11-Carbon-txt_ch11-pp 21/9/04 8:16 am Page 233

11

The carbon balance of forest soils: detectability of changes in soil carbon stocks in temperate and Boreal forests

F. Conen, A. Zerva, D. Arrouays, C. Jolivet, P.G. Jarvis, J. Grace and M. Mencuccini[Q2]

Symbols

$\hat{\gamma}(h)$, spatial semivariance at lag h
h, spatial distance or lag separating two samples
N, number of pairs of soil samples separated by a lag h
x, spatial coordinate locating soil sample
z, a regionalized variable of interest, in this case soil carbon content
δ, minimum detectable difference
μ_t, estimate of mean soil carbon content at time t
α, level of probability for type I error
β, level of probability for type II error
s^2, estimate of the population variance (σ^2)
n, sample size
$t_{\alpha,\upsilon,}$ critical t-value for type I error with υ degrees of freedom
$t_{\beta,\upsilon,}$ critical t-value for type II error with υ degrees of freedom

1. Introduction to soil carbon stocks

Vegetation sits at the interface between the global atmosphere and the soil, where it effectively transfers carbon as organic detritus from the stock in the atmosphere to the stock in the soil. Forests in particular bring about this transfer in a very effective manner, with the result that large amounts of carbon have accumulated in the soil within many forests over thousands of years. Globally, there is about three times as much carbon in soils as in vegetation, with the largest proportions in the northern

The Carbon Balance of Forest Biomes, edited by H. Griffiths & P.J. Jarvis.

temperate and Boreal forests (Schlesinger, 1997; Watson *et al.*, 2000). Thus, changes in the stocks of soil carbon, as a result of climate, land-use changes and management practices, in particular, may be very important components of local and regional carbon budgets.

The increase in atmospheric CO_2, and the likely increase in global temperatures that will result, is expected to alter the distribution of carbon between atmosphere, vegetation, and soils (Watson *et al.*, 2000). However, both the direction and the magnitude of the possible changes in soil carbon stocks are still being discussed, because of their dependence on several inputs and outputs. For example, rising atmospheric CO_2 concentration and increased nitrogen availability are expected to enhance CO_2 uptake by trees from the atmosphere and lead to enhanced growth and detritus production. By contrast, it is commonly supposed that an increase in temperature will lead to an increase in rates of respiratory oxidation of soil organic matter and a reduction of the soil carbon stock. In neither case is the long-term acclimation of the processes involved sufficiently well understood or the feedbacks adequately appreciated for reliable projections to be made.

The responses of soil carbon pools to climate change can be viewed as the results of an experiment lacking a control treatment. Consequently it is important to establish a firm baseline against which later observations can be compared. Because climate change is expected to occur on a time-scale of several decades, repeated direct measurements of soil carbon stocks at intervals of several years seem a good option. The choice of methods for determining changes in soil carbon stocks is large (Post *et al.*, 2001) but repeated direct measurements are relatively cheap and applicable in all parts of the world where soil carbon stocks are currently in a quasi-steady-state. Where soils are not in a quasi-steady-state as a result of land-use change, the same method could be used to assess the combined effects of land-use and climate change. Well-established methods for measuring changes in soil carbon pools will be one of several tools used to monitor and verify changes in soil carbon stocks as part of the Kyoto Agreement and resulting initiatives such as carbon trading.

The aim of this paper is to assess the site-to-site variability in the detectability of changes in soil carbon stocks. We do this by looking at 24 datasets of soil carbon in temperate and Boreal forests, from either published studies or our own recent measurements. It is not our aim to derive optimal sampling strategies that minimize the sample size to determine either area-averaged stocks or stock changes over time. Rather, we want to summarize the available literature and attempt to derive general principles that could be of use elsewhere. To do this, we explore the form of the relationship between mean and variance across the studies published thus far and we discuss its implication for future studies. Also, we propose a generalized scaling relationship between reductions in area sampled and proportion of variance remaining to identify the degree of spatial dependency of forest soil carbon densities. Finally, to illustrate the consequences of our findings, we present the case of the simplest sampling strategy (i.e., simple random sampling), and we calculate the minimum detectable differences in soil carbon stocks across this sample of forest sites.

First, the study sites are described and the methodology presented, then the concept of minimum detectable difference is briefly introduced and explained. Finally, the major results are presented and discussed.

2. Background to sites compared in this study

2.1 *Perthshire, UK*

Samples were collected in Griffin Forest, on the north-facing slope of the Tay Valley near Aberfeldy in Perthshire, Scotland, at an elevation of about 340 m (56°36' N, 3°48' E). Annual precipitation averages 1200 mm and mean annual temperature is 8.2 °C. The formerly grazed heathland on a podsolic soil of a sandy loam texture was ploughed and planted with Sitka spruce (*Picea sitchensis* (Bong.) Carr.) in 1981. The ploughing regime used to establish the forest inverted the material from furrows onto adjacent strips of land to form elevated ridges, resulting in three distinct surface zones: furrows, ridges, and undisturbed land. Ridges occupied about 50% of the total area, furrows and undisturbed ground about 25% each. There was no understorey and only little vegetation on the forest floor of occasional patches of mosses and grasses. Soil cores of 5.7 cm diameter were taken in a stratified random design from a 0.85 ha (1 hectare (ha) = 10^4 m^2) plot. Sampling depth was determined by the lower boundary of the original A horizon. This boundary was identified by the sudden change from dark brown to bright pale color. In the deepest parts of the furrows, where the A horizon had been completely removed by the plough, sampling depth was limited to the thickness of the L and O horizons formed since ploughing.

Samples were dried to constant mass at 60 °C and crushed in a mortar. Roots (larger than 1 mm) and the few stones (greater than 4 mm) were removed and the remaining sample was weighed, milled, and mixed. A sub-sample (approximately 1 g) was further ground and mixed in an agate mortar before being sub-sampled again (approximately 10 mg) for analysis in an elemental analyzer (Carlo-Erba 1106).

2.2 *Northumberland, UK*

Samples were collected at several localities within Harwood Forest (55°10' N, 2°3' W), in Northumberland, England. Harwood forest mostly consists of even-aged stands of pure Sitka spruce (*Picea sitchensis* (Bong.) Carr.), although in poorly drained sites lodgepole pine (*Pinus contorta* Dougl. ex Loud.) is also present. The forest was created over several decades during the period between the two world wars, by ploughing and subsequent planting of trees on ericaceous moorland and upland pasture. The dominant soil type is peaty gley. The area rises from 200 m in the southeast to 400 m in the northwest. Average annual precipitation is 950 mm, mean annual temperature is 7.6 °C. Understorey and forest floor vegetation were generally sparse or absent in the closed-canopy stands; however, a dense cover of heather, juncos, and graminoids was present in the clearfelled and replanted areas.

In several of the age classes, strata associated with the original ploughing could no longer be identified (either because of time or because of subsequent preparation practices), so that random samples were taken within plots nested within stands (Anderson and McLean, 1974). Soil samples were taken from eleven stands of pure Sitka spruce of varying age: 40 years (first rotation) and 12, 20, and 30 years old (second rotation), all located on peaty gley soils. Samples were also collected from one recent (3-year-old) clearfell and from one unplanted moorland plot outside the forest.

The entire area of the forest from which samples were taken was about 578 ha. In each stand (area varying between 40 and 60 ha) four or five plots were randomly selected and in each plot samples were taken from eight to ten points at randomly selected distances from the centre. The randomly selected distance was chosen such that the points would fall within a circle of 10 m radius around the center of the plot. The samples were taken with a soil auger of 5.7 cm diameter to a depth of 45 cm. Each sample was divided into L, O, and A layers for later analysis.

The samples were oven-dried at 105 °C for 24 h. Coarse fragments were removed by hand, and the soil was ground to pass a 0.5 mm mesh. Thirty per cent of all samples were analyzed both by C/N analyzer and by loss on ignition and three separate regressions were calculated for the litter, organic, and mineral layers (all $r^2 > 0.98$, $p < 0.001$). The carbon content of the other samples was measured by loss on ignition only, and the results converted to total carbon by using the previously calculated regression coefficients.

2.3 *Les Landes, France*

Soil samples were collected from a 9 ha mature forest (44°38' N, 1°14' W) of maritime pine (*Pinus pinaster* Ait.). The forest understorey vegetation is mainly composed of *Molinia caerulea* (L.) Moench., *Pteridium aquilinum* (L.) Kuhn, *Erica cinerea* L. and *Calluna vulgaris* (L.) Hull. This area receives about 900 mm per year of rain, and has a mean annual temperature of 12.7 °C. The land surface consists of a succession of dunes/interdunes, with the top of the dunes ranging from about 0.30–1.50 m in height and 10–50 m in width. Soils are hydromorphic sandy podsols, developed from Quaternary coarse sand, aeolian deposits. The vegetation, soil profile development, and organic matter storage are directly related to the microrelief and to seasonal variation of the superficial water-table level, with reference to the soil surface. Podsols with iron pans or rather cemented, enriched B-horizons are found in the well-drained upper parts of the ridges. More humus-rich podsols, with friable enriched B-horizons, sometimes overlaying a hydromorphic Cg-horizon, occur in the lower areas and poorly drained situations.

A stratified sampling design was applied. The forest was divided into 49 square subplots (strata). Half of the samples were collected along the lines of trees, the other half between the lines. The location of the sampling point was randomly selected, on the condition that at least one sample was located in each subplot. Samples were collected from 60 randomly selected sampling points. The O horizon samples were collected separately. Then, samples of the organo-mineral topsoil layer were collected down to depths of 0–0.4 m, 0.4–0.6 m, 0.6–0.8 m, and 0.8–1 m using a core sampler of 0.2 m inner diameter, to give samples of known volume for bulk density determination.

Bulk samples were oven dried to constant mass at 105 °C and an aliquot was finely crushed (less than 50 µm), to obtain reliable homogeneity before organic carbon and nitrogen microanalysis. Organic carbon and nitrogen contents were determined by dry combustion using an automated CNS[Q3] elemental analyzer (Fisons). Bulk densities were determined after oven-drying to a constant mass at 105 °C.

To estimate scales of spatial dependence, a classical estimator of the semivariogram, i.e. the spatial semivariance with lag h, was calculated:

$$\hat{\gamma}(h) = 1/2N(h) \sum_{i=1}^{N(h)} [z(x_i) - z(x_i + h)]^2 \quad (1)$$

where z is a regionalized variable, $z(x_i)$ and $z(x_i + h)$ are measured samples at points x_i and $x_i + h$, and $N(h)$ is several pairs separated by distance or lag h.

A spherical model describing the experimental semivariograms was fitted through a weighted least squares procedure. As recommended by Webster and Oliver (1990), the weighting used was the number of couples, $N(h)$.

3. Previous studies on soil carbon content

Growing interest in soil carbon as a likely variable in climate change has resulted in several published studies on the estimation of soil organic carbon content and its variability (e.g., Huntington *et al*., 1988; Fernandez *et al*., 1993; Liski, 1995; Homann *et al*., 2001; Conant and Paustian, 2002; Conant *et al*., 2003) or detailed descriptions of soil carbon stocks (Weber, 1999; Garten *et al*., 1999). Depending on the aim of the study, sampled areas have ranged from 10 m^2 (Liski, 1995) to at least 126 ha, with sampling points as far as 8 km apart (Homann *et al*., 2001). Only one study (Palmer *et al*., 2002) appears to have reported data on the variability in carbon stocks for several plots across an entire state (30 plots across Georgia, USA). In most studies the forest floor material (litter layer and O horizon) and several underlying layers within the upper 30 cm of mineral soil have been analyzed separately. Liski (1995) sampled to 40 cm, Huntington *et al.* (1988) to 54 cm, and Fernandez *et al.* (1993) to about 70 cm depth. Samples were taken with corers of cross-sections varying from 4.5 cm^2 (Garten *et al*., 1999) to 150 cm^2 (Homann *et al*., 2001) or excavated from 0.5 m^2 pits (Huntington *et al*., 1988; Fernandez *et al*., 1993). In the study using the smallest diameter corer, composite samples were made from two to four cores (Garten *et al*., 1999). Sample numbers in each study ranged from 18 (Weber, 1999; Garten *et al.*, 1999) to 271 (Homann *et al*., 2001) (*Table 3*). All these sites were in old-growth forests except for one site studied by Conant *et al.* (2003), which was in a 40-year-old second rotation forest, planted after harvesting of the previous old-growth stand.

4. Estimation of the mean and variance in soil carbon stocks across studies

As explained above, available data covered both our own field sites in Europe (two in the UK and one in France) and data obtained from the literature. No two examples are directly comparable to each other as they all employed different methods. However, all the examples had in common that soil carbon concentration per cent carbon) and soil mass per unit area (kilograms per square meter) had been determined on the same samples used for calculating soil carbon content per unit area (kilograms of carbon per square meter) for each sampling point. The more common practice of determining the two variables separately was deemed unsuitable for this type of investigation because the product of two averages is not necessarily the same as the average of the individual products, as will be shown later on.

The most important differences across studies were in the size of the sampled area, sampling depth, area covered by each sampling point, sampling design, and sample size. The magnitude of the area sampled was potentially the most important difference, as estimated variance should increase when larger areas are sampled. To correct for the differences across studies in the area sampled, we derived an empirical relationship between these two variables to enable us to scale the estimated variances to a common sampled area of one hectare.

To build this relationship, we identified those studies where estimates of variance were available at least at two different scales, i.e. where studies had deliberately been done with a 'nested' design (e.g., Homann *et al*., 2001; Palmer *et al*, 2002; Conant and Paustian, 2002; Conant *et al*., 2003; Northumberland, this study) or studies where an explicit geostatistical approach had been followed and estimates of variance derived in relation to distance (e.g., Liski, 1995; Les Landes, this study). For each of these studies, we calculated the ratios of plot sizes and variances of the smaller plots relative to the largest areas given (e.g., stands, forests, region), whereby a variance of 1 and a plot size of 1 were attributed to the largest spatial scale sampled and proportionally lower values of variance and area were attributed to the lower spatial scales. For Palmer *et al.* (2002), we calculated plot areas based on the given distances for which variances were indicated (7.85×10^{-5},[Q4] 1, and 1 200 000 ha). By taking relative measures of size and variance, we postulated that a general relationship could be found across all studies, indicating the generality of the scaling of variance with distance across several orders of magnitude.

Variance estimates may also be affected by choice of the sampling design (see, for example, Papritz and Webster, 1995a, b). This is far more difficult to assess and account for. When we reviewed the available literature data, we assumed that the estimates of population variance were directly comparable. Finally, sample size can also affect estimates of variance. For instance, Conant and Paustian (2002) showed that a different allocation of a fixed sample size between plots and sub-plots resulted in different estimates of within-plot versus among-plots variance. We assessed the impact of initial sample size by regressing it, alone or in combination with other variables, against the corresponding estimates of variance. As mentioned already, our intention was not to determine the absolute sample sizes required at specific sites to achieve a specified aim, but rather to investigate the nature and magnitude of the variability reported for several studies across temperate and Boreal forests. The design of optimal sampling schemes, and the consequences of this choice for sample size, was outside the scope of this study.

5. Estimation of minimum detectable difference

We define the minimum detectable difference, δ, as the statistically significant difference between two estimates of mean soil carbon content at the same site on two different occasions (μ_{t1} and μ_{t2}). As significance limits, we set the probability for falsely rejecting the null hypothesis at 5% ($\alpha = 0.05$). The probability for falsely accepting the null hypothesis, we set at 10% ($\beta = 0.10$, i.e., the statistical power = 0.9). Assuming simple random sampling, δ can then be estimated in a one-sample *t*-test (Zar, 1999) from the following:

$$\Delta = \sqrt{\frac{s^2}{n}}\,(t_{\alpha,\upsilon} + t_{\beta,\upsilon}) \qquad (2)$$

where s^2 is an estimate of the population variance (σ^2), n is the sample size ($n_{t1} = n_{t2} = n$) ($\sqrt{s^2/n}$ is the estimated standard error of the mean), and $t_{\alpha,\upsilon}$ and $t_{\beta,\upsilon}$ are the critical t-values for the specified values of α and β with $n - 1$ degrees of freedom (υ). The estimate of δ assumes a normal distribution, that repeated sampling is performed by the same method (simple random sampling), and that the variance is equal on successive occasions.

For practical purposes, we may want to test the hypotheses that soil carbon either increases over time ($\mu_{t1} < \mu_{t2}$, for example soil carbon gain in recently afforested sites) or that it decreases over time ($\mu_{t1} > \mu_{t2}$, for example soil carbon loss from old growth forests). In both cases we should apply a one-tailed t-test. However, in the case of the effects of climate change on soil carbon stocks, the direction of the prediction is more uncertain and both increases and decreases have been suggested as possible trends. In this case, a two-tailed t-test would be more appropriate. Our calculated sample sizes are reported based on a one-tailed t-test. Estimates for two-tailed t-tests will obviously be larger and can easily be obtained from them.

If we have an estimate of the population variance, we can estimate Δ as a function of sample size or, alternatively, estimate the sample size required to achieve a desired Δ. In both cases, the estimate is subject to error in the variance estimate. If sample sizes are required within certain prescribed confidence limits, they could be substantially larger, depending on the error estimate of the population variance (Johnson *et al.*, 1990).

6. Soil carbon stocks in temperate biomes

6.1 *Sampling procedure*

In this study, we have only included estimates of soil carbon content and its variance that were produced from direct measurements of soil carbon content on several individual samples (cf., Schwager and Mikhailova, 2002). The results for the site in Perthshire are used to demonstrate the bias introduced when carbon content is estimated as the product of average carbon concentration (grams of carbon per gram of soil) and average bulk density (or soil mass per unit area) and depth. The area-weighted mean soil mass of the three sampled strata was 119 kg m^{-2} and the mean carbon concentration within these layers was 9.3 g (C) g^{-1} soil. The area-weighted mean carbon content of the individual sampling points was 9.7 kg (C) m^{-2}. Had carbon content been calculated as the product of mean soil mass and mean carbon concentration from the same samples, the area-weighted carbon content would have been overestimated by 13% (*Table 1*). Overestimation is a general characteristic arising from the negative correlation between carbon concentration (grams of carbon per gram of soil) and soil bulk density (grams of soil per cubic centimeter), which again is a result of different densities (grams per cubic centimeter) of organic and mineral soil particles.

6.2 *Scaling of variance with plot size*

Because of the complexity of the processes leading to carbon accumulation, soil carbon stocks vary from place to place and over many spatial scales, from the very local

Table 1. Mean depth, soil mass, carbon concentration, and content of the organically rich layers (L, O, A) in different strata in a ploughed and afforested podsol in Perthshire, UK. Carbon content was calculated in two different ways. Numbers in brackets indicate one standard error.

Stratum	Fraction of total area	Sample size, *n*	Sample depth (cm)	Soil mass (kg m^{-2})	Carbon concentration (g (C) g^{-1} soil)	Carbon content (kg (C) m^{-2}) Mean of individual samples	Mean soil mass × mean of carbon concentration
Furrows	0.25	20	10.7 (1.3)	65 (10)	6.5 (0.7)	3.8 (0.65)	4.2 (0.79)
Undisturbed	0.25	20	19.8 (1.1)	87 (9)	13.1 (1.3)	9.8 (0.65)	11.4 (1.63)
Ridges	0.5	40	30.1 (1.5)	162 (10)	8.7 (0.8)	12.6 (0.90)	14.2 (0.17)
Total area	1	80	22.7 (0.9)	119 (6)	9.3 (0.5)	9.7 (0.51)	11.0 (0.46)

micropatch to the landscape and the regional scales. Geostatistical tools have often been applied to determine the spatial dependency of soil physical and chemical properties, but most studies have focused on changes occurring over only two or three orders of magnitude of changes in distance (e.g., from meters to hundreds of meters). In fewer cases, attention has been focused on the variability occurring at the regional scale (e.g., from kilometers to hundreds of kilometers). For logistical reasons, a sampling scheme investigating changes in soil carbon stocks over several orders of magnitude of distance has not, to our knowledge, been implemented yet.

The study made in the forest at Les Landes provides an example of the spatial dependency of soil carbon stocks in the range from tens of meters to hundreds of meters. In the upper 40 cm, where most of the carbon was located, 72% of the total variance of the 9 ha plot was nugget variance (i.e., the estimated vertical intercept of the regression model of semi-variance against distance) and the remaining 28% of the variance occurred over a range of 35.7 m (*Table 2*). Thus, compared with the size of the sampled plot (9 ha), subplots two orders of magnitude smaller (0.1 ha, i.e. a circle with diameter of up to 36.9 m) yielded lower estimates of variance. However, because of the nugget effect, even the smallest subplots still contain more than two-thirds of the total variance estimated at the plot scale. Although the nugget variance at greater depths was smaller, the range was similar. Large nugget values can be reduced by bulk sampling and by enlarging the size of the sampling point. However, this example is indicative of the low spatial dependency of variance on plot size. Decreasing plot size does not necessarily result in large decreases in variance.[Place Table 2 near here]

The ranges of spatial dependence at Les Landes were larger than those found at a site in Finland (Liski, 1995). Liski (1995) identified scales of spatial dependence between 1.1 m for the 0–10 cm layer and 8.4 m for the 0–40 cm layer. On a larger scale, Homann *et al.* (2001) found that 81% (0–15 cm depth) and 83% (15–30 cm depth) of the total variance of a forest of over 126 ha in size was already contained inside 2 ha plots. Although the plots in this study were larger in absolute size, the relative change in size from larger to smaller plots (from 126 to 2 ha, i.e. a factor of about 60) was more modest than in the example from Les Landes. In the forest floor, only 36% of

Table 2. Geostatistical parameters for soil carbon stocks at different depth intervals at a forest site in Les Landes, France

Depth interval (cm)	Mean (kg (C) m^{-2})	Coefficient of variation (%)	Nugget variance (fraction of sill)	Range (m)	Sill ((kg (C) $m^{-2})^2$)
0–40	6.91	69.8	0.72	36.9	22.8
40–60	1.63	67.5	0.24	37.71	1.13
60–80	1.22	62.4	0.03	40.05	1.02
80–100	1.66	132.5	0.52	40.05	1.00

the total variance was found at the 2 ha scale. However, this was of minor importance as the forest floor only contributed about 13% of the total carbon measured.

Over a wider range of plot sizes, there can be larger changes in variance. At our site in Northumberland, for example, 0.03 ha plots had only 54% of the variance found within an area of 578 ha (four orders of magnitude change in plot size). Some of this difference might result from the difference in age of the trees among the plots in different forest compartments. Over an even wider range, Palmer *et al.* (2002) studied variability in forest soil carbon across Georgia (USA), and found on average 21% of the total variance over distances of '*generally less than one meter*', another 38% over 1 ha plots, and the remaining 41% over an area of more than one million hectares (i.e., 11 and 7 orders of magnitude change in plot size, respectively). Clearly, variability of soil carbon stocks occurs over several orders of magnitude (cf., Webster and Oliver, 2001).[Q8]

When relative variances are plotted against relative plot sizes for those studies where multiple estimates were available, a strong logarithmic relationship is indeed found between these two variables (*Figure 1*; $r^2 = 0.60$, $n = 10$, nonlinear regression with the intercept for the logarithmic relationship forced through the point (1;1)). One point (an old-growth Douglas fir (*Psuedotsuga menziesii* (Mirb.) Franco) forest where buried logs could occasionally be found in the sample micro-plots (Conant *et al.*, 2003)), was noticeably outside the scatter of the regression line. For this example, reducing the area sampled from around 300 ha to a micro-plot of 2 m × 5 m size only reduced the variance by about 11% (Conant *et al.*, 2003)! The regression without this point is shown in *Figure 1* ($y = 0.035 \ln(x + 1)$, $r^2 = 0.84$, $n = 9$). This relationship indicated an approximately halving of the variance only for every six orders of magnitude decrease in plot size. In other words, reducing the plot size from 100 ha to 0.0001 ha (i.e. 1 m^2) (or, equivalently, from 10^6 ha to 1 ha) would only decrease the variance by an estimated 50%. We used this relationship to correct the estimated variance (hence, the estimated δ as well) to account for differences in sampling area across studies (δ_{corr}, *Table 3*). This correction changed the estimated δ only marginally, confirming that across-study changes in plot size did not significantly affect our estimates of variance.

6.3 *Factors affecting variance*

Considering all the sites presented in *Table 3*, there was a sevenfold difference among the smallest and the largest mean carbon content (3.0–21.3 kg (C) m^{-2}) and

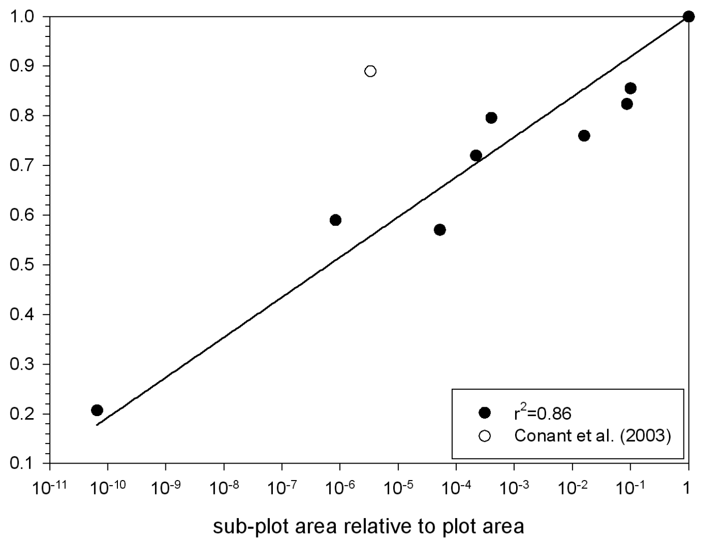

Figure 1. *Relationship between relative plot size and relative plot variance in soil carbon stocks across studies for which estimates were available at different spatial scales (Les Landes and Northumberland, this study; Homann et al., 2001; Palmer et al., 2002; Conant et al., 2003). Relative plot size was calculated as the ratio of the size of the sub-plots to the size of the plots. Relative variance was calculated as the ratio of the variance of the sub-plots to the total variance of the plots. The logarithmic regression line was forced to pass through the (1;1) point. The point marked by the white circle was excluded from the regression.*

coefficients of variation ranged from 10 to 112%. A strong curvilinear relationship was found between mean carbon content and variance corrected to 1 ha across all the sites reported in *Table 3* ($y = 0.0302\,x^{2.4463}$, $r^2 = 0.50$, $p < 0.001$). The relationship became even stronger ($y = 0.0208\,x^{2.5088}$, $r^2 = 0.63$, $p < 0.001$) when one site with a particular disturbance history was discarded from the analysis. One of the two Washington studies was heavily affected by the presence of buried wooden logs (old growth) that, as mentioned already, created high levels of very local variability. Some other 'disturbed' sites also showed a tendency towards larger scatter. For instance, the burning of the residues of the previous old-growth forest (second growth stand) (Conant *et al*., 2003) appeared to increase variance. Estimates of variance at the plot scale for both the Perthshire and Northumberland sites may have been affected by the ploughing before planting. At the site in Perthshire, ploughing increased the total variance fourfold (*Table 1*), assuming initial variance was the same as we found 20 years later in the still-undisturbed bands between the ridges. The site at Les Landes was characterized by uneven topography of sandy, C-poor dunes with low water holding capacity dunes, and shallow depressions where organic matter accumulated and water is more freely available. Such conditions create stark contrasts in conditions for plant growth and carbon mineralization rates. As mentioned in the site description, this leads to vegetation, soil profile and carbon storage being directly related to the micro-relief.

Table 3. *Estimated mean carbon content and its coefficient of variation (CV) for various forest sites. For each site, the minimum detectable change (Δ) for a sample size of 100 was estimated, as well as the sample size required to obtain a Δ of 0.5 kg (C) m^{-2}. The table is ranked by increasing variance. These calculations assume simple random sampling.*

Location	+Area (ha) [Q7]	Sample size (n)	Carbon content (kg (C) m^{-2})	CV (%)	Δ for n = 100 (kg (C) m^{-2})	Δ_{corr} for n = 100 (kg (C) m^{-2})	Sample size for Δ = 0.5 kg (C) m^{-2} (n)	Reference
Tennessee (stand)	0.1	3	3.0	18	0.16	0.17	13	Conant *et al.*, 2003
Tennessee, USA (235 m altitude)	0.02	18	4.0	10	0.11	0.12	8	Garten *et al.*, 1999
Tennessee (plot)	0.001	12	3.0	16	0.15	0.16	12	Conant *et al.*, 2003
Tennessee, USA (335 m altitude)	0.02	18	3.8	15	0.16	0.17	13	Garten *et al.*, 1999
Helsinki area, Finland	0.005	126	4.5	15	0.20	0.21	20	Liski, 1995
Tennessee, USA (1000 m altitude)	0.02	18	7.4	13	0.29	0.31	40	Garten *et al.*, 1999
Tennessee, USA (940 m altitude)	0.02	18	10.7	10	0.32	0.34	48	Garten *et al.*, 1999
Tennessee, USA (1670 m altitude)	0.02	18	9.6	12	0.34	0.36	53	Garten *et al.*, 1999
Oregon, USA	126	271	8.6	15	0.40	0.36	53	Homann *et al.*, 2001
Tierra del Fuego, Argentina	48.6	18	6.6	32	0.43	0.40	64	Weber, 1999
Washington (second growth, stand)	2.50	3	4.8	67	0.94	0.93	341	Conant *et al.*, 2003
Tennessee, USA (1650 m altitude)	0.02	18	8.9	18	0.47	0.50	100	Garten *et al.*, 1999
Washington (old growth, stand)	300.8	3	7.1	112	2.36	2.11	1752	Conant *et al.*, 2003
Perthshire, UK (undisturbed)	0.85[a]	20	9.8	30	0.86	0.86	292	This study
Maine, USA	0.4	24	11.1	26	0.85	0.87	295	Fernandez *et al.*, 1993
Washington (second growth, plot)	0.001	12	4.8	59	0.84	0.94	347	Conant *et al.*, 2003
Les Landes, France (0–40 cm)	9	60	6.9	70	1.43	1.37	739	This study
Northumberland, UK, (forest)	578	6	21.3	40	2.53	2.23	1951	This study
Northumberland, UK (stand)	50	5	21.3	37	2.31	2.15	1811	This study
Perthshire, UK, (ploughed)	0.85[a]	80	9.7	49	1.45	1.46	837	This study
New Hampshire, USA	23	55	16.0	38	1.86	1.76	1218	Huntington *et al.*, 1988
Northumberland, UK, (plot)	0.03	8	21.3	35	2.26	2.40	2260	This study
Washington (old growth, plot)	0.001	12	7.1	106	2.22	2.48	2415	Conant *et al.*, 2003

Equation 1 can be re-written as follows:

$$n = \frac{s^2}{\Delta^2}\,(\mathrm{t}_{\alpha,\upsilon} + \mathrm{t}_{\beta,\upsilon})^2 \qquad (3)$$

It is now apparent that the number of samples required, for a fixed minimum detectable difference, is a function of the population variance, not of its coefficient of variation or of its standard deviation. Because sample variance was strongly related to mean carbon content across sites, the number of samples required to estimate a Δ of 5 Mg (C) ha^{-1} (assuming simple random sampling, cf. equation 1) was positively correlated to mean carbon content ($y = 0.8967\,x^{2.4232}$, $r^2 = 0.63$, $p < 0.001$), i.e. carbon-rich sites require a far larger sampling effort than carbon-poor sites (*Figure 2*).

The estimated values of variance were finally regressed against the initial sample size, either alone or in a multiple regression including site carbon stocks. In neither case was a significant relationship found, suggesting that the available estimates of variance were reasonably robust.

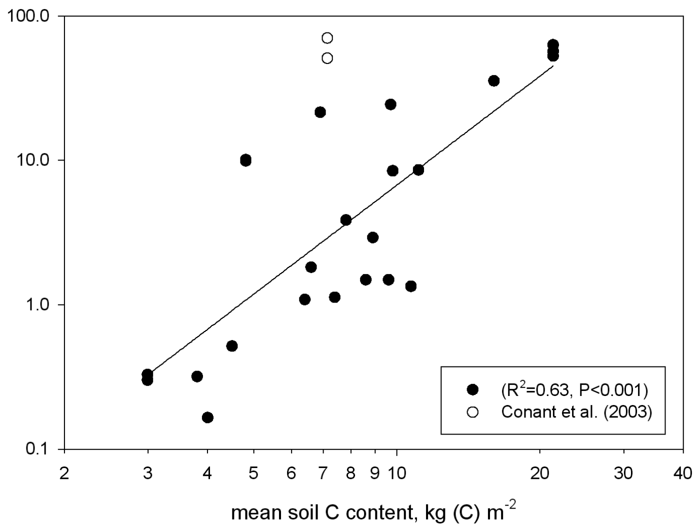

Figure 2. *Relationship between mean soil carbon stocks (kilograms of carbon per square meter) and estimated variance in soil carbon stocks across temperate and Boreal forests. For a change in soil carbon stocks between 3.0 and 21.3 kg (C) m^{-2}, the estimated population variance changed over three orders of magnitude (closed symbols). One site (Conant et al., 2003) fell outside the range of the regression between these two variables (labeled as outliers in the graph, open symbols; the two estimates are referred to plot-level and stand-level variance, respectively). At this old-growth Douglas fir site, spatial variability was large because of the presence of buried logs.*

6.4 *Sampling intensity and related minimum detectable changes*

Based on a sample size of 100, the minimum change that could be detected with a statistical significance ($\alpha = 0.05$, $\beta = 0.1$) ranged from 0.12 to 2.48 kg (C) m^{-2} and was lognormally distributed among the studied sites. For sites disturbed by wind-throw (Huntington *et al*., 1988; Weber, 1999) or ploughing (Perthshire and Northumberland, plot scale), we found the geometric mean δ for a sample size of 100 to be 1.37 kg (C) m^{-2} (median 1.60). At undisturbed sites, the geometric mean δ based on 100 samples would be 0.39 kg (C) m^{-2} (median 0.34).

To attain a δ of 0.5 kg (C) m^{-2}, the sample sizes that would be required are estimated to lie between 8 and 2415 depending on the particular site carbon content (*Table 3*). The geometric mean sample size for a δ of 0.5 kg (C) m^{-2} at the disturbed sites is 777 (median 1069). At the Perthshire site, the ploughing disturbance calls for a nearly threefold increase in sample size, despite the stratified sampling regime (*Table 3*). For all the other 'undisturbed' sites, the geometric mean sample size for achieving a Δ of 0.5 kg (C) m^{-2} is 63 samples (median 46).

7. Conclusions and summary

Estimating soil carbon content as the product of mean carbon concentration and bulk density can result in considerable overestimation. Carbon concentration and soil mass need to be measured on the same sample and carbon contents calculated for each individual sample before averaging. The effect of this bias is likely to be smaller (but still greater than zero) when the primary objective is in determining stock changes over time.

Variance and mean carbon content are significantly and positively related to each other, although some sites showed much higher variability than predicted by this relationship, likely as a consequence of their particular site history, forest management, and micro-topography. Because of the proportionality between mean and variance, the number of samples required to detect a fixed change in soil carbon stocks varied directly with the site mean carbon content from less than 10 to several thousands across the range of carbon stocks normally encountered in temperate and Boreal forests. This raises important questions about how to derive an optimal sampling strategy across such a varied range of conditions so as to achieve the aims of the Kyoto Protocol.

Overall, on carbon-poor forest sites with little or no disturbance to the soil profile, it is possible to detect changes in total soil organic carbon over time of the order of 0.5 kg (C) m^{-2} with manageable sample sizes even using simple random sampling (i.e. about 50 samples per sampling point). More efficient strategies will reveal even smaller differences. On disturbed forest sites (ploughed, windthrow) this is no longer possible (required sample sizes much greater than 100). Soils developed on coarse aeolian sediments (sand dunes), or where buried logs or harvest residues of the previous rotation are present, can also exhibit large spatial variability in soil carbon. Generally, carbon-rich soils will always require larger numbers of samples. On these sites, simple random sampling is unlikely to be the preferred method, because of its inherent inefficiency. More sophisticated approaches, such as paired re-sampling inside relatively small plots (see, for example, Ellert *et al*., 2001) are likely to reduce sample size significantly and lead to detection of smaller differences in carbon stocks over time. However, it remains to be shown that at these sites the application of efficient sampling designs will result in the detection of differences relevant for the objectives of the Kyoto Protocol (cf., Conant *et al*., 2003).

Finally, it should also be noted that, compared with the accuracy with which changes in atmospheric carbon content can be detected (less than 1 cm^3 cm^{-3}[Q5] CO_2), changes in soil carbon stocks are very uncertain. A release of 0.5 kg (C) from 1 m^2 of soil surface is equivalent to an increase in CO_2 concentration of about 125 cm^3 cm^{-3}[Q5] in the air column above the same area.

Acknowledgments

The EU-funded Carbo-Age project (contract ENV4-CT97-0577) and the NERC grant GR9/4806 provided the framework within which this study took place. Thanks are due to Mark Rayment for his encouragement to write this paper. Thanks are also due to Pete Smith (Aberdeen University, UK) and Pierre Bernier (Canadian Forest Service, Canada) for comments on an earlier version of the paper.

References

Anderson, V.L. and McLean, R.A. (1974) *Design of Experiments, A Realistic Approach*. Marcel Dekker, New York.

Conant, R.T. and Paustian, K. (2002) Spatial variability of soil organic carbon in grasslands: implications for detecting change at different scales. *Environmental Pollution* **116:** S127–S135.

Conant, R.T., Smith, G.R. and Paustian, K. (2003) Spatial variability of soil carbon in forested and cultivated sites: implications for change detection. *Journal of Environmental Quality* **32:** 278–286.

Ellert, B.H., Janzen, H.H. and McConkey, B.G. (2001) Measuring and comparing soil carbon storage. *In*: Lal, R., Kimble, J.M., Follet, R.F. and Stewart, B.A. (eds) *Assessment Methods for Soil Carbon*, pp. 131–146. CRC Press, Boca Raton.

Fernandez, I.J., Rustad, L.E. and Lawrence, G.B. (1993) Estimating total soil mass, nutrient content, and trace metals in soils under a low elevation spruce-fir forest. *Canadian Journal of Soil Science* **73:** 317–328.

Garten, C.T. Jr., Post, W.M.III., Hanson, P.J. and Cooper, L.W. (1999) Forest soil carbon inventories and dynamics along an elevation gradient in the southern Appalachian Mountains. *Biogeochemistry* **45:** 115–145.

Homann, P.S., Bormann, B.T. and Boyle, J.R. (2001) Detecting treatment differences in soil carbon and nitrogen resulting from forest manipulations. *Soil Science Society of America Journal* **65:** 463–469.

Huntington, T.G., Ryan, D.F. and Hamburg, S.P. (1988) Estimating soil nitrogen and carbon pools in a northern hardwood forest ecosystem. *Soil Science Society of America Journal* **52:** 1162–1167.

Johnson, C.E., Johnson, A.H. and Huntington, T.G. (1990) Sample size requirements for the determination of changes in soil nutrient pools. *Soil Science* **150:** 637–644.

Liski, J. (1995) Variation in soil organic carbon and thickness of soil horizons within a boreal forest stand – effect of trees and implications for sampling. *Silva Fennica* **29:** 255–266.

Palmer, C.J., Smith, W.D. and Conkling, B.L. (2002) Development of a protocol for monitoring status and trends in forest soil carbon at a national level. *Environmental Pollution* **116:** S209–S219.

Papritz, A. and Webster, R. (1995a) Estimating temporal change in soil monitoring. 1. Statistical theory. *European Journal of Soil Science* **46:** 1–12.

Papritz, A. and Webster, R. (1995b) Estimating temporal change in soil monitoring. 2. Sampling from simulated fields. *European Journal of Soil Science* **46:** 13–27.

Post, W.M., Izaurralde, R.C., Mann, L.K. and Bliss, N. (2001) Monitoring and verifying changes of organic carbon in soil. *Climatic Change* **51:** 73–99.

Schlesinger, W.H. (1997) *Biogeochemistry. An Analysis of Global Change*, 2nd edition. Academic Press, San Diego, CA.

Schwager, S.J. and Mikhailova, E.A. (2002) Estimating variability in soil organic carbon storage using the method of statistical differentials. *Soil Science* **167:** 194–200.

Watson, R.T., Noble, I.R., Bolin, B., Ravindranath, N.H., Verardo, D.J. and Dokken, D.J. (2000) *Land Use, Land Use Change and Forestry*. Special Report, IPCC. Cambridge University Press, Cambridge, UK.

Weber, M. (1999) Carbon budget of a virgin Nothofagus forest in Tierra del Fuego. *Forstwissenschaftliches Centralblatt* **118:** 156–166. (In German.)

Webster, R. and Oliver, M.A. (1990) *Statistical Methods in Soil and Land Resource Survey*. Oxford University Press.[Q6]

Zar, J.H. (1999) *Biostatistical Analysis*, 4th edition. Prentice Hall International, NJ.