Ecological Indicators 102 (2019) 401–409

Contents lists available at ScienceDirect

Ecological Indicators

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

Original Articles

How many samples? Soil variability affects confidence in the use of common agroecosystem soil indicators

Johannes Welsch[a,d,1], Cao Songling[b,c,1], Hannah L. Buckley[c], Niklas J. Lehto[d], E. Eirian Jones[d], Bradley S. Case[c,*]

[a] Environment Canterbury, PO Box 345, Christchurch, New Zealand
[b] Sichuan Agricultural University, College of Agronomy, Chengdu, China
[c] Auckland University of Technology, School of Science, Private Bag 92006, Auckland, New Zealand
[d] Lincoln University, Faculty of Agriculture and Life Sciences, PO Box 85084, Lincoln, Canterbury, New Zealand

ARTICLE INFO

Keywords:
Agroecosystem
Carbon
Estimation
Multi-functionality
Power analysis
Sampling intensity
Scaling up
Soil indicator
Spatial variation
Woody vegetation

ABSTRACT

There is a need for accurate and easily-measured indicators suitable for characterising and monitoring agroecosystem multi-functionality. This is particularly true in intensively-farmed landscapes where it is of interest to quantify the role of small, woody vegetation features in providing ecosystem services such as carbon sequestration. However, soil variability introduced by natural and management processes can interact with sampling designs to result in inappropriate sampling intensities and high levels of uncertainty in measured indicators. This can have consequences for upscaling of ecosystem quantities and decision making. Here, we present results from a pilot study aimed at quantifying and understanding variation in ten common indicators of soil condition and function, within shelterbelt and adjacent field soils, at four dryland sheep farms in Canterbury, New Zealand. Our results demonstrate a high level of spatially-structured soil variability, driven by (1) the effects of woody vegetation on shelterbelt soils relative to field soils, (2) differences in underlying soil types among sites, and (3) possible effects of grazing animals within fields. This soil variability had clear knock-on impacts for appropriate sampling effort, depending on the soil indicator in question, the original soil sampling density, and whether the aim was to estimate population mean values or to detect differences among sites with confidence. On the whole, confidence in soil indicator estimates was highest for soil condition indicators (pH, soil moisture, bulk density), variable for carbon quantities, depending on the measure used, and lowest for soil biological process indicators (tea bag index decomposition rate, bait lamina probe micro-invertebrate activity, and dehydrogenase enzyme activity); estimation confidence was also mostly lower for shelterbelt soils due to the effect of woody roots and inputs on soil variability. Based on our results, we present indicative sample size requirements to estimate population means for these different soil indicators. Ultimately, we advocate for the use of pilot studies, such as the one presented here, to facilitate understanding of variability in soil function indicators within different agroecosystems, and how this variability is partitioned spatially within and among vegetated features.

1. Introduction

The concept of agroecosystem multifunctionality is based on the idea that agricultural landscapes comprise both production and non-production land elements that contribute to economic, social, and ecological ecosystem service outcomes (Manning et al., 2018). Related to this, the idea of 'land sparing versus land sharing' has focussed debate on the degree to which intensively-managed farm landscapes, with minimal amounts of non-production vegetation, could realistically contribute toward ecosystem management goals (Kremen, 2015). From the land sharing perspective, small woody elements, such as individual large trees, shelterbelts, and riparian strips, can sometimes provide the only non-production vegetation in intensively farmed landscapes, and it remains unclear how these patches of non-production vegetation differ from the production matrix in terms of their contributions to ecosystem services and fine-scale ecological function.

The sequestration and storage of carbon in woody farm features, both above- and below-ground, can play a role in offsetting the effects of greenhouse gas emissions (Doran-Browne et al., 2017) and in the provision of a range of ecosystem services (Kort and Turnock, 1998,

* Corresponding author.
E-mail address: Bradley.Case@aut.ac.nz (B.S. Case).
[1] First-authorship of this work is shared by JW and CS.

https://doi.org/10.1016/j.ecolind.2019.02.065
Received 18 September 2018; Received in revised form 29 January 2019; Accepted 28 February 2019

Mize et al., 2008). In particular, agroecosystem soil carbon is an essential component of soil fertility and higher amounts of soil organic matter enhances the diversity and function of belowground faunal and microbial communities, through, for example, soil nutrient enrichment via decomposition processes (O'Donnell et al., 2001). In theory, soil carbon dynamics within woody, non-production features are expected to differ from adjacent farm fields due to contrasting levels of soil disturbance, microclimatic conditions and contributions to the carbon pool and soil development from woody plant biomass and roots. Soil carbon variation can be high, depending on the time over which woody vegetation has exerted an effect on the soil environment, and on how production and non-production vegetation elements are distributed spatially in agroecosystems (Söderström et al., 2014). Such variation can have consequences for carbon accounting in terms of up-scaling carbon estimates to landscape or national scales and for understanding potential future impacts of climatic and land management across agricultural landscapes (Godde et al., 2016; Gollany and Venterea, 2018). Locally, at the farm scale, soil carbon variation among vegetation features can also affect soil processes that drive belowground biodiversity and agroecological function.

While research has highlighted above-ground carbon variability among land uses and vegetation types (e.g., Williams et al., 2018), there remains a paucity of similar data at farm-to-landscape scales for soil carbon quantities and associated indicators of belowground soil condition and biological function. A better quantification and understanding of such soil variability at different spatial scales, in the context of different underlying environmental conditions, can inform the development of sampling designs for accurately quantifying soil carbon and related indicators of soil function within different vegetation elements on farms. Here, we present results from a New Zealand case study that quantified variability in ten soil variables indicative of soil condition, carbon and nitrogen status, and biological function, within and between pine shelterbelts and adjacent farm fields underlain by different types of soils. We ask: (1) How is variation in soil indicators partitioned spatially within and among shelterbelts and fields across the four sites? (2) What are the consequences of variability and sample size for estimating landscape-scale, population mean values for soil indicators? and (3) What are the consequences of variability and sample size for detecting significant differences in soil indicator values among vegetation elements on farms?

2. Materials and methods

2.1. Study sites

The study was conducted at four farm sites in a *c.* 2300 km^2 area located on the lowland Canterbury Plains, near Christchurch, New Zealand (Fig. 1). The farms had all been managed for at least the previous 20 years as low-intensity, dryland sheep farms typical of the region. Agriculturally, the region comprises a mixture of sheep, high-intensity dairy, and arable farms, with a low proportion (5–10%) of woody vegetation, typically occurring as small, linear shelterbelt elements (Welsch et al., 2014). Each farm site is underlain by soils belonging to one of four New Zealand soil orders (NZ Soil Classification system; Hewitt, 2010; hereafter referred to as "soil types") that dominate the local landscape, together comprising *c.* 80% of soils in this region (Fig. 1a, Table 1). The farms contained at least one regularly-trimmed pine (*Pinus radiata*) shelterbelt (Fig. 1b, Table 1), the most common type of shelterbelt across the Canterbury Plains, adjacent to sheep grazing fields. Thus, we were able to assess variability in soil condition, carbon and function measures between shelterbelt and adjacent field elements, and how these differed among farm sites underlain by different soil types, while controlling for large-scale land use and management effects, and shelterbelt species.

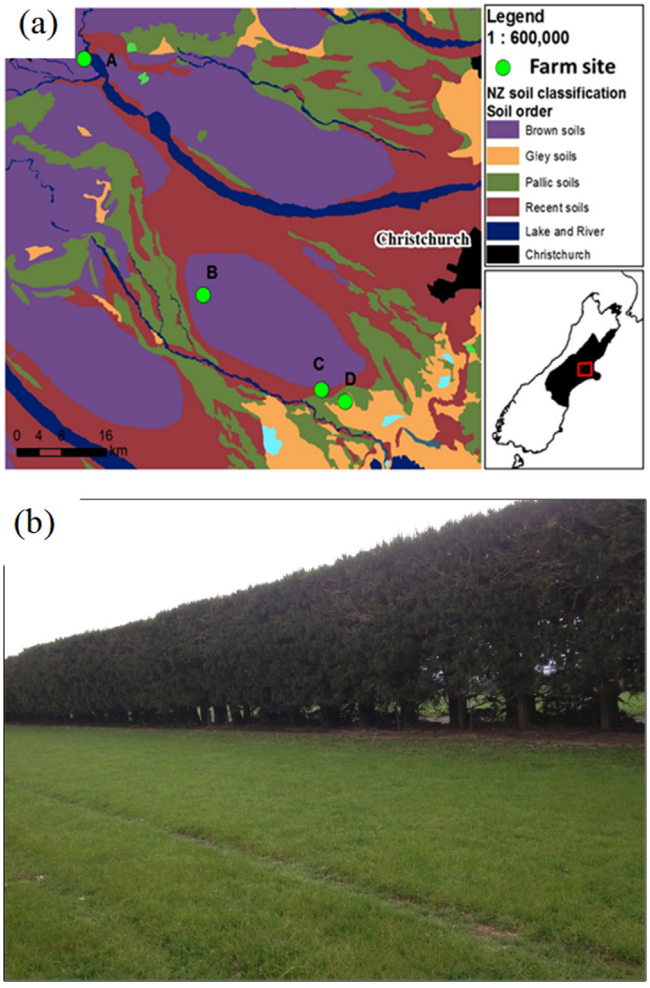

Fig. 1. (a) Distribution of study sites and underlying soil types across the case study area, and (b) an example of the shelterbelt and field conditions at the farm sites. The map in (a) was created in a GIS using the NZ Land Resource Inventory dataset (Newsome et al., 2000).

3. Sampling scheme and soil indicator quantification

We examined variation in ten soil variables that are indicative of soil condition, soil carbon and nitrogen status, and soil function. Soil pH, moisture content, and bulk density were measured as general indicators of soil condition. Soil carbon and nitrogen status was assessed by quantifying total carbon, organic carbon, permanganate oxidisable carbon (POXC) as a measure of labile carbon, and total nitrogen. Soil biological function was assessed via common indicators of soil microbial activity using dehydrogenase activity (DHA; Stevenson, 1959), decomposition rate using the "tea bag index" method (Keuscamp et al., 2013), and micro-invertebrate activity using bait lamina probes (e.g., Bispo et al., 2009).

To examine variation in the ten soil variables at the within-shelterbelt scale at each farm, we established three parallel, equally-spaced 120-m transects in each shelterbelt; depending on the shelterbelt width (Table 1), transect spacing varied from *c.* 0.5 to 1.5 m apart (Fig. 2). For comparison, an analogous sampling design was applied in the field adjacent to each shelterbelt, but situated 25 m away from the shelterbelt edge to minimise shelterbelt effects on field soils (Fig. 2). Along transects, we spaced sampling locations at 20-m intervals, thus defining the maximum sampling resolution along that axis. We quantified soil organic carbon, moisture, pH and bulk density for five of these locations along each of the three transects, resulting in 15 measurements of these variables for each shelterbelt/field. Total carbon, total nitrogen, and

Table 1
Environmental and shelterbelt characteristics at the four farm sites.

Farm Site	NZ Soil Class	Rainfall (mm/y)	Age (y)	Shelterbelt width (m)
A	Acidic Orthic Brown Soil	1000	18	2.5
B	Pallic Firm Brown Soil	650	25	2.5
C	Weathered Orthic Recent Soil	633	60	4.5
D	Typic Orthic Gley Soil	633	60	5.0

POXC were quantified at seven locations along the central transect within each shelterbelt/field; the tea bag and bait lamina probe assays were also carried out at these locations. Finally, DHA was originally quantified for every second sampling location along the three transects (nine samples per shelterbelt/field), but problems with subsequent lab analyses resulted in an average of eight samples for each shelterbelt and companion field. Thus, we did not quantify each soil variable for all soil sample locations, but at an interval that reflected the costs associated with a given measurement, the priority of a given variable in terms of quantifying variability, and the logistical difficulties with carrying out the measurements. All soil sampling was carried out in the first two weeks in December (the early austral summer), using a 10-cm step-on soil coring device, with a radius of 3 cm (volume of 283 cm^3). During this time, there were no significant rainfall events that would have greatly modified soil moisture conditions. Soil cores were stored in sealed plastic bags in a cool box in the field, after which they were transferred to a laboratory refrigerator (2–5 °C) for up to one week before air-drying. Soil samples were ground to pass through a 2-mm sieve in preparation for the laboratory measurements.

We measured soil moisture (%) using 10 g of ground sample as the relative percent decrease in weight after drying at 105 °C for 48 h (Blackmore et al., 1987). Oven dry bulk density was quantified for each sampling location by removing a 10 × 10-cm undisturbed sample using the "driving hammer" method (Blake and Hartge, 1986), oven-drying each sample, and calculating the weight of the sample relative to the initial volume. In addition, soil pH (1:2.5 soil:water w/v ratio) was measured for 10-g of air-dried and sieved soil sub-samples (Blackmore et al., 1987).

We analysed air-dried sub-samples of each soil core for total carbon and nitrogen using an Elementar Vario-Max CN Elemental Analyser (Elementar GmbH, Hanau, Germany). Organic carbon was quantified by measuring the loss-on-ignition (LOI) with a 10-g air-dried soil sample (Blackmore et al., 1987). Labile carbon was estimated using the active carbon technique for the determination of permanganate oxidisable carbon (POXC; Weil et al., 2003; Culman et al., 2012). We scaled all carbon and nitrogen values to tonnes per hectare (t ha^{-1}) quantities for the top 10 cm of mineral soil using bulk density measurements and the soil core volume.

We used the soil dehydrogenase enzyme activity (DHA) as an indicator of microbial activity (Alef, 1995; Thalmann, 1968), measured as the rate of reduction of triphenyltetrazolium chloride to triphenyl formazan and computed as per the equation provided by Alef (1995). Bait-lamina probes (BLPs) were used to indicate micro-invertebrate soil activity in the top 10 cm of the soil within the shelterbelts and adjacent fields on the farm sites that was contributing to the breakdown and decomposition of organic material (Torne, 1990). We placed bait-lamina probes vertically with the upper-most bait hole at the surface level of the mineral soil (Kratz, 1998); after a 90-day period, the probes were removed and micro-invertebrate activity was indicated as the percentage of bait eaten from the holes. The Tea Bag Index (TBI) method, following a standardised procedure introduced by Keuskamp et al. (2013), was used to quantify the rate of soil organic matter decomposition. For each of the shelterbelt and field tea bag sampling locations, five pairs of green and rooibos tea bags were buried at 8 cm depth. After 30 days, tea bags were collected and the replicate data were used to calculate k (decomposition rate) and S (stability) TBI parameters, as per Keuskamp et al. (2013).

3.1. Data analysis

All statistics and graphics were performed using R statistical software version 3.3.3 (R Core Team, 2017). Exploratory data analysis indicated no major departures from normality for all soil indicators (summary statistics provided in Appendix 1) and we therefore did not

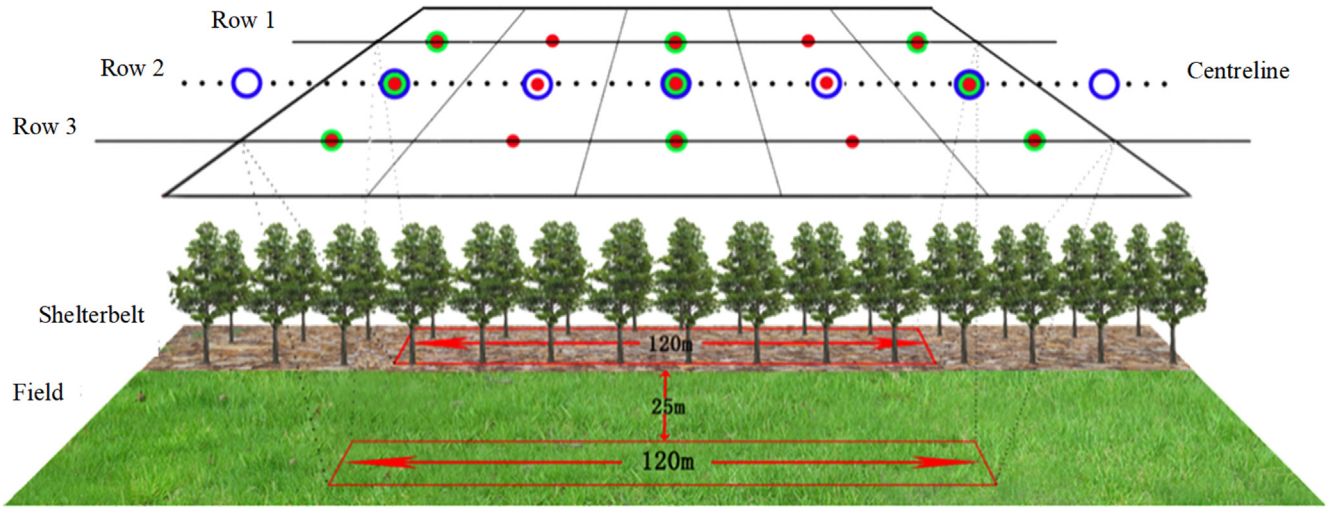

Fig. 2. Sampling design for quantifying soil condition, carbon and nitrogen, and process indicators at each farm site within both shelterbelts and adjacent farm fields. Red dots (●) indicate locations where soil moisture, bulk density, pH and organic carbon were measured. Blue rings (○) indicate locations where bait lamina probes were located, tea bags were placed, and permanganate oxidizable carbon, total carbon, total nitrogen were measured. Green rings (○) indicate locations where dehydrogenase activity (DHA) was measured. The finest distance between sampling points along the shelterbelt length is 20 m. In the width dimension, sampling point distances were equidistant and varied from 0.75 to 1.5-m depending on the shelterbelt width at each site (Table 1).

apply data transformations prior to analyses. We used box-plot graphs and principal component analyses (PCA – "prcomp" function in the R package "stats") to examine variability in, and associations among, measured indicators across sites and cover types. We used a one way analysis of variance (ANOVA) to test if measured soil indicators varied significantly within shelterbelts across the four sites, by comparing indicator values among transects established toward the fenced edge of the shelterbelt, at the shelterbelt centre, and toward the field side of the shelterbelt. For comparison, we similarly tested for difference among transects established in adjacent fields. Only soil indicators that were measured along all transects (moisture, pH, bulk density, organic carbon, and DHA) were compared for this analysis. Next, we used a two-way ANOVA to test whether the measured values differed among the four study farms, between fields and shelterbelts, and for interactions between these factors. The inclusion of an interaction term enabled us to determine if there was a combined effect of substrate and cover type on soil indicator values.

We carried out power analyses on the ten soil indicators for the farm field and shelterbelt subsets to compute, at three levels of confidence, minimum sample sizes required to: (i) estimate population mean values for soil indicators, and; (ii) detect a significant difference between shelterbelt and field indicator values. The power analysis for (i) was carried out based on the formula (Zar, 1999):

$$n = \left(\frac{Z\sigma}{E}\right)^2 \quad (1)$$

where, n is the estimated sample size, Z is a value from the standard normal distribution based on a given confidence level (e.g., $Z = 1.96$ for $\alpha = 0.95$ level of confidence), σ is the estimated population standard deviation for the given indicator (estimated from the field data), and E is the desired margin of error, which we set at 10% of the sample mean for each indicator. The estimate of sample sizes for (ii) used ANOVA variance partitioning results (i.e., statistics for within- and among-treatment effects) to carry out subsequent power analyses for the detection of differences among samples using the R package "pwr" (version 1.2–2; Champely, 2018). For both types of power analyses above, we estimated sample sizes for three levels of confidence: 0.8 ($p < 0.20$), 0.9 ($p < 0.10$), and 0.95 ($p < 0.05$).

4. Results

4.1. Spatial variability in soil indicators by cover type and farm site

For field soils, principal components analyses (PCA) showed a separation among the four sites in multivariate space in terms of measured indicators, as evidenced by the clustering of the different sites in different parts of ordination space (Fig. 3a). Sites A and D separated most strongly along PC axis 1, characterised by higher DHA (microbial activity) and pH for Site A to higher nitrogen and POXC quantities for Site D at the other end of the gradient. Sites B and C were found most strongly associated with PC axis 2, and particularly with higher organic carbon quantities and moisture levels, and lower decomposition and invertebrate activity rates. For shelterbelt soils (Fig. 3b), clustering of sites in the ordination was less strong, with some overlap for sites A and B and sites C and D on the ordination bi-plot. Sites A and B occurred in multivariate space on the opposite end of PC axis 1 from sites C and D; this axis was characterised by higher pH, organic carbon and process rates for the former pair of sites, and higher POXC and total carbon and nitrogen for the latter pair of sites.

Moisture levels were higher, and pH lower, in shelterbelt soils compared to field soils (Fig. 4). Both pH and moisture levels varied significantly between cover types (i.e., field vs. shelterbelt) and among sites, with a significant interaction among these factors in driving pH and moisture level variation (Table 2). Bulk density did not differ significantly between fields and shelterbelts, although there were significant differences among sites and in the interactions between site and

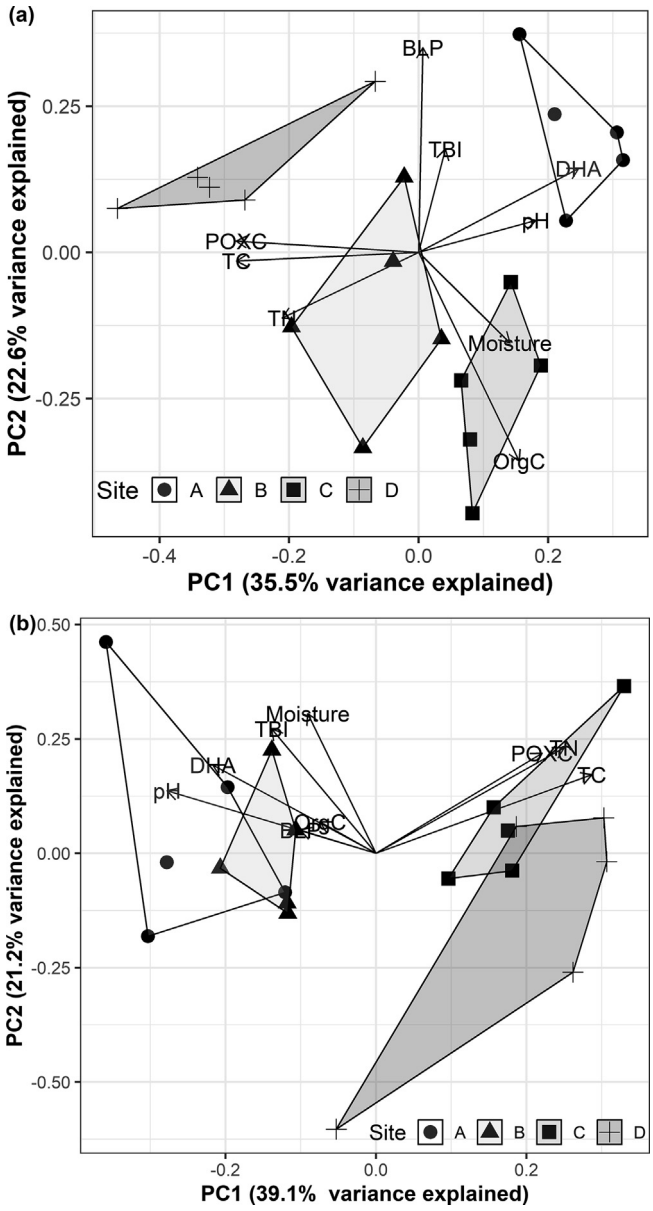

Fig. 3. Principal components analysis of soil characteristics, carbon stocks and carbon cycle processes within (a) fields and (b) shelterbelts in the four farm sites (acronyms defined in Table 2). Different symbols have been used to show the clustering of samples by the four farm sites, with convex hull polygons outlining samples of the same site to facilitate the visualisation of the clusters. Samples for field sites in (a) form four distinct farm clusters, while samples for shelterbelts in (b) show strong clustering for sites A and B and sites C and D on opposite ends of PC axis 1.

cover type (Table 2). For soil carbon and nitrogen quantities, there were no clear differences between fields and shelterbelts, with more variation occurring among sites (Fig. 5). Indeed, only total carbon differed significantly among cover types and sites, with significant interactions among these in determining total carbon levels (Table 2). Both organic carbon and total nitrogen varied among sites, and had significant cover type by farm site interaction effects (Table 2), while POXC only showed significant differences among sites (Fig. 5; Table 2). There were no significant ($p < 0.05$) differences in pH, soil moisture, DHA and organic carbon among the three parallel transects sampled within either shelterbelts or field soils.

Variation in biological process indicators was high overall (Fig. 6), with clear and significant differences between fields and shelterbelts for

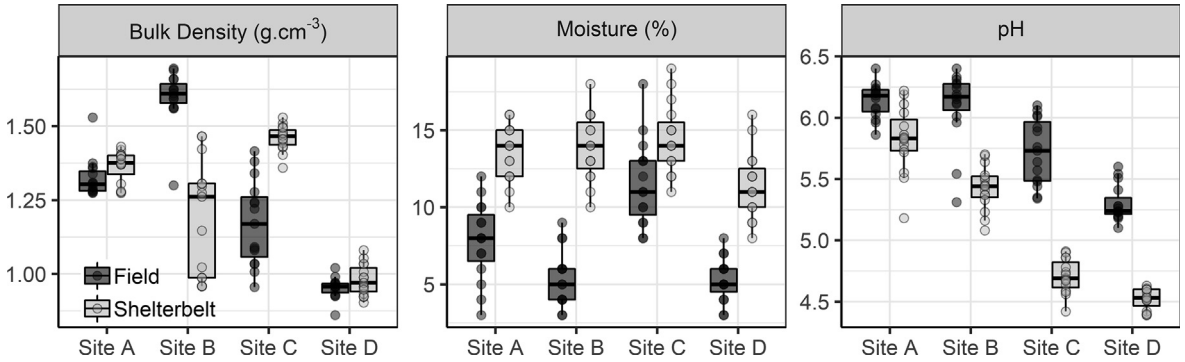

Fig. 4. Boxplot graphs showing variation in the three measured soil condition indicators (% moisture, pH, and bulk density) within and between fields and shelterbelts across the four farm sites. The boxplots show a summary of the data for each site, where the boxes represent the interquartile range (IQR, 25–75% of the data), median values are indicated by the bar within each box, and whiskers show the values within 1.5 times the IQR. The points overlaid onto the boxplots are data points for each individual sample taken within each site.

all three indicators, and a significant difference among sites as well for DHA (Table 2). Bait lamina probe feeding activity and decomposition rates were significantly higher, and DHA values lower, in shelterbelt soils compared to field soils.

5. Variability impacts on sample size requirements: power analyses

Power analyses showed that more-than-adequate samples had been taken to estimate population means for soil pH, moisture, bulk density with high confidence for both fields and shelterbelts across the study extent (Fig. 7a). Results for carbon and nitrogen measures varied by the type of C and N indicator and by cover type (fields vs. shelterbelts); for field soils, only POXC would require *c.* 20 additional samples to estimate mean values with 95% confidence for the study area, while for shelterbelt soils, from 10 to 60 more samples would be required to estimate all carbon and nitrogen indicators with 95% confidence, to within 10% of mean values. While more samples would be required to estimate population means for soil process indicators for both shelterbelt and field soils, the numbers of additional samples varied depending on the indicator and the cover type (Fig. 7a). For field soils, from *c.* 50 to 130 more samples would be required to estimate TBI decomposition rates and to measure BLP invertebrate activity with confidence, while only marginal increases in samples would be required to estimate these indicators within shelterbelt soils. By contrast, DHA was highly variable in shelterbelt soil samples compared to field soils, and thus *c.* 150 to 380 additional samples would have been required for accurate estimates of mean values for this indicator in shelterbelt soils (Fig. 7a).

On the whole, differences in the majority of soil indicators among both shelterbelts and fields across the four sites could be detected statistically with 90–95% confidence (Fig. 7b). For both shelterbelt and field soils, an additional 10 to 20 samples would be required to detect significant differences in TBI decomposition rates and BLP invertebrate activity rates with confidence, and an additional five or samples would be required to detect POXC differences in shelterbelt soils only.

6. Discussion

Many agricultural landscapes, including New Zealand's, are a product of initial large-scale forest clearance, followed by decades-to-centuries of ongoing management activities such as tilling, burning, fertilising and irrigating with the aim of generating and maintaining productive agricultural lands (Norton and Reid, 2013). The legacy of these activities on a broad scale is a major modification of the natural soil physical and chemical structure, modulated locally by the underlying soil type and ongoing management activities (e.g., nutrient

Table 2
Two-way analysis of variance (ANOVA) results comparing soil characteristics, carbon stocks and carbon cycle processes across the two vegetation cover types (fields and shelterbelts) and the four farm sites, and their interaction. Presented are the ANOVA degrees of freedom (*df* – treatments, interaction and residual), mean squares (MS), and *F*-values (*F*). Significance levels are $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***); NS = non-significant.

Indicator	Indicator code	No. samples	Residual *df*	Cover type (*df* = 1) MS (*F*)	Farm site (*df* = 3) MS (*F*)	Cover type × Farm site (*df* = 3) MS (*F*)
Soil moisture (%)	Moisture	120	112	1009.2 (214.1)***	117.5 (25.0)***	38.6 (8.2)***
pH	pH	120	112	14.4 (343.6)***	7.2 (172.3)***	0.7 (15.0)***
Bulk density (g·cm^3)	BD	120	112	NS	1.1 (113.6)***	0.6 (62.9)***
Total carbon (t·ha^{-1})	TC	40	32	1951.9 (21.2)***	869.1 (9.5)***	650.8 (7.1)***
Total nitrogen (t·ha^{-1})	TN	40	32	NS	2.1 (5.0)**	2.2 (5.2)**
Organic carbon (t·ha^{-1})	TOC	120	112	NS	492.3 (14.8)***	161.1 (4.8)**
Permanganate oxidisable carbon (t·ha^{-1})	POXC	40	32	NS	0.12 (6.6)**	NS
Dehydrogenase enzyme activity (ug·g^{-1}·hr^{-2})	DHA	60	52	170.4 (135.5)***	26.2 (20.8)***	NS
Bait-lamina probes (% bait removed)	BLP	40	32	42.0 (23.5)***	NS	NS
Tea bag index – *k* parameter	TBI	40	32	42.0 (23.5)***	NS	NS

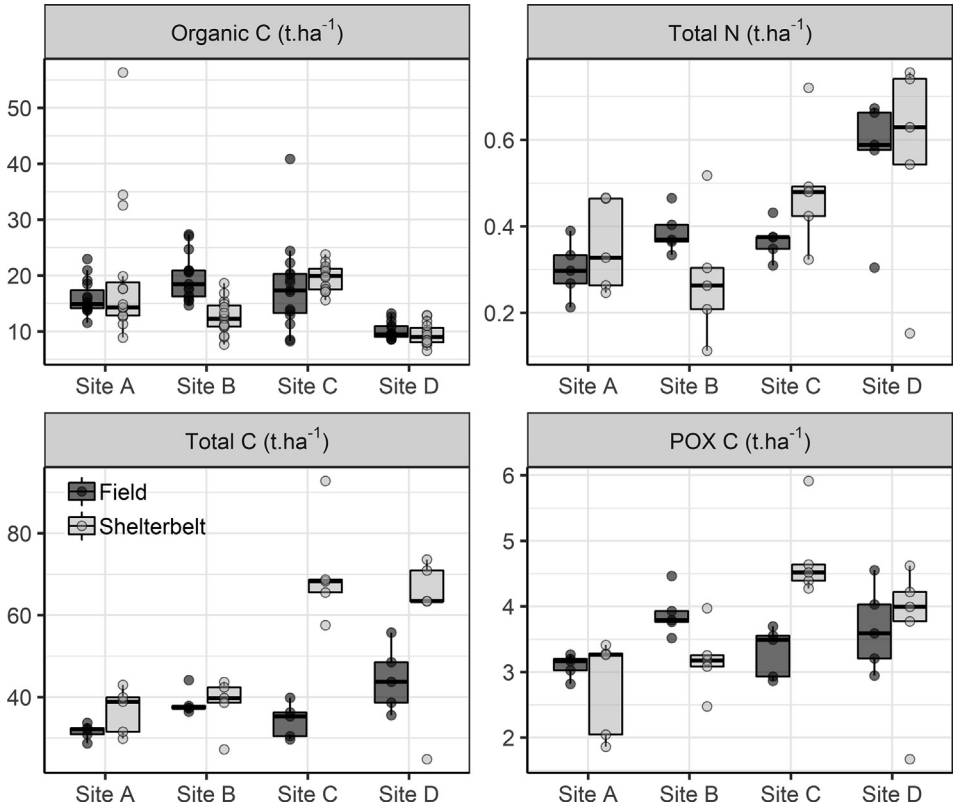

Fig. 5. Boxplot graphs showing variation in the four measured soil carbon and nitrogen measures (organic C, POX C, total C, and total N) within and between fields and shelterbelts across the four farm sites.

addition) associated with a given farming system (Stark et al., 2008). These effects, combined with the incorporation of woody vegetation elements such as shelterbelts and hedgerows on farms, generates spatial variability in soil attributes and processes at different scales. This study illustrates how: (a) variability results in differing levels of confidence in soil indicator measurements depending on the sampling design, intensity, and spatial structure of variation within a landscape, (b) shelterbelt soil indicator values differ from field soils, suggesting a modifying effect of above-ground woody vegetation; and (c) different soil indicators can vary substantially depending on the influence of aboveground woody vegetation and underlying substrate differences.

There were significant differences in a range of measured soil indicators among sheep farm locations, despite historically being subjected to similar management regimes, as well as between pine shelterbelts and adjacent fields on these farms. Field soils showed that measured soil indicators separated clearly in multivariate space by site, potentially in association with the differing underlying soil types among the sites. While we were unable to test for effects of shelterbelt age or size, due to lack of replication, soil indicator measurements under shelterbelts were nonetheless grouped in terms of the two younger (< 25 years), smaller (2.5 m wide) shelterbelts with less acidic and more biologically-active soils versus older (*c.* 60 years), wider (> 4.5 m wide) shelterbelts characterised by more acidic soils with higher total carbon, permanganate oxidizable carbon and total nitrogen quantities. The significance of ANOVA interaction terms further highlighted that variation in soil indicator values can be partitioned based on a combination of the effects of substrate and aboveground woody vegetation, or lack of, on soil processes. Our results support those of previous studies showing that the planting of semi-permanent woody vegetation elements, such as shelterbelts, can act to modify local ecological processes contributing to soil development and changes in nutrient cycling, thereby causing gradual shifts in the underlying soil environment over time (e.g., Dhillon and Van Rees, 2017, Nair et al., 2009). Indeed, soil biological function indicators suggested higher decomposition and

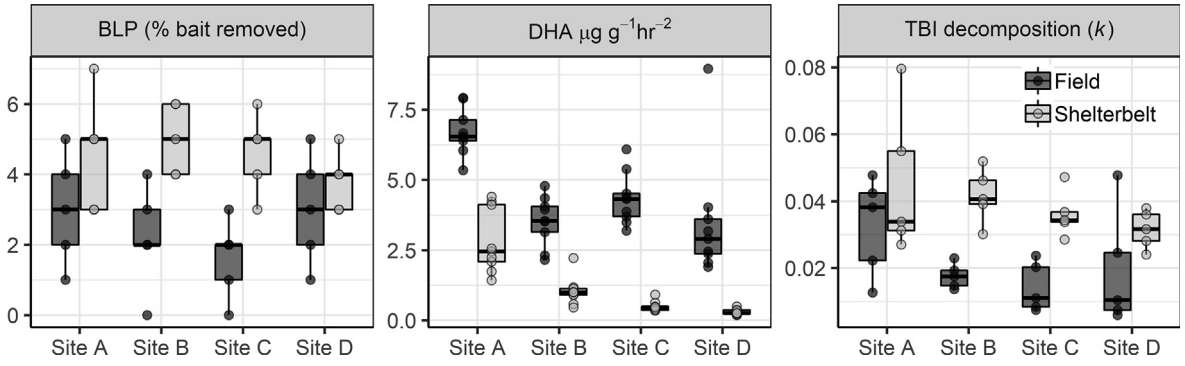

Fig. 6. Boxplot graphs showing variation in the three measured soil function indicators (BLP invertebrate activity, DHA, and TBI decomposition) within and between fields and shelterbelts across the four farm sites.

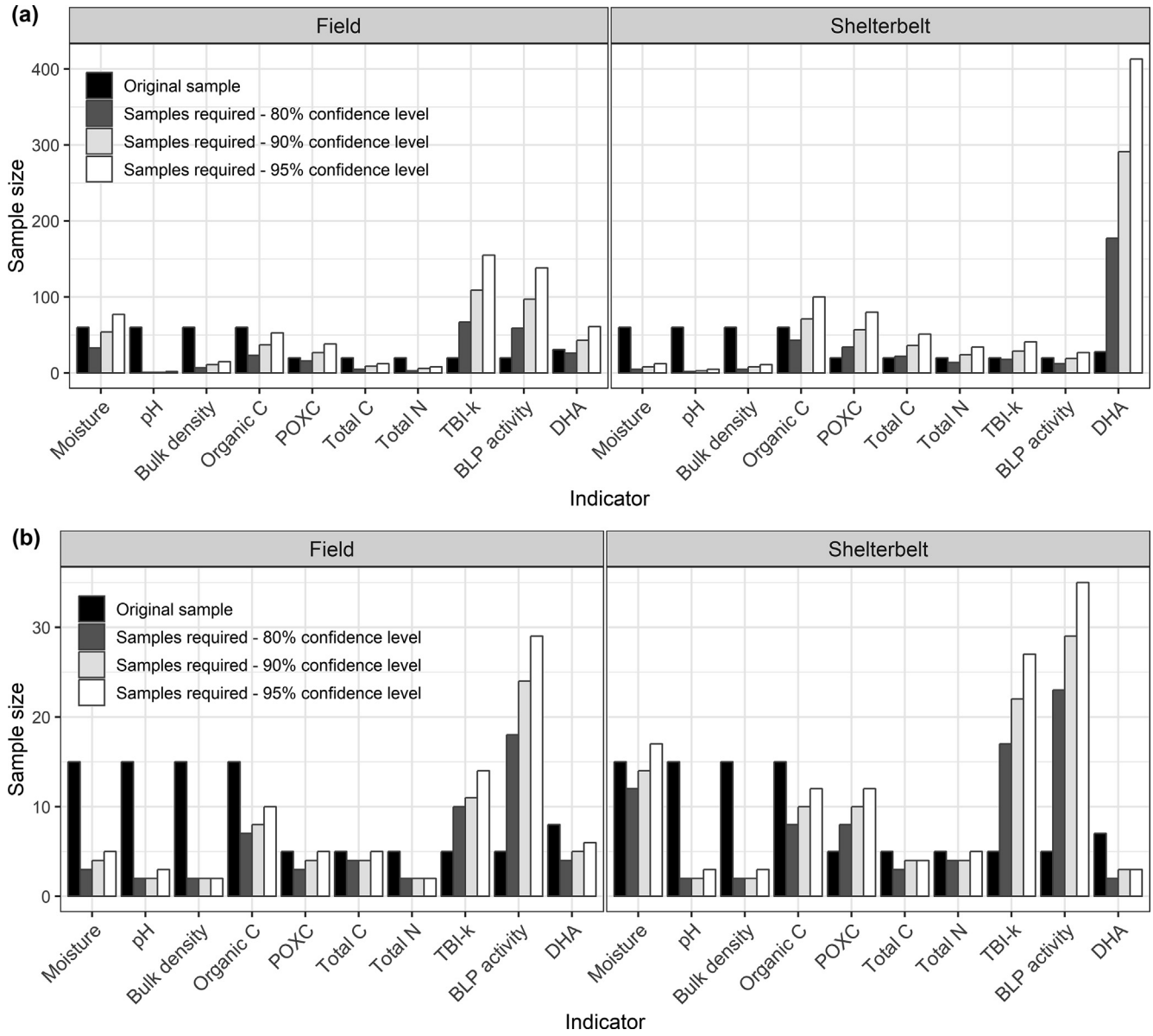

Fig. 7. Bar graphs showing the number of original samples taken for each soil indicator (black bars) and the predicted number of samples required to: (a) estimate the population mean for soil indicators in shelterbelt and field soils to within 10% of measured mean values, and (b) detect significant differences in indicator values in shelterbelt and field soils among sites. Results are presented for three levels of confidence: 80% (grey bars), 90% (black bars), and 95% (white bars).

micro-invertebrate activity rates in shelterbelt soils relative to field soils; DHA values were significantly higher in field soils, probably reflecting a stimulation of the soil microbial community via inputs of fertiliser and/or animal manure in production zones (e.g. O'Donnell et al., 2001) relative to shelterbelts.

Soils are naturally heterogeneous, yet soil sampling efforts do not always formally recognise or account for such variability (Grunwald, 2016). Our results indicate a high level of spatially-structured soil variability, with clear knock-on impacts for appropriate sampling effort, depending on the soil indicator in question, the original soil sampling density, and whether the aim was to estimate population mean values or to detect differences among sites with confidence. Generally, population means for moisture, bulk density, and pH, and the differences in these indicators among sites, could be estimated with confidence with fewer samples than used in our study design, as reflected by their relatively smaller measured coefficients of variation. Soil carbon and nitrogen measures were more variable across shelterbelt sites compared to field soils, likely due to tree-related effects (roots, above-ground inputs, shading, etc.) on soil development processes, as mentioned above. This resulted in a requirement for marginally larger sample sizes for these indicators, particularly in terms of estimating population mean values with confidence for shelterbelt soils. Indictors of biological function showed the greatest sensitivity to high inherent variability in terms of estimated sample size requirements, likely because invertebrate and microbial processes are operating at highly-localised scales (O'Brien et al., 2016). Results for these process indicators suggest that field soils may be even more variable than shelterbelt soils in terms of soil activity measures, likely as a consequence of management (e.g., fertilisers) and animal-related (e.g., urine) inputs (e.g., Moir et al., 2006, Hendrickson and Sanderson, 2017).

Natural variability, and low sample sizes, can impact on the ability to accurately upscale soil quantities, such as carbon and nitrogen, to obtain regional totals. For example, our results suggested that our sampling regime was inadequate for estimating mean shelterbelt soil total carbon with confidence in our case study landscape, based on measured natural variation and the requirement that estimates be

Table 3
The main drivers of variability for the ten soil indicators examined in this study, and indicative samples sizes for estimating population mean values for each indicator to within 10% of the true mean with a confidence level of 95% or higher. The figures presented are based on the sites investigated in this study and we recommend that pilot studies always be carried out to inform sampling designs in any new situation.

Soil indicator	Main influences on indicator variability	Indicative sample size range (n)	
		Field soils	Shelterbelt soils
Moisture	Soil physical properties, rainfall, sampling method, and time of year sampled	70–130	10–20
pH	Soil physico-chemical properties, vegetation, management practices	2–5	5–10
Bulk density	Soil physical properties, vegetation, management practices	15–20	10–15
Organic C	Soil physico-chemical properties, moisture (rainfall, topography), vegetation, and management practices	50–90	100–180
POX C	Soil physico-chemical properties, temperature, moisture, vegetation, management practices	40–60	80–100
Total C	Soil physico-chemical properties, vegetation, soil temperature and moisture, management practices	10–20	50–90
Total N	Soil physico-chemical properties, vegetation, management practices	10–15	30–60
TBI	Soil physico-chemical properties, soil moisture and temperature, vegetation, management practices	150–250	40–70
DHA	Soil physico-chemical properties, soil moisture and temperature, vegetation, management practices	60–100	400–700
BLP	Soil physico-chemical properties, soil moisture and temperature, vegetation, management practices	130–230	30–50

within 10% of mean values. Indeed, based on our current number of samples and design, we estimate that we would have to accept a margin of error of *c.* 50% to ensure that we are capturing the mean value with 95% confidence. The resultant accuracy band at this margin of error ($51 \pm 25.5\,t\,ha^{-1}$) compared to, e.g., a 10% error margin ($51 \pm 5.1\,t\,ha^{-1}$) could potentially be problematic if used as a basis for decision making or inference. The typical upscaling procedure for carbon accounting purposes would involve taking a mean shelterbelt (or field) soil carbon quantity estimate in tonnes per hectare and multiplying this value by the total area for that feature type in a landscape. Using this procedure, the range of shelterbelt carbon estimates for our study extent, with 5% of the area comprising shelterbelts, and based on our current accuracy, could range anywhere from 5.9 to 17.6 Mt of shelterbelt C. Thus, at the current level of confidence in the mean estimate, it is evident how carbon accounting efforts could lack reliability and result in unsubstantiated inferences or conclusions. This issue is also relevant for the use of soil biological process indicators, which were highly variable in this study, for monitoring and for making inferences about differences in ecological function across a landscape or among features in a landscape.

While our study covers a relatively small range of possible agroecosystem conditions, we propose in Table 3 some general sample size guidelines to estimate population mean values for the soil indicators examined. We reiterate that the number of samples required will ultimately depend on the amount of variability inherent to the agroecosystem being sampled, as influenced by a number of factors including: substrate and topographic heterogeneity, ongoing and historical management impacts, and the nature of the vegetation cover (Table 3). Thus, we present these sample sizes as a starting point only. Nonetheless, our study clearly demonstrates the utility of carrying out a pilot study in advance of any large sampling effort, to facilitate understanding of variability in soil function indicators within a given agroecosystem and how this variability is partitioned spatially within and among vegetated features. Such information is critical for informing the *a priori* development of appropriate field sampling designs and intensities, leading to greater confidence in results from investigations of agro-ecosystem multi-functionality.

Acknowledgements

The authors acknowledge funding support provided to BSC by the Brian Mason Science and Technology Trust, and to JW by the James Bruce Smith Memorial Scholarship and Lincoln University. The authors would also like to thank the four farm landowners for permission to carry out the research on their properties.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecolind.2019.02.065.

References

Alef, K., 1995. Dehydrogenase activity. In: Alef, K., Nannipieri, P. (Eds.), Methods in Applied Soil Microbiology and Biochemistry. Academic Press, London.
Bispo, A., Cluzeau, D., Creamer, R., Dombos, M., Graefe, U., Krogh, P.H., Sousa, J.P., Peres, G., Rutgers, M., Winding, A., Römbke, J., 2009. Indicators for monitoring soil biodiversity. Integrated Environ. Assess. Manage. 5, 717–719.
Blackmore, L.C., Searle, P.L., Daly, B.K., 1987. Methods for chemical analysis of soils. NZ Soil Bureau Scientific Report 80.
Blake, G.R., Hartge, K.H., 1986. Bulk Density. In: Klute, A. (Ed.), Methods of Soil Analysis. Part 1 – Physical and Mineralogical Methods, second ed. American Society of Agronomy, Madison, WI.
Champely, S., 2018. pwr: Basic Functions for Power Analysis. R package version 1.2-2. https://CRAN.R-project.org/package=pwr.
Culman, S.W., Snapp, S.S., Freeman, M.A., Schipanski, M.E., Beniston, J., Lal, R., Drinkwater, L.E., Franzluebbers, A.J., Glover, J.D., Grandy, A.S., Lee, J., 2012. Permanganate oxidizable carbon reflects a processed soil fraction that is sensitive to management. Soil Sci. Soc. Am. J. 76, 494–504.
Dhillon, G.S., Van Rees, K.C., 2017. Soil organic carbon sequestration by shelterbelt agroforestry systems in Saskatchewan. Can. J. Soil Sci. 97, 394–409.
Doran-Browne, N., Wootton, M., Taylor, C., Eckard, R., 2017. Offsets required to reduce the carbon balance of sheep and beef farms through carbon sequestration in trees and soils. Animal Product. Sci. 58, 1648–1655.
Godde, C.M., Thorburn, P.J., Biggs, J.S., Meier, E.A., 2016. Understanding the impacts of soil, climate, and farming practices on soil organic carbon sequestration: a simulation study in Australia. Front. Plant Sci. 7, 661.
Gollany, H.T., Venterea, R.T., 2018. Measurements and models to identify agroecosystem practices that enhance soil organic carbon under changing climate. J. Environ. Qual. 47, 579–587.
Grunwald, S., 2016. What do we really know about the space–time continuum of soil-landscapes? In: Environmental Soil-landscape Modeling. CRC Press, pp. 16–49.
Hendrickson, J., Sanderson, M., 2017. Perennial-Based Agricultural Systems and Livestock Impact on Soil and Ecological Services. In: Soil Health and Intensification of Agroecosytems, pp. 151–171.
Keuskamp, J.A., Dingemans, B.J.J., Lehtinen, T., Sarneel, J.M., Hefting, M.M., 2013. Tea Bag Index: a novel approach to collect uniform decomposition data across ecosystems. Methods Ecol. Evol. 4, 1070–1075.
Kort, J., Turnock, R., 1998. Carbon reservoir and biomass in Canadian prairie shelterbelts. Agrofor. Syst. 44, 175–186.
Kratz, W., 1998. The bait-lamina test: general aspects, application and perspectives. Environ. Sci. Pollut. Res. 5, 94–96.
Kremen, C., 2015. Reframing the land-sparing/land-sharing debate for biodiversity conservation. Ann. N. Y. Acad. Sci. 1355, 52–76.
Manning, P., Plas, F., Soliveres, S., Allan, E., Maestre, F.T., Mace, G., Whittingham, M.J., Fischer, M., 2018. Redefining ecosystem multifunctionality. Nat. Ecol. Evol. 2, 427.
Mize, C.W., Brandle, J.R., Schoeneberger, M.M., Bentrup, G., 2008. Ecological development and function of shelterbelts in temperate North America. In: In: Jose, S., Gordon, A.M. (Eds.), Towards Agroforestry Design: An Ecological Approach Vol. 4. Springer, Netherlands, pp. 27–53.
Moir, J.L., Fertsak, U., Cameron, K.C., Di, H.J., 2006, July. The spatial distribution and area coverage of urine depositions in grazed dairy or sheep and beef pastures in New Zealand. In: Proceedings of the 18th World Congress of Soil Science, vol. 160.
Nair, P.R., Nair, V.D., Kumar, B.M., Haile, S.G., 2009. Soil carbon sequestration in tropical agroforestry systems: a feasibility appraisal. Environ. Sci. Policy 12, 1099–1111.
Newsome, P.F.J., Wilde, R.H., Willoughby, E.J., 2000. Land resource information system spatial data layers. Landcare Research New Zealand Ltd, Palmerson North, NZ.

Norton, D., Reid, N., 2013. Nature and Farming: Sustaining Native Biodiversity in Agricultural Landscapes. CSIRO Publishing, Australia.
O'Brien, S.L., Gibbons, S.M., Owens, S.M., Hampton-Marcell, J., Johnston, E.R., Jastrow, J.D., Gilbert, J.A., Meyer, F., Antonopoulos, D.A., 2016. Spatial scale drives patterns in soil bacterial diversity. Environ. Microbiol. 18, 2039–2051.
O'Donnell, A.G., Seasman, M., Macrae, A., Waite, I., Davies, J.T., 2001. Plants and fertilisers as drivers of change in microbial community structure and function in soils. Plant Soil 232, 135–145.
R Development Core Team, 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org.
Söderström, B., Hedlund, K., Jackson, L.E., Kätterer, T., Lugato, E., Thomsen, I.K., Jørgensen, H.B., 2014. What are the effects of agricultural management on soil organic carbon (SOC) stocks? Environ. Evidence 3, 2.
Stark, C.H., Condron, L.M., O'Callaghan, M., Stewart, A., Di, H.J., 2008. Differences in soil enzyme activities, microbial community structure and short-term nitrogen mineralisation resulting from farm management history and organic matter amendments. Soil Biol. Biochem. 40, 1352–1363.
Stevenson, I.L., 1959. Dehydrogenase activity in soils. Can. J. Microbiol. 5, 229–235.
Thalmann, A., 1968. Zur methodik der bestimming der dehydrogenaseaktivitaet im boden mittels triphenyltetrazoliumchlord TTC). Landwirtschaftliche Forschung 21, 249–258.
Torne, E., 1990. Assessing feeding activities of soil-living animals I. Bait-Lamina tests. Pedobiologia 34, 89–101.
Weil, R.R., Islam, K.R., Stine, M.A., Gruver, J.B., Samson-Libig, S.E., 2003. Estimating active carbon for soil quality assessment: a simplified method for laboratory and field use. Am. J. Altern. Agric. 18, 3–17.
Welsch, J., Case, B.S., Bigsby, H., 2014. Trees on farms: investigating and mapping woody re-vegetation potential in an intensely-farmed agricultural landscape. Agric. Ecosyst. Environ. 183, 93–102.
Williams, D.R., Phalan, B., Feniuk, C., Green, R.E., Balmford, A., 2018. Carbon storage and land-use strategies in aricultural landscapes across three continents. Curr. Biol. 28, 2500–2505.
Zar, J.H., 1999. Biostatistical Analysis, fourth ed. Prentice Hall International Inc., New Jersey, USA.