value of $\overline{O}$ for the *Ospats* stratification with H strata. Similarly $\overline{O}_1$ and n_1' are the $\overline{O}$ for the *Ospats* stratification and optimal sample number where stratification number equals 1.

After optimizing the strata and sample size for each site, we adjusted the actual sample number upwards slightly (in most cases) to ensure that at least four samples were collected from each strata (it is necessary to have at least two samples from each strata in order to estimate the within strata variance). Neyman allocation was used to assign the optimal number of samples to each stratum. This is effectively an allocation proportional to the product of stratum area and within stratum standard deviation.

In New Zealand, an operational constraint of the project meant that there were limited time and resources available to optimize soil sampling for SCS auditing. The sites used in New Zealand had existing data from recent surveys. This did not affect the integrity of the process as these existing sites were assigned to the derived *Ospats* strata.

8. Soil sampling and statistical inference

With the sampling configurations produced for each site, a soil sample was collected. At each site, duplicate samples were collected very close to each other. The first duplicate was used for estimation of soil carbon concentration, and the other duplicate was used for estimation of soil bulk density. Samples consisted of 0.3 m deep cores where possible (with known volume) which were extracted using a vehicle mounted hydraulic coring instrument in Australia. In addition to estimating SCS to 0.3 m, in Australia, SCS estimates were also derived for the 0–5 cm, 5–15 cm, 15–30 cm depth intervals. In New Zealand SCS estimates at NZCarb_S1 were derived for the 0–10 cm, 10–20 cm, and 20–30 cm depth intervals, while at NZCarb_S2 the depth intervals were 0–5 cm, 5-10 cm, 10–20 cm, and 20–30 cm.

Estimation of soil bulk density was made by dividing the oven dried (105 °C) soil mass of the sample (2nd duplicate sample) by the volume of the soil. Total soil carbon concentration of the samples (1st duplicate sample) was determined via dry combustion method using a vario MAX CNS analyser (Elementar, Germany) in Australia, and Leco CNS Analyser in New Zealand (© 2016 LECO Corporation, MI, USA). The analysed samples were a priori subjected to air drying before 2-mm sieving. Any rocks/gravel were removed and weighed during the sieving process so that carbon stocks could be adjusted for gravel content. Sieving was by mechanized mortar and pestle grinding for 2 minutes for Australian soils, and by mechanised 2-mm roller sieve for New Zealand soils. Approximately 750 mg of fine grinded soil was analysed for total carbon for all analysed samples in this project. Carbon stocks are given in terms of tonnes per hectare for specified depth. Carbon stock is estimated as:

$$\begin{aligned} \text{Carbon Stock (t ha}^{-1}) = {} & \text{Bulk Density (g cm}^{-3}) \\ & \times \text{Carbon Concentration (g kg}^{-1}) \\ & \times \text{depth of soil (cm)} \times (1 - \text{Gravel (g kg}^{-1})) \end{aligned} \tag{5}$$

For the Australian sites, we then applied the following depth interval weighted equation:

$$\begin{aligned} \text{Carbon Stock:0} - 0.3\text{ m} = {} & (\text{Carbon Stock:0–5 cm} \times 0.167) \\ & + (\text{Carbon Stock:5} - 15\text{ cm} \times 0.333) \\ & + (\text{Carbon Stock:15–30 cm} \times 0.500) \end{aligned} \tag{6}$$

Subsequently we can also provide estimates of proportion of carbon stocks for each depth relative to the whole auditing depth (0–0.3 m).

Parameters of the depth interval weighting in eq. 6 were adjusted accordingly for the NZ sites.

Estimation of total on-farm carbon stocks, together with associated sampling variance of the mean is derived via Stratified Simple Random Sampling design inference. From de Gruijter et al. (2006) the overall mean is equated as:

$$\hat{\bar{z}}_{St} = \sum_{h=1}^{H} a_h \bullet \hat{\bar{z}}_h \tag{7}$$

where $\hat{\bar{Z}}_{St}$ is the overall mean, H is the number of strata, $\hat{\bar{Z}}_h$ is the within stratum mean and a_h is the relative area of the stratum. The sampling variance of the estimated mean is:

$$\hat{V}(\hat{\bar{z}}_{St}) = \sum_{h=1}^{H} a_h^2 \bullet \hat{V}(\hat{\bar{z}}_h) \tag{8}$$

where $\hat{V}(\hat{\bar{Z}}_{St})$ is the sampling variance and $\hat{V}(\hat{\bar{Z}}_h)$ is the sampling variance of the within stratum mean, which is estimated as:

$$\hat{V}(\hat{\bar{z}}_h) = \frac{1}{n_h(n_h - 1)} \sum_{i=1}^{n_h} (z_{hi} - \hat{\bar{z}}_h)^2 \tag{9}$$

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

$$\hat{\bar{z}}_{St} \pm t_{1-\alpha/2} \bullet \sqrt{\hat{V}(\hat{\bar{z}}_{St})} \tag{10}$$

where $t_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of the Student distribution with $n - 1$ degrees of freedom where n is the number of collected samples from the site.

In order to assess whether the stratified sampling design used at each of the sites were deemed to be efficient, we used the Brus (1994) method that was designed for testing the stratification effect using the collected sample data. This entails estimating the spatial variance within a farm from the StSRS (Eq. 4, Brus, 1994). With this value, it is divided by the number of sampling locations (sample size) to obtain the sampling variance of the estimated mean for simple random sampling (i.e. no stratification). This sampling variance of the estimated mean divided by the sampling variance of StSRS, is the stratification effect. If this value is larger than 1, it means that StSRS is more efficient than simple random sample. Below zero means stratification was not effective and a simple random sample would have sufficed.

9. Results and discussion

9.1. Spatial downscaling of national digital soil maps

9.1.1. Australia

The SCS map from Viscarra Rossel et al. (2014) was clipped to the bounding extents for each landholding. Fig. 3A shows the national scale mapping with associated prediction standard deviation for AusCarb_S1. Fig. 3B shows the mean and standard deviation estimated by stacking the maps that were generated by Gaussian simulation of the national mapping for the same study area. Finally Fig. 3C shows the downscaled carbon stock map with associated standard deviation. These composite of digital soil maps have also been produced for the other AusCarb sites and are provided as Supplementary material 1. A visual analysis of the maps in Fig. 3, shows that the spatial pattern of the national mapping (both predictions and uncertainties) is adhered to via the Gaussian simulations. Table 1 summarizes the comparative results between mapping products in terms of the averages of the map predictions and the standard deviation. For example, for AusCarb_S1 the mean value from the national mapping predictions was 75.32 t ha^{-1} while for the mean of the Gaussian simulations the average was 75.01 t ha^{-1}. Note that these estimates are model based estimates of the mean, while estimates of the mean based on StSRS that follow further on are design based

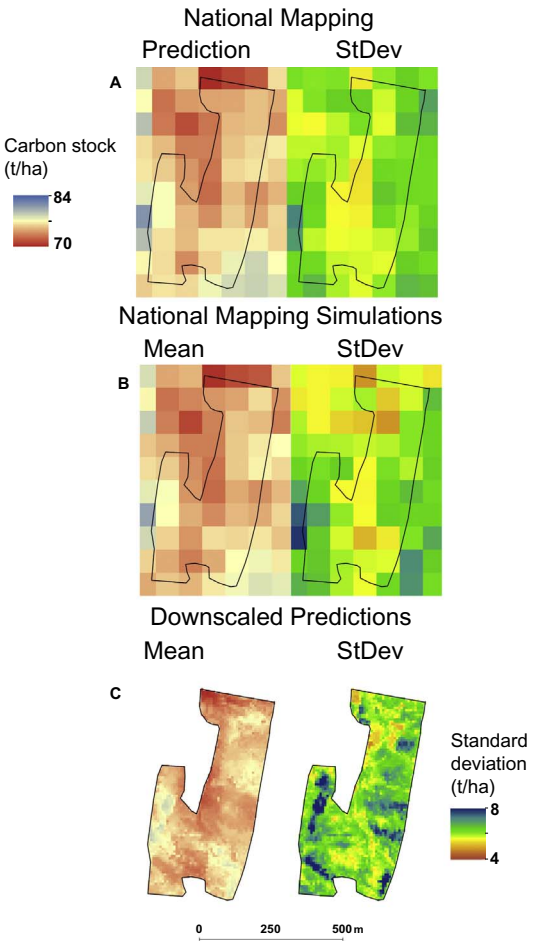

Fig. 3. Digital maps of soil carbon for the AusCarb_S1 site. Mapped predictions and standard deviation from Viscarra Rossel et al. (2014), which corresponds to the national scale mapping product (A). The mean and standard deviation of simulated realisations made from Gaussian simulations of the national scale mapping product (B). Downscaled soil carbon mapping showing the mean and standard deviation of predictions (C).

estimates. We calculated the model based means here purely for simple comparison between the national mapping products and associated downscaled outputs. In terms of the standard deviation maps, the means were 6.10 t ha^{-1} and 6.02 t ha^{-1} respectively. The concordance between both map products (predictions) was 0.95, while for the standard deviation maps, it was 0.65. This general result was observed for the other Australian study sites too. Ascertaining the reason why restoring the uncertainty estimates was not as closely adhered to when compared with the target variable predictions warrants further investigation. It is possible that the spatial relationship between the target variable and associated uncertainty is non-linear, meaning that the effectiveness of Gaussian simulation is slightly compromised. However, some initial investigative steps could first look at the possibility of generating significantly more simulations to better sample from the underlying uncertainty, or considering spatial models that incorporate anisotropic variation.

Table 1. Comparisons of means of carbon stocks for each site on the basis of the available national scale mapping, mean of the simulations derived from the national scale mapping (and their correspondence), and of the mapping outputs derived from spatial downscaling. The last column indicates the estimated deviation, in terms of a mean square error between simulated national scale mapping outputs and the downscaled maps derived from them. Bracketed numbers for Australian sites are standard deviations while NZ sites report a mean with upper and lower 95% confidence level; na: not applicable.

In terms of the digital map derived from spatial downscaling there is

Table 1
Comparisons of means of carbon stocks for each site on the basis of the available national scale mapping, mean of the simulations derived from the national scale mapping (and their correspondence), and of the mapping outputs derived from spatial downscaling. The last column indicates the estimated deviation, in terms of a mean square error between simulated national scale mapping outputs and the downscaled maps derived from them. Bracketed numbers for Australian sites are standard deviations while NZ sites report a mean with upper and lower 95% confidence level; na: not applicable.

Site	A. National mapping	B. National MAPPING simulations	Concordance between A and B	C. Downscaled mapping	Downscaling deviation MSE
AusCarb_S1	75.32 (6.10)	75.01 (6.02)	0.95 (0.65)	74.95 (6.38)	2.76
AusCarb_S2	75.01 (5.86)	74.60 (5.77)	0.94 (0.71)	74.51 (6.24)	4.24
AusCarb_S3	73.77 (5.58)	73.48 (5.64)	0.98 (0.73)	73.97 (6.36)	1.88
AusCarb_S4	75.02 (5.81)	74.71 (5.88)	0.94 (0.65)	74.44 (5.96)	1.17
NZCarb_S1	77.40 (63.47–91.34)	na	na	64.40 (61.68–67.12)	na
NZCarb_S2	93.77 (54.40–133.15)	na	na	93.26 (84.19–102.32)	na

quite strong correspondence relative to the national mapping. Obviously the downscaled mapping shows much more fine-grained variation, which is attributed to the covariates that were used in the *dissever* algorithm. Results summarized in Table 1 confirm this correspondence where the mean of the downscaled map predictions are very close to that for the national mapping at each Australian site. The relatively higher uncertainty estimate associated with the downscaling is attributed to the incorporation of the downscaling uncertainty which was quantified as the mean square error (MSE) of deviation between national mapping and downscaled predictions. It is intuitive to expect higher uncertainties on average in this regard because predictions are more localized compared to the coarser grid of the national mapping. Furthermore, there has not been any additional observed data to support the predictions at the finer scale, which as a consequence the predictions at the fine scale will inevitably be more uncertain. On average, over the 100 downscaling realisations, the deviation MSE ranged between 1.17 (AusCarb_S4) and 4.24 (AusCarb_S2). The relative difference can be attributed to the differentiated relationship of the target variable (carbon stock) to the predictive covariates between sites. For AusCarb_S4 the considered covariates appear to be adequate to the task, but for AusCarb_S2 we would perhaps need to consider further covariate data to improve the downscaling procedure. As the land use and management appears to be homogenous across this site, a land use map alone would not be helpful. Related information in terms of vegetation indices such as plant cell density estimates or time series estimates of NDVI could perhaps be useful for further consideration.

9.1.2. New Zealand

The NZAGRC national carbon map (NZAGRC, 2016) was clipped to the bounding extents for each landholding. Fig. 4A shows the national scale mapping with associated prediction standard deviation for NZCarb_S2. Fig. 4B shows the downscaled carbon stock map with associated standard deviation. Similar maps were also produced for NZCarb_S1 and are provided in Supplementary Material 1.

What can be observed on Fig. 4 is a good correspondence between the national mapping and downscaled mapping in terms of the mean SCS predictions. On the other hand, the spatial variation of the uncertainty from the national map is not retained through to the downscaled estimated of the uncertainty — it appears the uncertainty is relatively lower for the downscaled map compared to the national map. This is attributed to the method used for downscaling the national maps, where the approach used by Malone et al. (2017) propagates the uncertainty via simulation through to the downscaled maps. Using the *dissever* R package directly as was done for the New Zealand sites, the uncertainties are estimated via a bootstrapping procedure that considers only the target variable data and not the associated uncertainty of those predictions. In summary, estimation of the uncertainties of the downscaled mapping can be achieved via the approach of Malone et al. (2017) or using the *dissever* R package (Roudier et al., 2017). However, the approach of Malone et al. (2017) is recommended as it explicitly deals with the uncertainty of the map that is to be downscaled.

9.1.3. Ospats

At each site, the downscaled maps were first processed using aerial imagery to clip out areas where soil sampling could not possibly be performed. For example at AusCarb_S2, a creek runs the north-south direction through the landholding. We subsequently focused only on the areas dedicated to vineyard production. Road obstacles and water catchment areas were similarly excluded from the auditing area at AusCarb_S4 and other sites. A similar process was followed for NZ sites, clipping the road, housing and waterways from the downscaled maps. For NZCarb_S2 in particular, accessibility due to difficult terrain meant to only a sub-catchment of the property could be audited at the time of the study.

For *Ospats*, to optimize the strata number for each site we considered outcomes for between 2 and 10 strata. Optimal sample number was estimated using eq. 4. Table 2 summarizes the outcomes from running *Ospats* for each site. Two strata were optimal for AusCarb_S2, three strata were optimal for AusCarb_S1, AusCarb_S3, and AusCarb_S4, and four strata were optimal for NZCarb_S1 and NZCarb_S2. To ensure at least 4 samples be collected from each strata, we increased the total number of samples slightly to meet this criteria when required. In total we selected, 16, 10, 20, 30, 50 and 100 samples from AusCarb_S1, S2, S3, S4 and NZCarb_S1 and S2 respectively. Fig. 5 shows the stratifications made for each site, upon which the sampling locations are overlaid.

Table 2. Strata configurations, recommended sample sizes, and sample allocations for each site. Note that New Zealand sites did not implement steps for optimal stratification and sample number. For all sites, Neyman allocation was used to determine the number of samples to be taken from each stratum.

9.1.4. Statistical inference

Soil carbon stocks for each of the sites are summarized on Tables 3 and 4 for Australia and New Zealand respectively. In Australia, in terms of the 0–0.30 m depth, the highest stocks were attributed to AusCarb_S3 with 25.93 t ha^{-1} of carbon. However, relative to the other sites, there was also the highest amount of uncertainty associated with this mean carbon stock estimate, where it ranged between 16.58 t ha^{-1} and 35.28 t ha^{-1}. One contributor for this relatively higher amount of uncertainty is that there is a significant amount of inorganic carbon present around this site by way of the calcium carbonate rich marl parent materials. The presence of marl was not observed for any of the other Australian sites. At AusCarb_S3, the marl does not occur homogenously across the site spatially, or even at a specific depth. Rather it occurs in seemingly discrete parcels whose spatial pattern is difficult to determine, and it can occur in any depth of the profile.

Table 3. Estimated soil carbon stocks and upper and lower confidence bounds for each Australian site for 0–0.3 m and specified depth intervals. A proportion of the total carbon stock is provided for each depth interval. Stratification effect is given for the 0–0.3 m depth interval,

The relatively higher uncertainty about the mean carbon stocks at AusCarb_S3 seems to be an exception when compared with the other

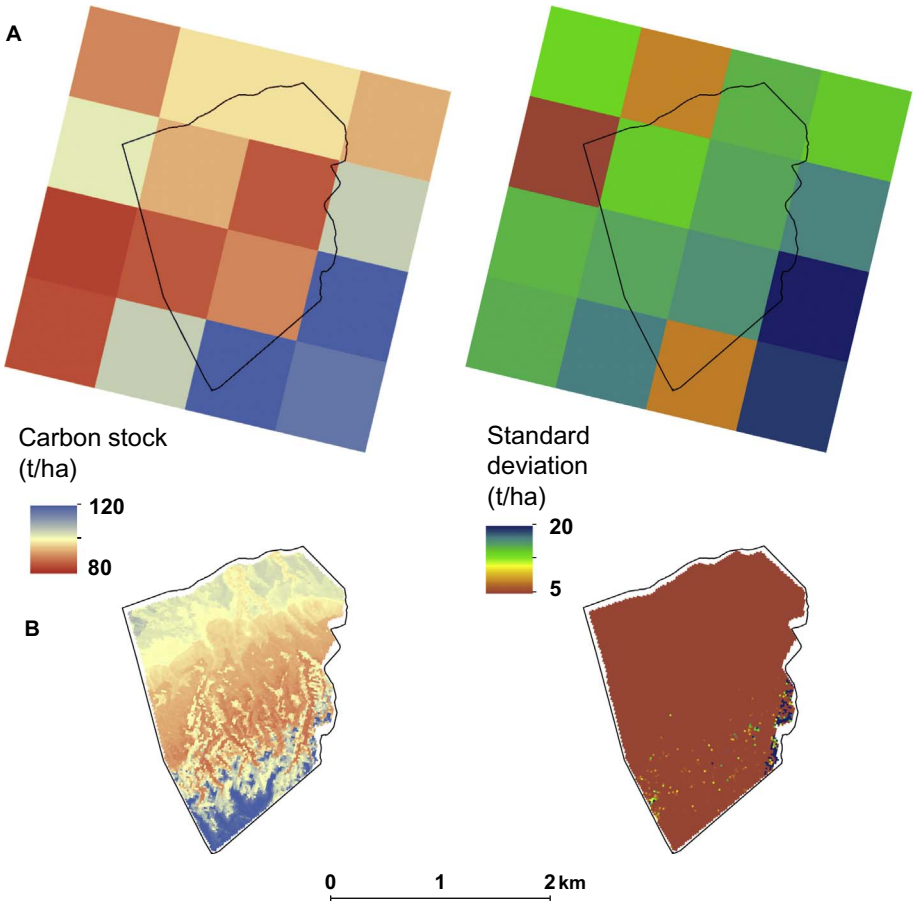

Fig. 4. Digital maps of soil carbon for the NZCarb_S2 site. Mapped predictions and standard deviation from NZAGRC (2016), which corresponds to the national scale mapping product (A). Downscaled soil carbon mapping showing the mean and standard deviation of predictions (B).

Table 2
Strata configurations, recommended sample sizes, and sample allocations for each site. Note that New Zealand sites did not implement steps for optimal stratification and sample number. For all sites, Neyman allocation was used to determine the number of samples to be taken from each stratum.

Site	Optimal strata number	Optimal sample size	Actual sample allocation	Actual total sample size
AusCarb_S1	3	16	6, 6, 4	16
AusCarb_S2	2	6	6, 4	10
AusCarb_S3	3	27	12, 11, 7	30
AusCarb_S4	3	17	10, 6, 4	20
NZCarb_S1	4	na	25, 25, 25, 25	100
NZCarb_S2	4	na	13,8,19,10	50

Australian sites. The carbon stocks at the other sites are relatively lower, but the confidence intervals are much narrower, indicating more certainty about the means. AusCarb_S2 has less than half the stocks than that at the nearby AusCarb_S1. This is likely due to distinctive soil types between these two sites where soils are predominantly clayey at AusCarb_S1 compared to AusCarb_S2 which are predominantly fine-sandy clays. Consequently the relative capacity of these different soils to store carbon will also be quite different. Vineyard management is also likely to play a significant role in the amount of carbon stocks. It is known that the between vine management at AusCarb_S1 involves no tillage and maintenance of perennial pastures, While at AusCarb_S2, tillage twice a year is common practice. But because these two sites have distinctive soils, the scope of this research does not permit an analysis of the relative differences in SCS under different management regimes. A basic analysis though between AusCarb_S1 and AusCarb_S4 – which do have similar soil types, but quite different vineyard management practices shows quite distinct differences in carbon stocks when we consider the depth variable. AusCarb_S4 has a long history of annually till between vine row spaces. At this site, we have estimated that 53% of the carbon stocks are found between 15 and 30 cm, and 11% in the top 5 cm. At AusCarb_S1 we estimated 40% and 23% respectively for the same two depth intervals. Note that these percentage estimates are based on the mean stocks and do not consider the uncertainty about the means.

Comparison between SCS at each of the Australian sites as determined by spatial downscaling with those stocks estimated from samples collected (by StSRS) in the field, show some significant differences. The impression from the downscaled outputs is that the carbon stocks at each site are relatively the same (See Table 1). From the soil sampling, it is evident that this situation is not true. From this comparison, obviously, the national mapping for this area (the Lower Hunter Valley in general) is biased positively in a large way, improperly inflating the estimated carbon stocks. Unfortunately without any prior data, the spatial downscaling methodology that was used in this study does not correct for the bias; rather the objective of the downscaling is to achieve mass balance between coarse-grained and fine-grained mapping. The implication of this outcome is that the stratification is performed on data that do not represent the site-specific conditions of soil carbon variation. This will inevitably result in sampling inefficiencies. Indeed, Table 3 indicates that for the Australian sites, StSRS is not as efficient compared to a simple random sample.

Workflows for incorporating a bias correction step in the spatial downscaling procedure has been show previously for soil carbon mapping by Malone et al. (2017) where field observations of carbon concentration were incorporated into the dissever algorithm. This

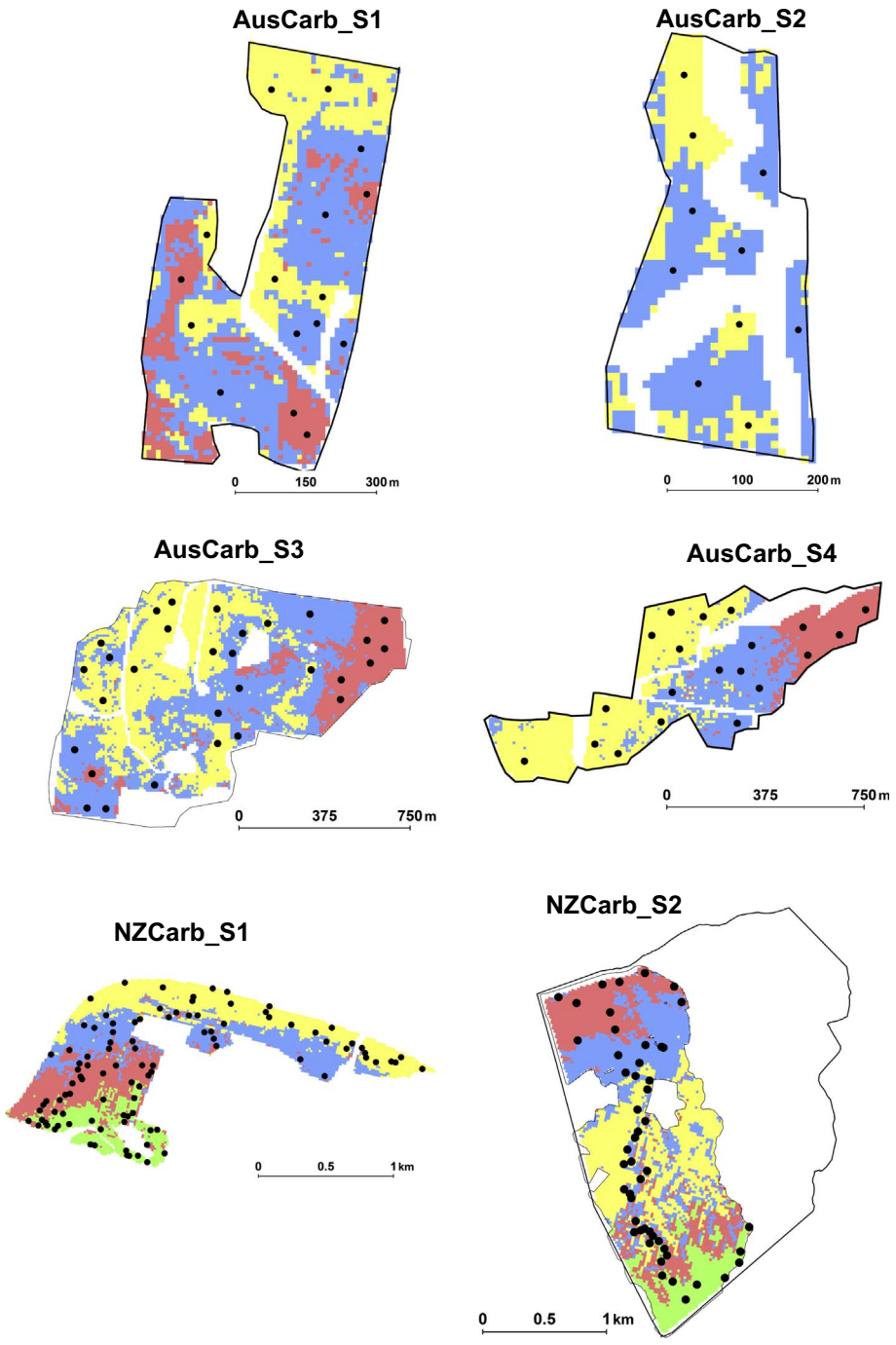

Fig. 5. Stratifications and sample configurations for each Australian and New Zealand site. Color represents stratum. See Table 2 for details.

evidently corrected the substantial bias that was apparent in the national mapping for the area that was investigated the north-western NSW, Australia. Similar bias correction procedures for downscaling have been demonstrated in other studies, particularly those concerned with climatic data applications e.g. Déqué (2007), Wilcke et al. (2013). One limitation of using this additional bias correction step is that there needs to be an availability of existing observations to adjust the downscaled outputs, which was not the case for the Australian sites.

At each NZ site, a comparison between SCS determined by downscaling with those estimated by collecting field samples (by StSRS) show a good match (Table 1 and Table 4) for the 0.3 m soil depth. This is explained by the fact that covariates used to develop the national model and drive the disaggregation process relate quite well to processes influencing the accumulation of carbon in surface soil horizons. Stratification was clearly warranted at the NZ sites as indicated by the stratification effect values shown on Table 4 all being greater than 1. While we are confident in this result, note that due to operational circumstances, the samples were not selected with probability sampling. Rather, pre-existing samples (collected via simple random sampling) were allocated to the *Ospats*-derive strata. Consequently, in strict adherence to sampling theory, the statistical parameters derived from StSRS inference are potentially biased. We doubt this to be significant due to the density of sampling that was carried out at each of the two sites however.

Table 4. Estimated soil carbon stocks and upper and lower confidence bounds for each New Zealand site for 0–0.3 m and specified

Table 3
Estimated soil carbon stocks and upper and lower confidence bounds for each Australian site for 0–0.3 m and specified depth intervals. A proportion of the total carbon stock is provided for each depth interval.

Site	Depth (m)	Soil carbon stocks (t/ha)	Upper stocks (t/ha)	Lower stocks (t/ha)	Percentage of total carbon stocks	Stratification effect
AusCarb_S1	0–0.30	22.42	27.64	17.20	100	0.7
	0–0.05	30.43	39.57	21.28	23	
	0.05–0.15	24.88	31.06	18.70	37	
	0.15–0.30	18.10	22.80	13.41	40	
AusCarb_S2	0–0.30	9.93	14.56	5.30	100	0.9
	0–0.05	9.24	13.43	5.05	16	
	0.05–0.15	11.66	17.69	5.63	39	
	0.15–0.30	9.01	13.25	4.77	45	
AusCarb_S3	0–0.30	25.93	35.28	16.58	100	0.6
	0–0.05	18.81	23.60	14.01	12	
	0.05–0.15	25.14	32.74	17.55	32	
	0.15–0.30	24.64	42.61	15.05	56	
AusCarb_S4	0–0.30	12.77	14.64	10.90	100	0.6
	0–0.05	8.57	11.03	6.10	11	
	0.05–0.15	13.82	17.94	9.70	36	
	0.15–0.30	13.48	16.88	10.08	53	

depth intervals. A proportion of the total carbon stock is provided for each depth interval.

10. General discussion

Public availability of soil carbon mapping via national carbon models means that most people are able access information specific to their own landholding. This study has tested a protocol for deriving on-farm estimates of soil carbon stocks using a combination of spatial downscaling and optimal stratification. A generalized protocol flowchart is provided in Fig. 1. Here, this includes the acquisition of national scale soil carbon mapping, and relatedly the acquisition of relevant farm scale environmental data. This is followed by implementation of spatial downscaling where the objective is to derive relevant farm scale soil carbon mapping with associated uncertainty quantification. *Ospats* is then used to derive the sampling strata, followed by the associated field sampling then laboratory analysis and statistical inference to derive a baseline audit of the on- farm soil carbon stocks.

A rather major issue encountered for the Australian component of this study was the use of downscaled national mapping being positively biased and ultimately inaccurate. This led to an inappropriate optimized stratification and sampling recommendations. The result being that the stratification effect contributed to the overall uncertainty about the mean estimates of SCS at each location. Despite promising results for the tested protocol in NZ, some further consideration is warranted for improving the sampling efficiency in the baseline establishment round of sampling. Perhaps the initial sampling should be optimized based on available financial resources, or aim for geographical coverage or spatially balanced sampling (Brus et al., 1999; Stevens and Olsen, 2004; Brus, 2015).

Some consideration needs to be made with respect to the auditing process once current baseline carbon stocks have been established on the farm. As indicated in de Gruijter et al. (2015) an updated farm-scale map is used for input in *Ospats*. If digital soil mapping is used, this step simply entails incorporating the samples collected from the first round of sampling with any other existing samples; after which the new digital soil map is produced. However, even with these additional samples, the density of the sampling may still be impractical for deriving an acceptable digital soil map to use in the second round of sampling. Therefore we propose a number of possible routes to follow if digital soil mapping alone is not going to be suitable. The selected option is determined on the basis of the quality of available information.

1. Perform spatial downscaling as was done for the first round. It is inevitable that with national soil mapping being digitally based, the underlying models will be dynamic in the sense that they can be updated pending the acquisition of new data. This will achieve the result of a new realisation of the national mapping and associated prediction variance which may or may not be improved. It is likely in the future that nationally directed efforts for updating and improving soil mapping will occur as a globally collective response to addressing environmental issue and land resource monitoring. With this option, some consideration of the feasibility of using *Ospats* needs to be made particularly if it is thought the national mapping is going to be unreliable.
2. Perform downscaling together with incorporation of observed data. This is akin to data assimilation, but more specifically as bias correction (Poggio and Gimona, 2015). The work performed by Malone et al. (2017) demonstrates how to implement this in particular for soil carbon mapping.
3. Use a model averaging approach to fuse together downscaled outputs with digital soil mapping outputs. Model averaging is a useful approach if one has multiple outcomes from different models and

Table 4
Estimated soil carbon stocks and upper and lower confidence bounds for each New Zealand site for 0–0.3 m and specified depth intervals. A proportion of the total carbon stock is provided for each depth interval.

Site	Depth (m)	Soil carbon stocks (t/ha)	Upper stocks (t/ha)	Lower stocks (t/ha)	Percentage of total carbon stocks	Stratification effect
NZCarb_S1	0.0–0.3	65.30	68.26	62.33	100	1.2
	0.0–0.1	29.75	31.45	28.04	46	
	0.1–0.2	22.72	24.06	21.37	35	
	0.2–0.3	12.83	13.87	11.79	20	
NZCarb_S2	0–0.3	101.88	111.57	92.29	100	1.1
	0–0.05	24.51	26.11	22.87	24	
	0.05–0.1	22.31	24.66	19.96	22	
	0.1–0.2	33.76	37.37	29.77	33	
	0.2–0.3	22.75	24.76	18.32	22	

wants some way of consolidating them into one output (Diks and Vrugt, 2010). Most approaches preferentially weight the outcomes based on accuracy such that the final outcome is one that is at least as good as or better than the best individual model.

To assess what the best option is, one could follow each of the 3 steps and select the one which achieves the lowest prediction variance. Ultimately the third step will probably be the best option as the mapping will be a collective of the models derived from both national and local sources.

Acquisition of high resolution ancillary data is a necessary requirement for driving the downscaling towards the fine resolution mapping of soil carbon. The availability of these resources will be variable. Many landholdings where management is based on precision technologies, will likely have some or a combination of high resolution data such as yield mapping, very high resolution elevation data, as well as radiometric and EMI data. At the other end of the spectrum such information will not be immediately available. This was the case for the investigated sites in the Australia where we needed to collect the information ourselves. While high resolution elevation data is available for this area in general (25 m resolution), other key variables such as gamma radiometric data is poorly resolved. This situation will inevitably change in the future as outputs from new sensing technologies coupled with finer resolutions will become available for public use. Using some examples is the availability of sub-metre LIDAR data that is available for much of the coverage of New Zealand (https://data.govt.nz/dataset/show/4709). A publically available 10 m resolution DEM is available for the conterminous USA (http://nationalmap.gov/elevation.html). Similar developments in large extent earth resource data acquisition will or are inevitably happening in Australia to date.

In terms of SCS, there will always be a need to investigate new technologies that will make SCS measurement more efficient and cheaper. Direct measurement of soil carbon in the field via inelastic neutron scattering (Yakubova et al., 2016) shows particular promise in terms of efficiency. Infrared technology is a natural candidate for that purpose too (Bellon-Maurel and McBratney, 2011). A further advancement to this is to bring the technology into the field and perform the measurement in real time in the field or in situ (Roudier et al., 2015). Such technology will enhance the ability to collect more data. Yet this needs to be balanced by taking into account the expected higher measurement error associated with using infrared measurements. Synergising *Ospats* with infrared carbon measurement is likely to be an avenue of further research.

Further in regards to measurement is that of bulk density, the necessary input for SCS. It is a soil property that is difficult to measure, because it is quite labour intensive. Measurement of bulk density at soil depth also presents logistical problems when there is no desire to excavate soil pits to gain physical access to the soil beyond 0.3 m. In fact, soil coring as was used in this research is likely to be the mainstay for soil carbon auditing research because it allows an efficient means to retrieve soil sample at depth. Nevertheless, how to obtain bulk density measurements from the core samples has and is likely to attract substantial research interest. The approach used in Australia was one approach, but could still be seen as labour intensive because substantial field and laboratory effort is required to extract duplicate soil cores, estimate soil wetness, and weigh out samples etc. Recently, Lobsey and Viscarra Rossel (2016) have developed a sensing method based on gamma-ray attenuation and vis-NIR spectroscopy. It was shown by comparison that this approach is far more efficient than traditional approaches to estimating bulk density, and can be used in the field too. A step further is a method proposed by Pallasser et al. (2015) where measurement of SCS of a whole soil core can be measured in entirety and in one step i.e. no need to estimate bulk density separately. The method of analysis is based on combustion, where a large aliquot of soil is combusted, and a detector evaluates how much CO_2 is evolved as a result. The approach is advantageous because large sample volumes (500–600 g) can be used, as opposed to very small aliquots (e.g. 500 mg) that are more traditionally used with the instrumentation currently available. Measurement of carbon from volume of this size reduces bias and error associated with using very small aliquots. Currently the custom made measurement instrument – soil carbon bench – is situated in a lab. It will have far greater utility if it were taken into the field.

11. Conclusions

Research undertaken in Australia and New Zealand has tested and helped to develop a possible soil carbon auditing method appropriate for wide farm-scale application. We expect on-farm soil carbon auditing to play a significant role in the future of agriculture to meet:

(i) international monitoring commitments of greenhouse gas emissions, and
(ii) regulations placed on landowners to sustain the soil resource.

The method takes advantage of existing national and local environmental datasets that help to define the likely variability of soil carbon on a farm, so that the area can be stratified for on-going monitoring. The method we have proposed would also be applicable irrespective of the spatial extent of the mapping to be downscaled. In this study we have investigated national mapping extents. But if the spatial resolution of the source map (where the spatial extent could be state-level, regional, or even smaller) is seemingly too coarse i.e. pixels are too large, for on-farm application, then our proposed downscaling, then sampling method would be equally applicable.

However, while the proposed approach provides a statistically robust and transparent method for estimating on-farm soil carbon stocks, an appraisal of the reliability of the national mapping to be used for spatial downscaling needs to be performed initially. Unreliable mapping without any site-specific information to correct it will ultimately lead to relatively imprecise soil carbon stock estimates when establishing a baseline. After a reliable baseline has been established, our approach will be beneficial in subsequent sampling auditing campaigns.

Acknowledgements

In New Zealand, funding support was provided by the New Zealand Government to support the objectives of the Livestock Research Group of the Global Research Alliance on Agricultural Greenhouse Gases. The Australian-based researchers acknowledge the support of the Australian Government Department of Agriculture and Water Resources (FTRG 1194105-66) via the Filling the Research Gap Round 2 soil carbon research funding scheme project entitled “Farm scale assessment of SOC from disaggregated national/regional scale models”.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.geodrs.2018.02.002.

References

Arrouays, D., Marchant, B.P., Saby, N.P.A., Meersmans, J., Jolivet, C., Orton, T.G., Martin, M.P., Bellamy, P.H., Lark, R.M., Louis, B.P., Allard, D., Kibblewhite, M., 2014a. On soil carbon monitoring networks. In: Hartemink, A.E., McSweeney, K. (Eds.), Soil Carbon. Springer International Publishing, Cham, pp. 59–68.
de Forges, A.R., 2014b. In: Arrouays, D., McKenzie, N.J., Hempel, J., McBratney, A.B. (Eds.), GlobalSoilMap: Basis of the Global Spatial Soil Information System. CRC Press, London.
Bellon-Maurel, V., McBratney, A., 2011. Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils – critical review and research perspectives. Soil Biol. Biochem. 43, 1398–1410.
Breiman, L., 2001. Random forests. Mach. Learn. 41, 5–32.
Brus, D.J., 1994. Improving design-based estimation of spatial means by soil map

stratification. A case study of phosphate saturation. Geoderma 62, 233–246.
Brus, D.J., 2015. Balanced sampling: a versatile sampling approach for statistical soil surveys. Geoderma 253, 111–121.
Brus, D.J., Spätjens, L.E.E.M., de Gruijter, J.J., 1999. A sampling scheme for estimating the mean extractable phosphorus concentration of fields for environmental regulation. Geoderma 89, 129–148.
Bunting, P., Armston, J., Clewley, D., Lucas, R.M., 2013. Sorted pulse data (SPD) library—part II: a processing framework for LiDAR data from pulsed laser systems in terrestrial environments. Comput. Geosci. 56, 207–215.
Chiles, J., Delfiner, P., 1999. Geostatistics: Modeling Spatial Uncertainty. Wiley, New York.
Cook, S., Corner, R., Groves, P., Grealish, G., 1996. Use of airborne gamma radiometric data for soil mapping. Soil Res. 34 (1), 183–194.
de Gruijter, J.J., Minasny, B., McBratney, A.B., 2015. Optimizing stratification and allocation for design-based estimation of spatial means using predictions with error. J Surv Stat Methodol 3 (1), 19–42.
de Gruijter, J.J., McBratney, A.B., Minasny, B., Wheeler, I., Malone, B.P., Stockmann, U., 2016. Farm-scale soil carbon auditing. Geoderma 265, 120–130.
Déqué, M., 2007. Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: model results and statistical correction according to observed values. Glob. Planet. Chang. 57 (1–2), 16–26.
Diks, C.G.H., Vrugt, J.A., 2010. Comparison of point forecast accuracy of model averaging methods in hydrologic applications. Stoch. Env. Res. Risk A. 24 (6), 809–820.
FAO, 1998. World Reference Base for Soil Resources. Food and Agriculture Organization of the United Nations, Rome.
Friedlingstein, P., Meinshausen, M., Arora, V.K., Jones, C.D., Anav, A., Liddicoat, S.K., Knutti, R., 2014. Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks. J. Clim. 27 (2), 511–526.
Goidts, E., Van Wesemael, B., Crucifix, M., 2009. Magnitude and sources of uncertainties in soil organic carbon (SOC) stock assessments at various scales. Eur. J. Soil Sci. 60, 723–739.
de Gruijter, J.J., Brus, D.J., Bierkens, M.F.P., Knotters, M., 2006. Sampling for Natural Resource Monitoring. Springer-Verlag, Berlin Heidelberg, The Netherlands.
Hedley, C.B., Yule, I.J., Eastwood, C.R., Shepherd, T.G., Arnold, G., 2004. Rapid identification of soil textural and management zones using electromagnetic induction sensing of soils. Australian J Soil Res 42 (4), 389–400.
Hedley, C.B., Payton, I.J., Lynn, I.H., Carrick, S.T., Webb, T.H., McNeill, S., 2012. Random sampling of stony and non-stony soils for testing a national soil carbon monitoring system. Soil Res. 50, 18–29.
Hedley, C.B, Roudier Peterson P, P., 2016. Gamma soil surveys – investigating soil patterns for nutrient and water management. In: Currie, L.D., Singh, R. (Eds.), Integrated nutrient and water management for sustainable farming. Occasional Report No. 29 Fertilizer and Lime Research Centre, Massey University, Massey University, Palmerston North, New Zealand 8 pages. http://flrc.massey.ac.nz/publications.html.
Hewitt, A.E., 2010. In: Lincoln (Ed.), New Zealand Soil Classification, 3rd ed. Manaaki Whenua Press, New Zealand 136p.
Jobbágy, E.G., Jackson, R.B., 2000. The vertical distribution of soil organic carbon and its relation to climate and vegetation. Ecol. Appl. 10 (2), 423–436.
Lobsey, C.R., Viscarra Rossel, R.A., 2016. Sensing of soil bulk density for more accurate carbon accounting. Eur. J. Soil Sci. 67 (4), 504–513.
Malone, B.P., McBratney, A.B., Minasny, B., Wheeler, I., 2012. A general method for downscaling earth resource information. Comput. Geosci. 41, 119–125.
Malone, B.P., Hughes, P., McBratney, A.B., Minasny, B., 2014. A model for the identification of terrons in the lower Hunter Valley, Australia. Geoderma Reg 1, 31–47.
Malone, B.P., Styc, Q., Minasny, B., McBratney, A.B., 2017. Digital soil mapping of soil carbon at the farm scale: a spatial downscaling approach in consideration of measured and uncertain data. Geoderma 290, 91–99.
McNeill, S.J.E., Golubiewski, N., Barringer, J., 2014. Development and calibration of a soil carbon inventory model for New Zealand. Soil Res. 52 (8), 789–804.
Meinshausen, N., 2006. Quantile regression forests. J. Mach. Learn. Res. 7, 983–999.
Minasny, B., Malone, B.P., McBratney, A.B., Angers, D.A., Arrouays, D., Chambers, A., Chaplot, V., Chen, Z.-S., Cheng, K., Das, B.S., Field, D.J., Gimona, A., Hedley, C.B., Hong, S.Y., Mandal, B., Marchant, B.P., Martin, M., McConkey, B.G., Mulder, V.L., O'Rourke, S., Richer-de-Forges, A.C., Odeh, I., Padarian, J., Paustian, K., Pan, G., Poggio, L., Savin, I., Stolbovoy, V., Stockmann, U., Sulaeman, Y., Tsui, C.-C., Vågen, T.-G., van Wesemael, B., Winowiecki, L., 2017. Soil carbon 4 per mille. Geoderma 292, 59–86.
Minty, B.R.S., McFadden, P., Kennett, B.L.N., 1998. Multichannel processing for airborne gamma-ray spectrometry. Geophysics 63 (6), 1971–1985.
Morgan, M.G., Henrion, M., Small, M., 1990. Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. Cambridge University Press, USA.
NZAGRC (New Zealand Agricultural Greenhouse Gas Research Centre), 2016. Soil Organic Carbon Stocks Model for New Zealand, Developed by Stephen McNeill, Landcare Research - Manaaki Whenua, under Contract to NZAGRC. (Access by permission from the New Zealand Agricultural Greenhouse Gas Research Centre via http://www.nzagrc. org.nz/).
Orton, T.G., Pringle, M.J., Bishop, T.F.A., 2016. A one-step approach for modelling and mapping soil properties based on profile data sampled over varying depth intervals. Geoderma 262, 174–186.
Pallasser, R., Minasny, B., McBratney, A.B., 2015. Carbon determination system for whole soil cores. Commun. Soil Sci. Plant Anal. 46 (2), 221–234.
Paustian, K., Andrén, O., Janzen, H.H., Lal, R., Smith, P., Tian, G., Tiessen, H., Van Noordwijk, M., Woomer, P.L., 1997. Agricultural soils as a sink to mitigate CO2 emissions. Soil Use Manag. 13, 230–244.
Poggio, L., Gimona, A., 2015. Downscaling and correction of regional climate models outputs with a hybrid geostatistical approach. Spatial Stat. 14, 4–21.
Rawlins, B.G., Marchant, B.P., Smyth, D., Scheib, C., Lark, R.M., Jordan, C., 2009. Airborne radiometric survey data and a DTM as covariates for regional scale mapping of soil organic carbon across Northern Ireland. Eur. J. Soil Sci. 60 (1), 44–54.
Roudier, P., Hedley, C.B., Ross, C.W., 2015. Prediction of volumetric soil organic carbon from field-moist intact soil cores. Eur. J. Soil Sci. 66 (4), 651–660.
Roudier, P., Malone, B.P., Hedley, C.B., Minasny, B., McBratney, A.B., 2017. Comparison of regression methods for spatial downscaling of soil organic carbon stocks maps. Comput. Electron. Agric. 142, 91–100.
Saby, N.P.A., Bellamy, P.H., Morvan, X., Arrouays, D., Jones, R.J.A., Verheijen, F.G.A., Kibblewhite, M.G., Verdoodt, A.N.N., ÜVeges, J.B., Freudenschuß, A., Simota, C., 2008. Will European soil-monitoring networks be able to detect changes in topsoil organic carbon content? Glob. Chang. Biol. 14 (10), 2432–2442.
Sanchez, P.A., Ahamed, S., Carre, F., Hartemink, A.E., Hempel, J., Huising, J., Lagacherie, P., McBratney, A.B., McKenzie, N.J., Mendonca-Santos, M.D., Minasny, B., Montanarella, L., Okoth, P., Palm, C.A., Sachs, J.D., Shepherd, K.D., Vagen, T.G., Vanlauwe, B., Walsh, M.G., Winowiecki, L.A., Zhang, G.L., 2009. Digital soil map of the world. Science 325 (5941), 680–681.
Sanderman, J., Baldock, J.A., 2010. Accounting for soil carbon sequestration in national inventories: a soil scientist's perspective. Environ. Res. Lett. 5, 034003.
Spencer, S., Ogle, S.M., Breidt, F.J., Goebel, J.J., Paustian, K., 2011. Designing a national soil carbon monitoring network to support climate change policy: a case example for US agricultural lands. Greenhouse Gas Measurement and Manag. 1 (3–4), 167–178.
Stevens, D.L., Olsen, A.R., 2004. Spatially balanced sampling of natural resources. J. Am. Stat. Assoc. 99, 262–278.
Stockmann, U., Adams, M.A., Crawford, J.W., Field, D.J., Henakaarchchi, N., Jenkins, M., Minasny, B., McBratney, A.B., Courcelles, V.d.R.d., Singh, K., Wheeler, I., Abbott, L., Angers, D.A., Baldock, J., Bird, M., Brookes, P.C., Chenu, C., Jastrow, J.D., Lal, R., Lehmann, J., O'Donnell, A.G., Parton, W.J., Whitehead, D., Zimmermann, M., 2013. The knowns, known unknowns and unknowns of sequestration of soil organic carbon. Agric. Ecosyst. Environ. 164, 80–99.
Viscarra Rossel, R.A., Webster, R., Bui, E.N., Baldock, J.A., 2014. Baseline map of organic carbon in Australian soil to support national carbon accounting and monitoring under climate change. Glob. Chang. Biol. 20 (9), 2953–2970.
Wilcke, R.A.I., Mendlik, T., Gobiet, A., 2013. Multi-variable error correction of regional climate models. Clim. Chang. 120 (4), 871–887.
Wood, A.T.A., Chan, G., 1994. Simulation of stationary Gaussian processes in [0,1]. J. Comput. Graph. Stat. 3 (4), 409–432.
Yakubova, G., Kavetskiy, A., Prior, S.A., Torbert, H.A., 2016. Benchmarking the inelastic neutron scattering soil carbon method. Vadose Zone J. 15.