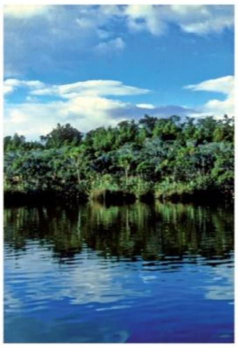
Utilising the soil quality monitoring data set: multivariate and spatial analysis of soil quality data

Envirolink Advice grants 1109-GSDC101 and 1110-ESRC250



Landcare Research
Manaaki Whenua

Utilising the soil quality monitoring data set: multivariate and spatial analysis of soil quality data

Envirolink Advice grants 1109-GSDC101 and 1110-ESRC250

Bryan Stevenson, Stephen McNeill, Allan Hewitt

Prepared for:

Gisborne District Council

P.O. Box 747
Gisborne 4010
New Zealand

And

Southland Regional Council

Private Bag 90116
Invercargill 9840
Southland

July 2012

Landcare Research, Gerald Street, PO Box 40, Lincoln 7640, New Zealand, Ph +64 3 321 9999, Fax +64 3 321 9998, www.landcareresearch.co.nz

Reviewed by:	*Approved for release by:*
Craig Ross Capability Leader Landcare Research	Alison Collins Portfolio Leader, Characterising Land Resources Landcare Research
Landcare Research Contract Report:	LC 1060

Contents

Summary

Project and Client

A method for soil quality monitoring was developed through the 500 Soils Project to meet RMA requirements for regional authorities to provide State of Environment (SoE) reporting on soils. A database of the soil quality data utilising the ‘500 Soils’ approach has recently been compiled under an Envirolink Tools grant. This database is potentially of value to smaller Regional Councils who do not have active soil quality programmes, as the soil order/land use combinations established in other regions could be applicable across regional boundaries. However, spatial statistics of the data set need to be verified and gaps in soil-order/land-use combinations assessed. To assess these needs Gisborne District Council and Southland Regional Council co-sponsored this report through two Envirolink advice grants.

Objectives

These linked projects will provide:

1. Quantitative statistical information on the “robustness” of the soil quality data set using spatial autocorrelation models.

2. Re-testing initial assumptions around soil order and land use associations generated from the initial 500 Soils Project.

3. Assessment where there are gaps in the current monitoring.

Methods

Values for the seven soil quality indicators (pH, total carbon (C) and nitrogen (N) concentration, Olsen P, anaerobically mineralisable N, bulk, macro-porosity(measured at –5 kPa tension) and bulk density), along with spatial location, land use and soil order, were extracted from 811 soil quality site records. In addition to these seven soil quality factors, the carbon-to-nitrogen ratio (C/N) was also used as a pseudo-indicator. The most recent sampling data available at time of analysis was used for sites that have been visited more than once. Spatial statistics on the individual indicators and multivariate statistical tests (principal components analysis and fuzzy logic clustering) were employed to characterise the data set as a whole.

Results and Conclusions

- Indicator means from the current dataset differed slightly from the original 500 soils data but measures of variance were similar, supporting the underlying assumptions made from variance calculations in the original dataset (i.e. sample number calculations made based on variance).
- Principal component analysis indicated that although the organic, physical, chemical, and fertility components derived in the initial dataset were still present to an extent, the distribution of soil quality indicators in the principal components is more evenly

distributed in the present dataset than in the initial 500soils analysis. The soil-order and land-use factors are still highly significant, but the variation explained by these factors is somewhat less (about 3% less for land use and about 9% less for soil order) than in the initial analysis.

- Spatial autocorrelation analysis of the dataset suggests essentially no autocorrelation for AMN and pH, medium range autocorrelation for macroporosity and Olsen P (28–40 km), and longer range autocorrelation for Total C and N (~100 km). Comparison with a larger MfE dataset indicates that the range for C (~23.4 km) was considerably shorter than our dataset and that use of multiple data sets would be preferable. The autocorrelation ranges indicate that samples taken within the range would show some degree of correlation. This does not negate the validity of sites within the range but indicates that corrections may need to be applied to statistical testing to validate assumptions that samples are independent (i.e. not correlated).
- For managed land uses (cropping, dairy pasture, drystock pasture, horticulture, and forestry) our initial cluster analyses indicated that a six-cluster configuration yielded clusters that gave meaning separation between soil order and land use interaction. The clusters formed primarily around the organic (total C, total N, AMN) and physical (bulk density and macroporosity) components of the soil quality indicators.
- The overall statistical basis of the 500 soils sampling strategy appears to be sound, though gaps exist in the stratification of soil-order/land-use combinations (particularly for Raw, Anthropic, Ultic, Podzol, Organic, Semi-arid, and Melanic Soils). Additionally, the spatial distribution of sampling sites throughout the country is patchy and non-uniform.
- The cluster analysis suggests that soil “phenotypes” exist and these groupings may prove useful in simplifying soil order/land use stratification. Further work would be needed, however, to derive optimal groupings (e.g. groupings that yield the lowest total variance).

Recommendations

- The soil quality dataset (and soil quality results from other regions particularly regarding land use trends) are valuable in assisting smaller regions in planning soil quality programs. However, there are still gaps that need be filled.
- Utilisation of other databases (NSD, soil carbon monitoring, LMI) are likely to be useful for filling gaps in land-use/soil-order interactions and spatial coverage, though conversion factors for various sampling depths would need to be derived.
- Refining the cluster analysis groupings after utilising data within other databases (above) could potentially simplify stratification of soil-order/land-use combinations and therefore the sampling strategy for smaller regional councils who do not have active soil quality programs.
- Currently, soils are grouped only by soil order. For most soil orders this appears to be sufficient, however, Brown, Recent, and Gley Soils appear to be somewhat more variable than most other soils (discounting soils orders with <15 samples). Use of subgroup designations or linking the particular soil series to S-MAP sibling attributes could potentially provide greater resolution for theses soils orders in particular.

- The ‘500 soils program’ was designed to give regions flexibility in sampling sites that were of greatest concern in that particular region, however, more coordination in selection of sites between regions would also assist in filling gaps in the land use/soil type stratification. The Land Management Forum (LMF) could assist in this role.
- The goals of the ‘500 Soils’ approach to soil quality monitoring should be reviewed in the advent a “national SOE reporting” program is announced. There are likely to be differences in scale and approach. For example, additional indicators may be desired for national reporting and selection criteria for sites (particularly of managed land use sites versus indigenous sites) could differ.

1 Introduction

To increase soil quality understanding in New Zealand a Sustainable Management Fund Project (#5089), Implementing Soil Quality Indicators for Land was initiated in 1999. The project, referred to as the "500 Soils Project", collected new soil quality data from approximately 500 sites (508 sites, roughly one site per 25 km^2) selected by the various participating Regional Authorities from April 1999 to June 2001.Before the 500 Soils Project there was no nationally consistent or scientifically based soil quality monitoring data for New Zealand. Soil quality sampling by regional councils has continued to the present and the methods developed through the 500 Soils project have been used to meet RMA requirements for regional authorities to provide State of Environment (SoE) reporting on soils (Lilburne et al. 2002, 2004; Sparling et al. 2004;Sparling &Schipper 2004).

Principal component analysis by Schipper and Sparling (2000) related the variance of indicator values to acidity, organic resource, physical and chemical components. Sparling et al. (2004) also used an analysis of variance approach to estimate the proportion of total variance explained by soil order (12–49%), land use (21–39%), while the total variation explained by the factors (and interactions) was 50–68%. Consequently, the SINDI system for interpretation of soil quality indicator values (Sparling et al. 2003, http://sindi.landcareresearch.co.nz/) used land-use type and soil order to categorize soil quality indicator target values.

A database of the soil quality data utilising the 500 Soils approach (including the original 500 soils data and subsequent sampling) has recently been compiled under an Envirolink Tools grant. This dataset may be of value to smaller Regional Councils who do not have active soil quality programmes, as the soil-type/land-use combinations established in other regions could be applicable across regional boundaries. Progress has been made in understanding the variation of key dynamic soil properties by soil quality researchers, and it had been envisioned that periodic review of the 500 soils programme should be undertaken (Hill et al. 2003). In this study we utilise a number of different multivariate analysis and clustering approaches to examine the expanded soil quality dataset.

2 Objectives

Using the soil quality dataset:

- Compile quantitative statistical information on the "robustness" of the soil quality data set using spatial autocorrelation models.
- Retest initial assumptions around soil type and land use associations generated from the initial 500 Soils project.
- Assess where there are gaps in the current monitoring.

3 Methods

Values for the seven soil quality indicators (pH, total carbon (C), total nitrogen (N), Olsen P, anaerobically mineralisable N, bulk density, and macroporosity), along with spatial location, land use, and soil type were extracted from the soil quality database. There were 809 site records available at the time of analysis. For indicators measured gravimetrically, the data were expressed on area (total C and total N) or volumetric (AMN, Olsen P) basis. Macroporosity measured at –5 kPa was used in this analysis as there were sites that did not have the –10 kPa macroporosity measurement. Though technically not an indicator, the C/N ratio was also included for two reasons. First, it is easily derived from total C and total N, and is widely accepted as a useful indicator of organic matter condition. Second, in the 500Soils dataset, C and N are strongly correlated (correlation 0.81), while the C/N ratio is essentially uncorrelated with C (correlation 0.08) and only moderately correlated with N (correlation –0.44), so the C/N ratio provides useful additional information over and above that provided by C and N separately.

We analysed the SINDI soil quality dataset in R(R Development Core Team2012). Data from the latest available sampling date were used for sites that have been visited more than once, and records from previous visits were excluded. Where applicable, values were expressed on a volumetric (or area basis for C and N) to facilitate comparison to the earlier statistical analysis of the 500 soils dataset (Hill et al 2003, Sparling et al. 2004). Additionally, records were only retained for analysis if all seven indicators were available. Transformations were carried out on Olsen P, anaerobically mineralisable N (AMN), and macroporosity (MP) to avoid the strong skewing of the data towards small values. A log-transformation was applied to Olsen P, and square-root transformations to AMN and MP, and the transformed quantities were quite close to Gaussian distributions. Although the subsequent analysis does not strictly require Gaussian-distributed variables, the analysis is simpler and generally more reliable if the analysed quantities have a more-or-less symmetric distribution.

For Olsen P, one zero value was replaced by the minimum of the non-zero values, to avoid difficulties with the log-transformation. The data were split into a spatial and an aspatial dataset, with 722 and 794 rows in each. The spatial dataset, which is a subset of the aspatial dataset, was used for spatial autocorrelation analysis, while the aspatial dataset was used for the cluster and principal component analysis.

To examine spatial autocorrelation, empirical and model semi-variograms were fitted to the spatial data for each (possibly transformed) quantity. For the model variograms, only an exponential shape (with a nugget) was considered, as trying an alternative was not thought to provide any obvious advantage.

Principal component analysis (Manly 2005) was used for three reasons. First, it was hoped that the variability of the various quality factors would be largely explained by a few of the principal components (PCs), thus providing a slightly more economical way of expressing the relationships. Second, the PC analysis provides information concerning which indicators provide the greatest variation in the dataset, and gives some clues as to their meaning. Finally, the PCs provide a simplified way to describe the variation of soil and land-use classes.

In order to investigate whether the soil records exhibited natural grouping we used cluster analysis, which aims to partition the soil observations into one of a fixed number of clusters.

In this project we used fuzzy *c*-means clustering (Bezdek 1981), which is a method of natural data clustering very similar to the *k*-means clustering algorithm (Lloyd 1982), but which has better convergence properties. In fuzzy clustering, each point has a degree of membership of belonging to all clusters (as in fuzzy logic) so that points at the edge of a cluster will have a lower degree of membership when compared with points in the centre of a cluster. The fuzzy *c*-means method requires an initial estimate of the number of clusters, begins with a random assignment of points to clusters, and also requires a fixed fuzzification factor (greater than 1) that describes the degree of membership variability of the points in the cluster. In this study, we tested candidate fuzzy *c*-means clustering for 4–7 clusters, with a range of different fuzzification factors.

4 Results

4.1 Basic statistics and distribution of soil types

Table 1 presents mean and variance measures for the current dataset and the original 500 Soils dataset. Mean indicator values differ to some extent but variance parameters are generally similar for the two data sets. Additionally, current distribution of sites by soil order and land use is shown in Appendix 1, and mean and variance measures by soil order shown in Appendix 2.

Table 1 Basic statistics for the current soil quality dataset and the original 500 soils dataset

	pH	Tot C	Tot N	lg C/N[1]	sr AMN[1]	lg Olsen P[1]	BD	sr MP[1]
		Mg ha^{-1}	Mg ha^{-1}		ug cm^{-3}	ug cm^{-3}	Mg m^{3}	%
Current data set								
Mean	5.81	54.2	4.42	2.53	10.17	3.06	0.94	3.52
				(12.5)	(103)	(21.3)		(12.4)
CV (%)	9.4	37.8	39.9	9.8	32.0	35.4	27.8	34.0
Se	0.02	0.72	0.06	0.01	0.11	0.04	0.01	0.04
lower quartile	5.45	40.09	3.00	2.37	8.25	2.38	0.74	2.67
Original 500 Soils date set								
Mean	5.73	53.4	4.23	2.56	10.16	2.87	0.89	3.69
				(12.9)	(103)	(17.6)		(13.6)
CV (%)	9.1	37.9	41.5	9.2	28.4	37.9	27.5	31.3
Se	0.02	0.90	0.08	0.01	0.13	0.05	0.01	0.05
lower quartile	5.41	39.33	2.81	2.40	8.28	2.14	0.72	2.80

[1]For transformed (lg-log, sr-square root) indicators the back-transformed mean is shown in parenthesis.

Distribution of sites in the original 500 soils dataset by land use was 12%, 24%, 27%, 13%, 8%, 11%, for cropping, dairy, drystock, forestry, horticulture, and indigenous forest respectively, compared with 15%, 24%, 26%, 14%, 10%, 9% in the current dataset.

Distribution of sites in the original 500 soils dataset by soil order was 22%, 18%, 7%. 5%, 2%, 2%, 8%, 2%, 11%, 18%, 5% for Allophanic, Brown, Gley, Granular, Melanic, Organic, Pallic, Podzol, Pumice, Recent, and Ultic soils respectively compared with 19%, 20%, 8%, 5%, 2%, 2%, 12%, 2%, 9%, 17%, 4% in the current dataset.

4.2 Spatial Analysis

Most statistical analyses assume that samples are uncorrelated. Spatial analysis calculates the range under which there is some degree of correlation between samples. Empirical and model semi-variograms were fitted to the spatial data for each (possibly transformed) quantity. For the model semi-variograms, only the exponential shape (with nugget) was considered, as it was not thought there was any obvious advantage trying an alternative. Semi-variograms were successfully fitted to all factors, except for bulk density (BD), which is curious, as statistically it is usually a well-behaved parameter.

Figure 1 shows the empirical and model semi-variograms for six parameters, using 5km bins. The results suggest significant spatial correlation for C, N, log(Olsen P) and sqrt(MP). There is perhaps some evidence of spatial autocorrelation for sqrt(AMN) and pH, but the autocorrelation distance for both is so short that the estimates may be unreliable. In these latter cases, it would be prudent to reduce the bins size to (say) 1 km, to get better resolution towards zero point-to-point distance. This reduction in bin size causes an increase in scatter, which makes the variogram difficult to interpret, so some subjective trade-off is required. A subsequent analysis with 1-km bin size for sqrt(AMN) and pH suggested that the autocorrelation distance is no more than a few kilometres for pH and sqrt(AMN). However, as these estimates are based on a relatively small number of samples with small point-to-point distance, the autocorrelation distance estimate would be subject to high uncertainty.

The values for the variogram fits are shown in Table 2. Note that the value for the correlation range for carbon (97.8 km) is considerably larger than the figure calculated using a combination of NSD, MfE soils and LMI data of 24.1 km (McNeill 2012). Presumably the much larger dataset in the latter study (with approximately three times the number of samples as the 500-soils) provides better mathematical support for short point-to-point distances, and probably yields values of higher accuracy for the spatial autocorrelation distance. Also note that the figures for sqrt(AMN) and pH were for a much smaller bin size than was used for the other quality factors, and the estimates are likely to have high uncertainty.

Given the short correlation range distances for sqrt(AMN) and pH (1.9 and 2.2 km respectively) when compared with the other quality factors, and the likely uncertainty of these estimates, it can be assumed that these two factors are essentially spatially uncorrelated. This conclusion would not be true for C and N concentration, Olsen P or MP.

Table 2 Summary of parameters from the spatial autocorrelation analysis, showing the parameters for an exponential kernel with nugget effect

Parameter	Nugget	Sill	Range (km)	Bin size (km)
Tot C	252	153	97.8	5
Tot N	1.99	1.58	95.6	5
sqrt(AMN)	2.52	4.18	1.85	1
Log(Olsen P)	0.624	0.707	40.2	5
pH	0.097	0.168	2.19	1
sqrt(MP)	0.867	0.587	28.4	5

It was not possible to fit a satisfactory semi-variogram to bulk density. The empirical semi-variogramusing5-km bins in Figure 2 shows what could be interpreted as short-distance autocorrelation, plus unusual large-distance correlation as well. This type of empirical semi-variogram is not well described by the exponential spatial kernel, or by any other commonly used kernel shape.

A better picture of the spatial behaviour for bulk density can be obtained by calculating the semi-variogram for eight different directions from 0 to 157.5 degrees (22.5 degree steps between 0 and 180 degrees). The result in Figure 3 shows considerable variation in the semi-variogram shape for large point-to-point distances, but reasonably consistent behaviour for point-to-point distances up to 100 km. However, it was not possible to fit a semi-variogram by excising point-to-point distances above 200km, suggesting that either the exponential model was faulty, and/or the spatial autocorrelation distance was too short for the available data. The variation in the semi-variogram distances of 250km or more for different directions suggests some complex contrasting relationship between regional groups of soil data. However, the semi-variogram is calculated using the limited 500-soils dataset. A much larger dataset would be required to determine if these regional contrasts were consistent, or merely an artefact of the dataset. In any case, the impression from the semi-variogram plots of Figure 3at very small point-to-point distances is that there is little evidence of spatial autocorrelation.

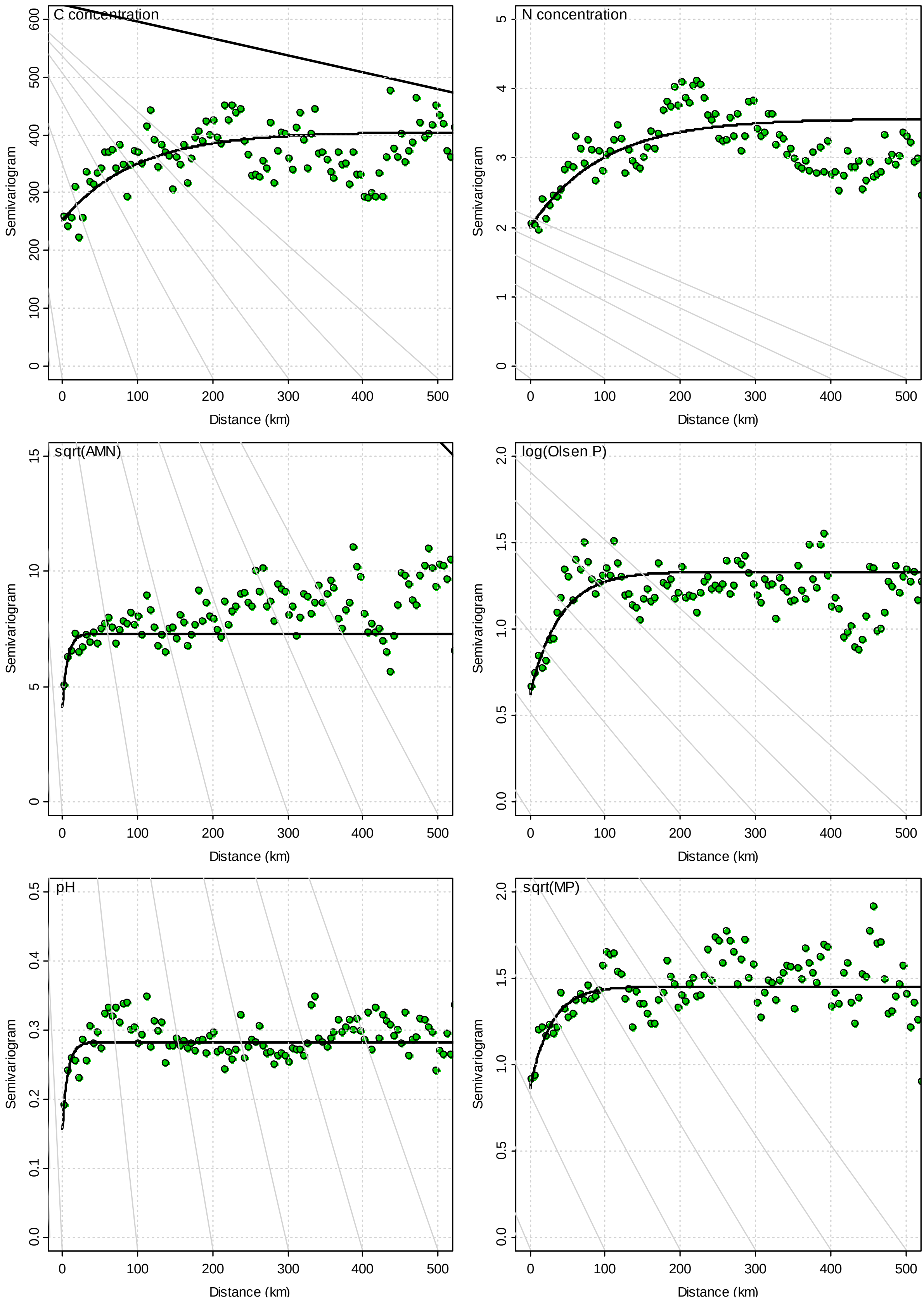


Figure 1 Semi-variograms for six of the soil quality factors (excluding bulk density). In each case, an exponential kernel was fitted with a nugget, and a 5-km bin size was used.

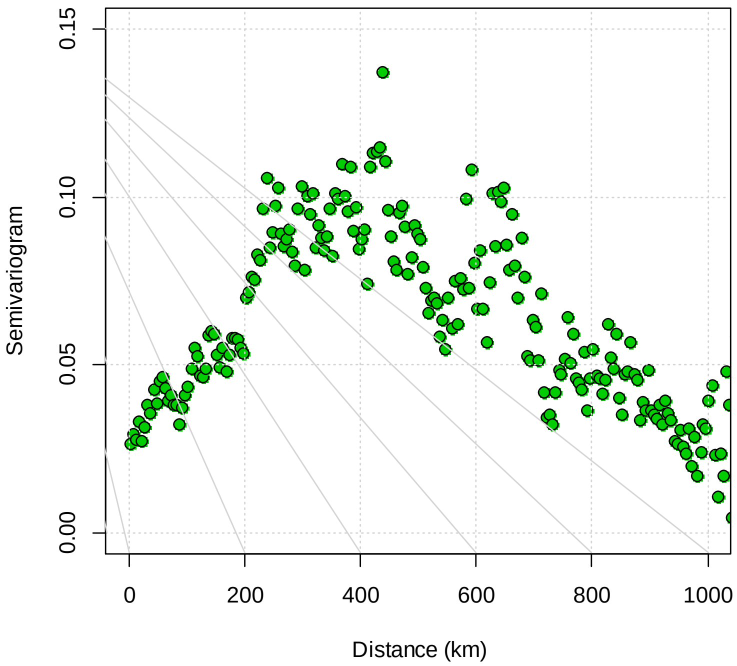


Figure 2 Plot of the empirical semi-variogram for bulk density, using 5-km bins.

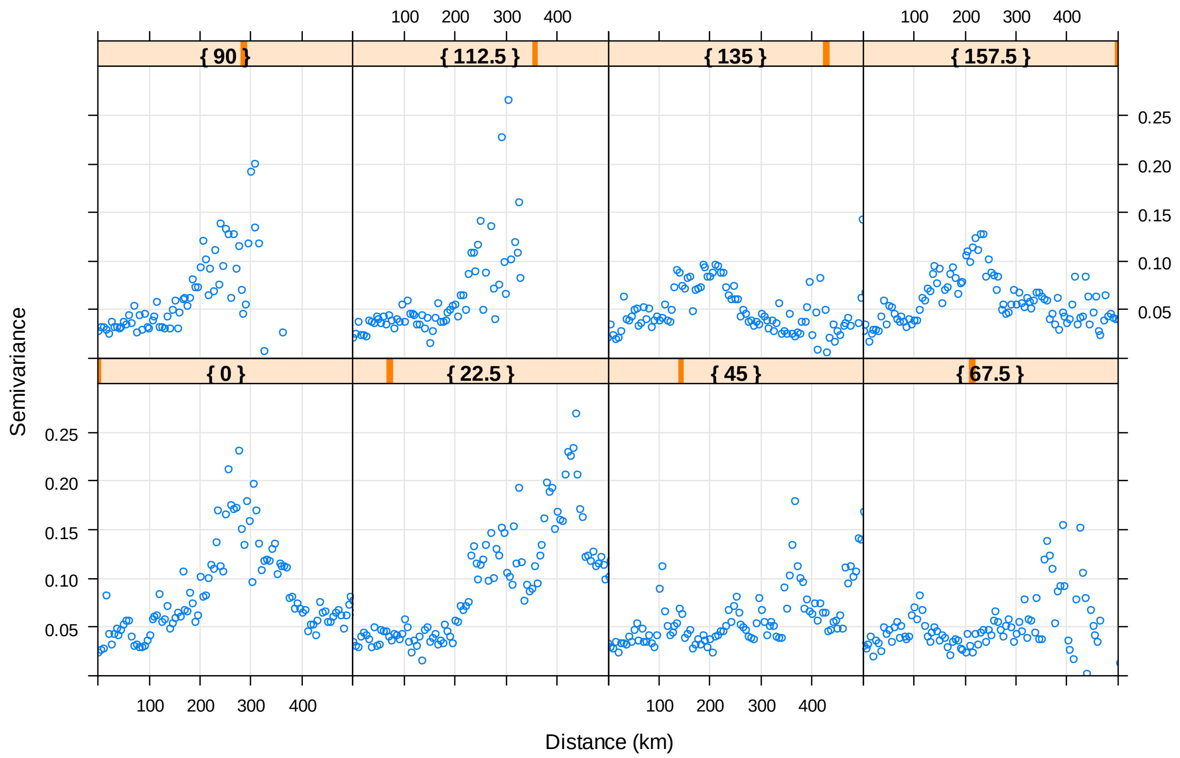


Figure 3 Plots of the empirical semi-variance of soil bulk density, for eight different directions from 0 to 180 degrees (22.5 degree steps). In each case, the orientation in degrees of the semi-variogram calculation is shown in the panel title.

4.3 Principal Components (multivariate analysis)

The objective of principal component analysis is to take the eight soil quality parameters (pH, C and N concentration, log(C/N ratio), log(Olsen P), sqrt(AMN), BD, sqrt(MP)), and find linear combinations of these to produce indices $z_1 \cdots z_8$(principal components) that are uncorrelated and arranged in decreasing order of importance of variation. In principal component analysis of the original 500 soils data, the soil quality indicators grouped within the principal components corresponded to organic (total C, total N, AMN), physical (bulk density and macroporosity), fertility (Olsen P) and chemical (pH) aspects of the soil.

By convention, since all the quality factors are measured with different scales, the quality factors are centred (the mean value is removed) and scaled by the standard deviation of the samples. The linear factors that are used to form the PCs are shown in Table 3 (all sites including indigenous and tussock) and Table 4 (managed sites only), while plots of the proportion of variance associated with each PC, and the cumulative total variance (for managed sites) are shown in Figure 4. The principal components analysis performed on the initial 500 soils data (Sparling et al. 2004) showed distinct components related to soil organic (total C, total N, AMN), fertility (Olsen P), physical (bulk density, macroporosity) and chemical (pH) attributes. Although these elements are still present to an extent, our analysis of the current dataset indicated a more complex, and less dominant mixture of the primary soil quality indicators within each component. Because of the complex mixture of components, axis rotation (e.g. Varimax) was not performed in the present analysis.

For all sites and for managed sites only, approximately 90% of the total variation is explained by the first five of the eight PCs, so the decomposition does not drastically reduce the dimensionality of the eight-variable dataset by much. However, the first two PCs are notably dominant when compared with the other six PCs. The first PC can be interpreted as a contrast between C and N concentration, log(Olsen P) and sqrt(AMN) (with pH contributing a minor effect) on one hand, against log(C/N ratio) and sqrt(MP) on the other. Bulk density appears to contribute little to the effect of this first PC. The second PC could be interpreted as a contrast between C concentration, N concentration, and log(C/N ratio) in one group, and pH, log(Olsen P), and BD in a second group. The effect of sqrt(AMN) and sqrt(MP) is quite small for this PC. Similar interpretations could be attached to higher order PCs, although they would have less relevance as the PC order is increased. It is interesting to note that the contribution of each factor to the PCs varies considerably. C concentration has a consistently strong effect for all PCs, while the effect of N concentration, log(C/N ratio), log(Olsen P), and sqrt(AMN) is somewhat less pronounced. BD has a strong effect in only a few PCs, and other factors have effects that differ from one PC to another.

Comparing the two sets of linear factors in Table 3and Table 4 (all sites versus only managed sites respectively) it appears that most of the coefficients are broadly the same. A notable exception is PC 4, which has significantly different factors applied to each of the quality factors; presumably these differences for PC 4 reflect the modest influence of indigenous sites.

Table 3 Coefficients of the principal components for eight soil quality factors on all sites

		Principal component number							
		1	2	3	4	5	6	7	8
Soil quality factor	pH	0.267	-0.366	-0.507	-0.106	-0.710	-0.141	-0.046	0.021
	Cconc	0.243	0.534	0.053	-0.402	-0.218	0.112	0.262	-0.603
	Nconc	0.444	0.372	-0.129	-0.038	0.019	0.273	0.297	0.695
	logCNratio	-0.410	0.223	0.345	-0.473	-0.348	-0.398	-0.096	0.388
	logOlsenP	0.377	-0.214	-0.151	-0.606	0.527	-0.317	-0.211	0.012
	sqrtAMN	0.403	0.233	0.176	0.474	-0.025	-0.725	0.022	-0.047
	BD	0.130	-0.528	0.470	-0.105	-0.037	-0.030	0.686	0.001
	sqrtMP	-0.428	0.118	-0.575	0.000	0.217	-0.329	0.563	-0.014

Table 4 Coefficients of the principal components for eight soil quality factors on managed sites

		Principal component number							
		1	2	3	4	5	6	7	8
Soil quality factor	pH	0.193	-0.403	-0.597	-0.099	-0.644	0.135	-0.021	0.019
	Cconc	0.319	0.480	-0.105	0.368	-0.222	-0.142	0.280	-0.611
	Nconc	0.490	0.307	-0.106	-0.024	0.012	-0.269	0.304	0.699
	logCNratio	-0.382	0.280	0.091	0.586	-0.370	0.382	-0.079	0.366
	logOlsenP	0.328	-0.275	-0.336	0.558	0.554	0.241	-0.167	0.015
	sqrtAMN	0.439	0.151	0.288	-0.314	-0.022	0.774	0.020	-0.052
	BD	0.028	-0.546	0.409	0.242	-0.085	0.017	0.684	0.001
	sqrtMP	-0.416	0.193	-0.500	-0.206	0.292	0.293	0.571	-0.017

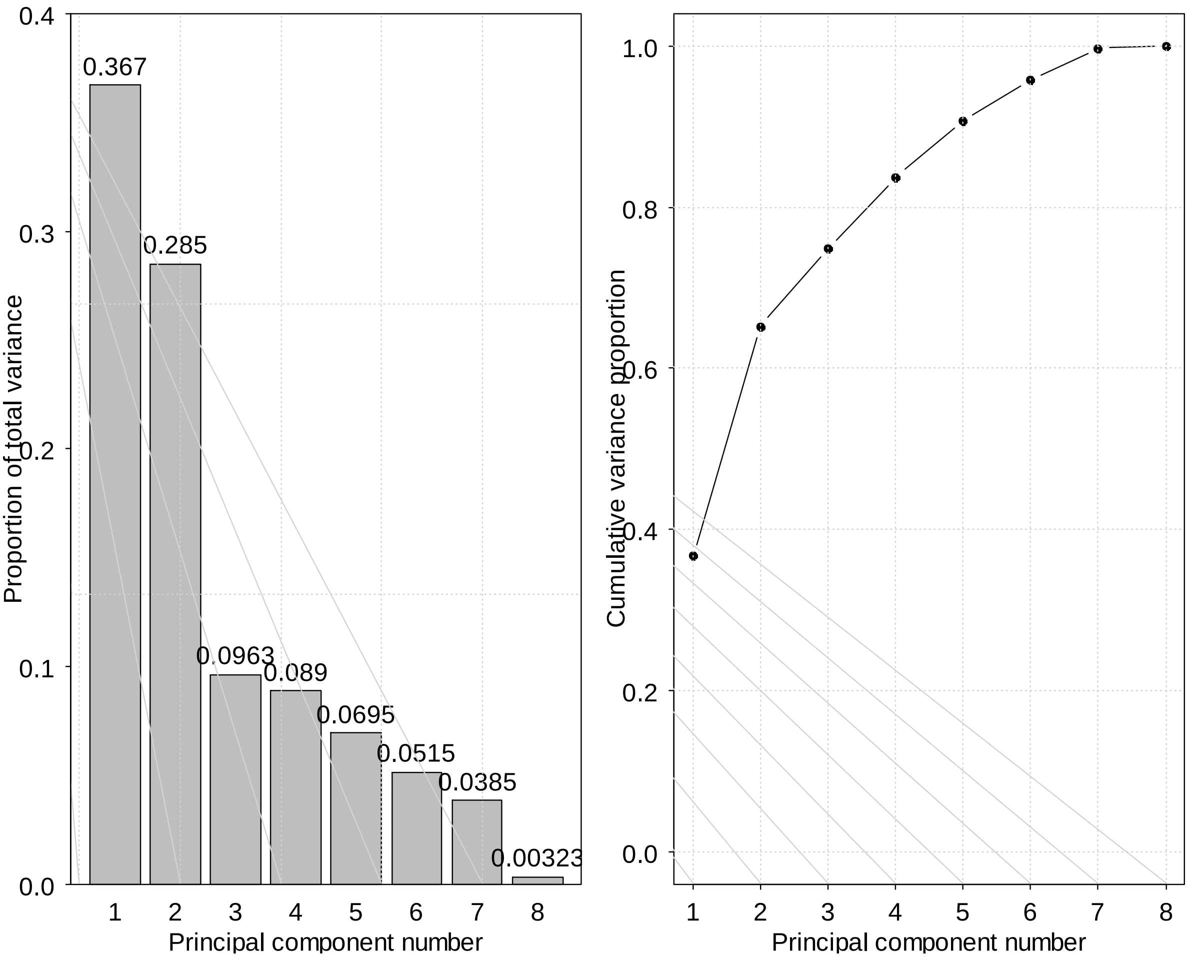


Figure 4 (Left) Bar plot of the proportion of the total variance explained by each principal component (managed sites only, see Table 4), with the proportion printed above each bar. (Right) Cumulative plot of the proportion of variance explained by principal components.

4.4 Variability as a function of land use and soil type

For each soil quality factor, it is of interest to know the degree to which the total variability is explained by land use, soil order, and the interaction between these two factors. This is a type of analysis of variance (ANOVA), where the principal assumption is that the residuals are Gaussian distributed. In this case, the distribution of the ANOVA residuals was tested by eye using Gaussian quantile plots, and in most cases the residuals conformed closely to an assumption of Gaussian residuals. The exception to this was C concentration, where there was a tendency for the residuals to have a long tail; this perhaps indicates that a transformation might have been required. In any case, the resultant F-value from the ANOVA was sufficiently high that it is unlikely the conclusions from the analysis would have changed.

Table 5 (all sites) and Table 6 (managed land-use sites only) provide values for the proportion of variance explained by each of land use, soil order, and their interaction, as well as the combination of all these terms. The last of these figures corresponds to the unadjusted R^2 of the corresponding linear model for a soil factor response, with land use, soil order, and their interaction as explanatory variables.

Table 5 For all sites, the proportion of variance associated with each soil quality factor that is explained by land use, soil order, or the interaction between land use and soil order. The furthest right column is the proportion of variance explained by all terms

	Land use	Soil order	Land use: Soil order	All terms
pH	0.349	0.063	0.060	0.471
Tot C	0.174	0.270	0.064	0.508
Tot N	0.266	0.183	0.062	0.510
log(CN ratio)	0.303	0.107	0.055	0.465
log(Olsen P)	0.358	0.053	0.057	0.469
sqrt(AMN)	0.356	0.066	0.064	0.486
BD	0.181	0.402	0.042	0.625
sqrt(MP)	0.276	0.088	0.056	0.420

Table 6 For managed sites, the proportion of variance associated with each soil quality factor that is explained by land use, soil order, or the interaction between land use and soil order. The furthest right column is the proportion of variance explained by all terms

	Land use	Soil order	Land use:Soil order	All terms
pH	0.356	0.068	0.067	0.491
Tot C	0.160	0.269	0.071	0.499
Tot N	0.289	0.182	0.063	0.534
log(CN ratio)	0.360	0.100	0.059	0.519
log(Olsen P)	0.417	0.051	0.057	0.525
sqrt(AMN)	0.339	0.077	0.066	0.481
BD	0.193	0.405	0.044	0.642
sqrt(MP)	0.269	0.092	0.052	0.412

Both land use and soil order variables were highly significant (p-value< 0.001). The range in total variance explained by all factors (0.42 – 0.63 for all sites) was similar but somewhat lower than analysis by Sparling et al. (2004) of 0.50 – 0.68. Change in the range of variance due to land use (0.21 – 0.39 in Sparling et al. (2004), compared with 0.17 – 0.36 in this data set) was generally smaller than changes in variance due to the soil order (0.12 – 0.49 in Sparling et al. (2004)) compared with 0.06 – 0.40 in the current dataset. The soil-order factor was greater than the land-use factor for total C and bulk density, whereas the land-use factor was greater for pH, Tot N, C/N ratio, Olsen P, and macroporosity. This is similar to the original 500 soils analysis, with the exception of total N, which was nearly equal in land-use and soil-order factors in that analysis.

4.5 Cluster Analysis

The objective of a cluster analysis is to use the (possibly transformed) values of the soil quality factors to devise a scheme for grouping the soil-order/land use combinations into classes so that similar combinations are in the same group. In general, the method used for clustering must be numerically based, and the true number of classes is not known beforehand. Since the number of clusters is somewhat arbitrary, this parameter must be specified before the analysis. There is usually some information that can guide this selection. For instance, the PC analysis described in Section4.3 suggests that at least five PCs are required to describe 90% of the total variability and only two PCs describe just under two-thirds of the total variability. These results suggest we might need as many as five or six groups to cover the vast majority of the points adequately; two would probably be too few. In practice, we tried as few as four clusters, and in later trials we tried up to 17. Cluster sizes greater than 7 appeared to offer no compelling advantages.

Using the fuzzy c-means approach also meant adopting a specific value for the fuzzification factor (a number greater than a value of one), which helps control the shapes of clusters and the manner in which points are allocated to cluster groups. We used a fuzzification factor of 2, which was the default value for the fuzzy c-means software we used (Dimitriadou et al. 2011), but found that larger or smaller values made very little difference to the result.

Specifying four clusters separated out Organic Soils, while five clusters split higher C soils on dairy and drystock. Six clusters separated out lower C soils (primarily Recent Soils) on cropping, which is a useful grouping. Our preliminary analysis indicated that using more than six clusters did not appear to break out meaningful clusters that could be interpreted on a soil-order or land-use basis, but this could be investigated further. Previous analysis had indicated that indigenous sites could be classified as largely separate from managed sites but indigenous sites can be quite variable (Giltrap & Hewitt 2004).It also was found that managed sites clustered better without indigenous sites included. Mean indicator values for the six cluster analysis are shown in Table 7.

Table 7 Mean values for the six clusters selected by fuzzy c-means clustering, for the managed sites

Cluster	pH	Tot C Mg ha^{-1}	Tot N Mg ha^{-1}	C/N ratio	AMN µg cm^{-3}	Olsen P µg cm^{-3}	BD Mg m^{-3}	MP %
1	6.13	25.1	2.19	11.8	53	27.1	1.19	14.6
2	5.90	39.0	3.22	12.6	83	19.7	1.05	13.1
3	5.84	52.8	4.53	11.9	124	26.1	0.94	12.0
4	5.84	67.6	5.75	12.0	144	26.9	0.88	10.4
5	5.65	86.1	7.08	12.4	150	26.7	0.79	8.8
6	5.72	137.0	7.67	18.2	105	28.9	0.55	9.6

Because the SINDI soil quality dataset was not designed to be an unbiased sample of land use classes and soil orders, it is not possible to reach absolute conclusions about the characteristics of the clusters. However, the relationships suggest that clusters do have functional meaning. Differentiation between clusters was greatest for the organic (total C, total N, AMN) and physical (bulk density, macroporosity) indicator properties that are less directly manageable. The pH and Olsen P indicators showed little variation between clusters.

As with the principle components analysis, this suggests a somewhat lesser impact for pH and Olsen P in the current dataset than was measured in the original data set. This may indicate that productive land uses are being managed more carefully for Olsen P and pH (through nutrient and liming management programs), or it may simply reflect the (generally small) change in distribution of land uses or even external factors such as the price of P fertilisers.

For most clusters, there were minor components of many soil orders and/or land uses present (Table 8). This can be explained in part by the general assumption that sites have been under the current land use for enough time for all properties to represent that particular land use; and also, in part, by the wide variation in land-use management (particularly for cropping and horticulture) where some specific management practices (e.g. manure or compost applications) would have large impacts on soil quality indicators. Specific soil orders, land uses or land-use/soil-order interactions, however, were generally more prominent within the various clusters. We make the following generalisations about clusters:

- Cluster 1 (*n*=74) contained the lowest soil C values of any group. It was largely represented by Recent Soils on cropping and drystock and also by Gley Soils on cropping.
- Cluster 2(*n*=168) had moderately low soil C values. It was represented most prominently by drystock (Brown, Pallic, Recent Soils) and cropping (most soils but particularly Pallic Soils).There were also significant components of Brown, Pallic, and Recent Soils in forestry. Soil quality values (in comparison with other groups) were characterised by relatively low TC and TN, moderately low AMN, and adequate BD and MP.
- Cluster 3 (*n*=219) had moderate soil C values. Soil orders were generally similar in this cluster to cluster2 but land-use distribution was shifted from crop/drystock/forestry (in cluster 2) to dairy/drystock/forestry, though Allophanic Soils on cropping was split between cluster 2 and 3. Interestingly, C/N ratio was lowest in this cluster.
- Cluster 4 (*n*=152) had moderately high soil C and was also represented by all land uses, but is dominated by dairy and dry stock on Allophanic, Ultic, Granular, and Brown Soils. Cluster 4 is characterised by moderately high TC and TN, high AMN, and adequate BD and MP.
- Cluster 5 (*n*=70) contained the highest soil C values for mineral soils. It was dominated by dairy and drystock, on Allophanic Soils but also had a component of Brown Soils on dairy and drystock. The mean bulk density in this group (0.79) was less than that in cluster 4, also suggesting a higher Allophanic Soils component than in cluster 4. AMN was high, BD is relatively low, and MP is low. These levels reflect the properties of Allophanic soils, in which allophane or other short-range-order minerals and associated soil organic matter promote low bulk density. Low MP values would not normally be expected for Allophanic Soils, but they are commonly observed in Allophanic Soils under dairy use as an indication of intensive grazing by heavy animals.
- Cluster 6 (*n*=11) was dominated by Organic Soils on most land uses. TC and TN are high in comparison with other groups as would be expected from Organic Soils, AMN is adequate, and BD low, which is characteristic of Organic Soils. As for cluster 5, the MP value is low, which is also not normally expected for Organic Soils. Again, low MP is common for Organic Soils under intensive dairy use with related to treading of heavy animals.

Droogers and Bouma (1997) have provided a useful conceptual framework to integrate inherent and dynamic soil properties. Borrowing from plant and animal ecology, they coined the term "genoform" for soil formed under native vegetation and "phenoform" for the equivalent soil with similar inherent properties but with dynamic properties modified by the impacts of a specific land use history. The reasoning behind this approach is that soil classifications and spatial soil databases that deliberately exclude dynamic soil properties are limited in their ability to support realistic spatial analyses of land-use issues involving dynamic soil properties. The clusters indicate that the land-use history of some soil orders may result in a similar "phenoform" that could be useful in the soil quality context.

The grouping of soil orders shown in Table 8 was derived from similarity of the mean total C values (see Appendix 2). Statistics for this grouping of soils (Organic, Allophanic, Granular/Ultic, Pumice/Gley/Brown/Melanic, Pallic/Podzol, and Recent) are shown in Appendix 3. Previously, Sparling et al. (2003) had groupings based on the lower quartile of total C of Organic, Allophanic, Brown/Granular/Melanic/Ultic, Gley/Pallic/Podzol/Recent; statistics for this grouping are also shown in Appendix 2. Comparison of the variance components of the groups is inconclusive and does not strongly indicate which grouping would yield the lesser variance.

Table 8 Cluster membership by land use and soil order groups

Cluster	Soil Order group	Crop	Dairy	Drystock	Forestry	Hort
1	Allophanic	0	0	1	2	0
	Granular, Ultic	3	0	0	0	0
	Organic	0	0	0	0	0
	Pumice, Gley, Brown, Melanic	15	0	3	4	2
	Pallic, Podzol	9	0	0	2	2
	Recent	14	2	9	0	4
2	Allophanic	6	0	0	3	1
	Granular, Ultic	7	0	0	0	1
	Organic	0	0	0	0	0
	Pumice, Gley, Brown, Melanic	11	9	26	22	9
	Pallic, Podzol	15	0	12	5	4
	Recent	4	3	13	9	8
3	Allophanic	6	5	6	2	11
	Granular, Ultic	1	3	5	6	4
	Organic	0	0	0	0	0
	Pumice, Gley, Brown, Melanic	5	38	34	15	6
	Pallic, Podzol	4	2	18	5	8
	Recent	3	15	8	4	3
4	Allophanic	3	28	11	6	7
	Granular, Ultic	0	11	11	4	0
	Organic	1	0	0	0	1
	Pumice, Gley, Brown, Melanic	1	17	13	11	4
	Pallic, Podzol	1	7	2	2	1
	Recent	1	4	2	2	1
5	Allophanic	0	17	9	2	1
	Granular, Ultic	0	0	4	2	0
	Organic	0	2	0	0	0
	Pumice, Gley, Brown, Melanic	3	10	8	4	0
	Pallic, Podzol	1	2	1	0	0
	Recent	1	1	0	0	2
6	Allophanic	0	1	0	0	0
	Granular, Ultic	0	0	0	0	0
	Organic	1	5	1	0	1
	Pumice, Gley, Brown, Melanic	0	1	0	1	0
	Pallic, Podzol	0	0	0	0	0
	Recent	0	0	0	0	0

5 Conclusions

The 500 Soils approach was a pragmatic programme to develop soil quality monitoring for regional and national SOE reporting. The basic assumptions of the programme appear to be sound. Indicator means from the current dataset differ somewhat from the original data but measures of variance were similar. This supports the underlying assumption that the inherent variation in soil quality indicators could be captured in a national dataset of about 500 samples. Additionally, the principal component analysis indicates that the general groupings of organic, physical, chemical, and fertility components are, to some extent, still present, but soil quality indicators in the principal components are more evenly distributed in the current dataset than in the initial 500Soils analysis. Soil-order and land-use factors are still highly significant factors, but the variation explained by these factors is somewhat less (about 3% less for land use and about 9% less for soil order) than in the initial analysis.

Spatial autocorrelation analysis of the dataset suggests essentially no autocorrelation for AMN and pH, which exemplifies within paddock variation due to phenomena such as urine spots; medium-range autocorrelation for macroporosity and Olsen P, exemplified by farm-scale phenomena such as fertilizer regime; and longer range autocorrelation for Total C and N due to environmental gradients. The autocorrelation range for C and N (about 100 km in this dataset) was considerably greater than results of a larger data set (McNeill 2012) of 24.1 km, and illustrates that different datasets can be combined to increase the coverage and statistical robustness of measurements. Shorter distance between samples does not negate those sampling points, but indicates there will likely be some correlation between samples (e.g. they are not uncorrelated, which many statistical models assume) and this may need to be taken into account when statistical analyses are performed.

Several general land use trends have emerged from regional soil quality monitoring. Dairy sites often have high nutrient levels and low macroporosity. Drystock sites can show divergent characteristics, with the more intensive sites similar to dairy whereas low intensity sites (i.e. marginal sites) can have relatively low nutrient and fertility values. Horticultural and cropping sites can have high nutrient levels, but declining C can also be a problem in cropping. Because of the smaller number of sites analysed for most regions, it has generally not been feasible to analyse soil-order/land-use combinations. Even at a national level, it has been acknowledged that obtaining statistically robust stratification for all land-use/soil-order interactions would be difficult if not impossible (Hill et al. 2003).Thus part of the uncertainty of the 500 Soils approach is balancing the proportions of soil orders and land uses on a national level with adequate statistical robustness of soil-order/land-use stratification. For instance, Brown Soils occupy approximately 43% of the land area on a national basis (Sparling et al. 2004) but compose only about 20% of the current soil quality monitoring sites. On the other hand, Allophanic and Recent Soils (covering about 5–6% of the land area) make up about 18 and 17% of soil quality sites respectively, although Allophanic Soils in particular are heavy utilised in managed land uses and considered among our most productive soils.

Though Brown Soils are under-represented by land area, the number of sites sampled (in excess of 150 sites) does give statistical power in contrasting land-use effects within the soil order (though distribution of land uses is not balanced). Allophanic and Recent Soils also have reasonably high sample numbers (~150 samples for Allophanic and ~135 samples on Recent), yet even within these groups land use distribution varies considerably. Gley, Granular, Pumice, and Pallic Soils are intermediate, with 65–100 samples. Raw, Anthropic, Podzol, Organic, and Melanic Soils all have less than 15 samples each, and Ultic Soils ~30 total samples. Land areas for these soil orders are relatively low (~3% for Ultic and Raw Soils and 1% or less for Granular, Melanic, Organic, and Anthropic). There are no Semi-arid or Oxidic Soils currently in the database. It could be argued that the majority of these soils have little importance to managed land uses, so a decision must be made on whether to sample these soils of lesser coverage.

The cluster analysis may prove useful in simplifying soil-order/land-use stratification, though further analysis is needed to ensure the chosen clusters are the optimal configuration (i.e. result in the least variance among groups). The cluster analysis suggests that soil "phenotypes" exist and these groupings may prove useful in the context of soil quality. For instance, the cropping land use on Recent and Gley Soils was predominant in cluster 1 (the cluster with the lowest C content), indicating these soils in particular are at risk for low soil C content. Clusters for managed land uses (cropping, dairy, drystock, forestry, and horticulture) vary primarily inorganic (Total C, total N, and AMN) and physical (bulk density, macroporosity) components, rather than chemical (pH) and fertility (Olsen P) measures though, in part, this may be due to the method of clustering.

Regional councils have been the prime agents responsible for implementing SOE reporting; thus the system has been developed to address the major soil quality concerns within the regions, and these concerns are primarily on managed lands. While this is obviously advantageous for regional reporting, the lack of central leadership does have some drawbacks for adequate stratification of soil-order/land-use combinations. The geographical coverage of monitoring is an issue as spatial coverage is not uniform throughout New Zealand (particularly in the South Island). There is much greater density of sites in regions where soil quality monitoring has been most active (Auckland, Waikato, and Wellington regions, which when combined make up nearly half the soil quality sites).Additionally, percentage of non-managed land uses (native forest and tussock grassland, for example) is considerably below the national occurrence. While the soil quality status of native sites is often not computed (target values are not well defined for native sites), these sites form a baseline from which to compare the effects of land-use management, and native site condition it its own right would likely form an important aspect of national soil quality reporting. Other data sets (e.g. NSD, CMS/soil C monitoring data) may be useful in filling some of these gaps, but scaling factors to compensate for the various sampling depths would need to be derived.

6 Recommendations

- The soil quality dataset (and soil quality results from other regions particularly regarding land use trends) are valuable in assisting smaller regions in planning soil quality programs. However, there are still gaps that need be filled.
- Utilisation of other databases (NSD, soil carbon monitoring, LMI) are likely to be useful for filling gaps in land-use/soil-order interactions and spatial coverage, though conversion factors for various sampling depths would need to be derived.
- Refining the cluster analysis groupings after utilising data within other databases (above) could potentially simplify stratification of soil-order/land-use combinations and therefore the sampling strategy for smaller regional councils who do not have active soil quality programs.

- Currently, soils are grouped only by soil order. For most soil orders this appears to be sufficient, however, Brown, Recent, and Gley Soils appear to be somewhat more variable than most other soils (discounting soils orders with <15 samples). Use of subgroup designations or linking the particular soil series to S-MAP sibling attributes could potentially provide greater resolution for theses soils orders in particular.
- The ‘500 soils program’ was designed to give regions flexibility in sampling sites that were of greatest concern in that particular region, however, more coordination in selection of sites between regions would also assist in filling gaps in the land use/soil type stratification. The Land Management Forum (LMF) could assist in this role.
- The goals of the ‘500 Soils’ approach to soil quality monitoring should be reviewed in the advent a “national SOE reporting” program is announced. There are likely to be differences in scale and approach. For example, additional indicators may be desired for national reporting and selection criteria for sites (particularly of managed land use sites versus indigenous sites) could differ.

7 References

Dimitriadou E, Hornik K, Leisch F, Meyer D, Weingessel A. 2011. e1071: Misc Functions of the Department of Statistics (e1071), TU Wien. R package version 1.6, http://CRAN.R-project.org/package=e1071.

DroogersP,Bouma J 1997.Soil survey input in exploratory modelling of sustainable soil management practices. Soil Science Society of America Journal 61:1704–1710.

Giltrap D, Hewitt AE. 2004. Spatial variability of soil quality indicators in New Zealand soils and land uses. NZ Journal of Agricultural Research 47:167–177.

Hill RB, Sparling G, Frampton C, Cuff J 2003. National soil quality review and programme design. Environmental reporting: signposts for sustainability. Technical Paper No 75, Land. Wellington, Ministry for the Environment. 34p.

Lilburne L, Sparling G, Schipper L 2004. Soil quality monitoring in New Zealand: development of an interpretative framework. Agriculture ecosystems and environment 104: 535–544.

Lloyd SP 1982. Least squares quantization in PCM. IEEE Transactions on Information Theory 28(2):129–137.

Manly BFJ 2005, Multivariate statistical methods:aprimer.Boca Raton, FL, Chapman & Hall/CRC.

McNeill S 2012.Respecification and reclassification of the MfE Soil CMS model.Landcare Research Contract Report LC975. Wellington, Ministry for the Environment. 29p.

R Development Core Team 2010.R: a language and environment for statistical computing.Vienna, Austria,R Foundation for Statistical Computing.http://www.R-project.org/

Sparling GP, Lilburn L, Vojvodic-Vucovic M 2003.Provisional targets for soil quality indicators in New Zealand.Palmerston North, Landcare Research. 60 p.

Sparling GP, Schipper LA 2002.Soil quality at a national scale in New Zealand. Journal of Environmental Quality 31: 1848–1857.

Sparling GP, Schipper LA 2004.Soil quality monitoring in New Zealand: trends and issues arising from a broad-scale survey. Agriculture Ecosystem & Environment 104: 545–552.

Sparling GP, Schipper LA, Bettjeman W, Hill R 2004.Soil quality monitoring in New Zealand: practical lessons from a 6-year. Agriculture Ecosystems & Environment104: 523–534.

Appendix 1 – Distribution of number of sites by soil order and land use of soil quality sites

The table below gives the distribution of number of sites by soil type and land use

Soil Order	Dairy	Drystock	Forestry	Crop	Horticulture	Indigenous	Tussock	Other	Sum	(%)
Allophanic	60	35	15	16	20	11			157	19.4
Brown	30	43	35	10	9	20	12		159	19.7
Gley	19	14	2	19	9	2		2	67	8.3
Granular	8	14	1	10	3	6			42	5.2
Melanic		5	2	5		1			13	1.6
Organic	7	1	0	2	2	2			14	1.7
Pallic	10	31	8	30	15	3			97	12.0
Podzol	1	2	6			3			12	1.5
Pumice	26	23	18	1	3	3			74	9.1
Recent	25	32	15	23	18	17	4	1	135	16.7
Ultic	6	6	11	1	2	6			32	4.0
Raw/Anthropic		3	2					2	7	0.9
Sum	192	209	115	117	81	74	16	5	809	100%
(%)	23.7	25.8	14.2	14.5	10.0	9.1	2.0	0.6	100%	

Note: The percentages in this table may not add to exactly 100%, due to rounding of figures.

Appendix 2 – Statistical summaries for soil quality factors, by soil order

The tables below provide summary statistics for the eight soil quality factors, using transformed values where applicable. In each case, the summaries are by soil order, where the soil order groupings are as follows:

- Organic
- Allophanic
- Granular/Ultic (GU)
- Pumice/Gley/Brown/Melanic (PGBM)
- Pallic/Podzol (PP)
- Recent

Mean for soil quality indicators by soil order

Soilorder	pH	Cconc	Nconc	log(CNratio)	sqrt(AMN)	log(OlsenP)	BD	sqrt(MP)
Allophanic	5.89	65.1	5.70	2.46	11.33	2.88	0.745	3.40
GU	5.75	60.9	4.58	2.60	10.84	2.94	0.943	3.39
Organic	5.70	116.1	6.72	2.88	9.65	3.30	0.517	3.22
PGBM	5.71	52.5	4.16	2.56	10.15	3.08	0.913	3.73
PP	5.88	47.3	4.15	2.45	10.44	3.06	1.101	3.27
Recent	5.96	43.4	3.55	2.51	10.07	3.30	1.127	3.40

Standard deviations for soil quality indicators by soil order

Soilorder	pH	Cconc	Nconc	log(CNratio)	sqrt(AMN)	log(OlsenP)	BD	sqrt(MP)
Allophanic	0.489	16.8	1.70	0.190	2.64	1.058	0.165	1.190
GU	0.666	16.8	1.52	0.276	3.07	1.154	0.157	0.991
Organic	0.723	35.1	2.14	0.354	2.17	0.665	0.191	1.357
PGBM	0.528	17.6	1.61	0.253	2.92	1.133	0.217	1.211
PP	0.566	14.1	1.39	0.232	2.99	1.022	0.284	1.165
Recent	0.483	16.5	1.40	0.210	2.75	1.072	0.224	1.142

Coefficient of variation for soil quality indicators by soil order

Soilorder	pH	Cconc	Nconc	log(CNratio)	sqrt(AMN)	log(OlsenP)	BD	sqrt(MP)
Allophanic	0.0831	0.258	0.297	0.0773	0.233	0.367	0.221	0.350
GU	0.1159	0.276	0.332	0.1060	0.283	0.392	0.166	0.292
Organic	0.1267	0.302	0.318	0.1228	0.225	0.201	0.368	0.422
PGBM	0.0925	0.335	0.388	0.0989	0.288	0.369	0.238	0.325
PP	0.0962	0.298	0.334	0.0945	0.287	0.334	0.258	0.357
Recent	0.0811	0.380	0.394	0.0838	0.273	0.325	0.199	0.336

The second group of statistics are for the following soil order groupings:

- Allophanic
- Brown/Granular/Melanic/Ultic (BGMU)
- Gley/Pallic/Podzol/Recent (GPPR)
- Organic

Mean for soil quality indicators by soil order

Soilorder	pH	Cconc	Nconc	log(CNratio)	sqrt(AMN)	log(OlsenP)	BD	sqrt(MP)
Allophanic	5.89	65.1	5.70	2.46	11.33	2.88	0.745	3.40
BGMU	5.70	55.7	4.35	2.58	10.40	2.85	0.963	3.62
GPPR	5.93	46.7	3.94	2.48	10.36	3.30	1.088	3.29
Organic	5.70	116.1	6.72	2.88	9.65	3.30	0.517	3.22

Standard deviations for soil quality indicators by soil order

Soilorder	pH	Cconc	Nconc	log(CNratio)	sqrt(AMN)	log(OlsenP)	BD	sqrt(MP)
Allophanic	0.489	16.8	1.70	0.190	2.64	1.058	0.165	1.19
BGMU	0.586	17.3	1.68	0.270	3.20	1.144	0.172	1.16
GPPR	0.507	17.3	1.45	0.209	2.81	1.029	0.253	1.12
Organic	0.723	35.1	2.14	0.354	2.17	0.665	0.191	1.36

Coefficient of variation for soil quality indicators by soil order

Soilorder	pH	Cconc	Nconc	log(CNratio)	sqrt(AMN)	log(OlsenP)	BD	sqrt(MP)
Allophanic	0.0831	0.258	0.297	0.0773	0.233	0.367	0.221	0.350
BGMU	0.1027	0.311	0.385	0.1049	0.308	0.402	0.178	0.320
GPPR	0.0855	0.371	0.369	0.0844	0.272	0.312	0.233	0.340
Organic	0.1267	0.302	0.318	0.1228	0.225	0.201	0.368	0.422

Finally, tables of the summary statistics for each individual soil orders:

Mean for soil quality indicators by soil order

Soilorder	pH	Cconc	Nconc	log(CNratio)	sqrt(AMN)	log(OlsenP)	BD	sqrt(MP)
Anthropic	6.82	21.79	1.91	2.41	6.87	2.97	1.470	2.67
Brown	5.62	53.62	4.23	2.58	10.22	2.74	0.964	3.80
Melanic	6.38	50.04	4.42	2.43	9.98	3.58	1.061	2.86
Gley	5.95	52.29	4.33	2.48	10.79	3.69	0.989	3.12
Allophanic	5.89	65.15	5.70	2.46	11.33	2.88	0.745	3.40
Pumice	5.57	51.06	3.82	2.63	9.46	3.12	0.717	4.31
Granular	5.97	62.77	5.01	2.52	11.24	3.23	0.941	3.44
Organic	5.70	116.11	6.72	2.88	9.65	3.30	0.517	3.22
Pallic	5.98	47.78	4.35	2.40	10.71	3.28	1.175	3.06
Recent	5.96	43.41	3.55	2.51	10.07	3.30	1.127	3.40
Ultic	5.46	58.48	4.02	2.71	10.31	2.57	0.946	3.33
Raw	5.98	8.65	0.42	3.04	1.72	2.19	1.230	6.10
Podzol	5.11	43.21	2.54	2.85	8.20	1.23	0.498	4.92

Standard deviations for soil quality indicators by soil order

Soilorder	pH	Cconc	Nconc	log(CNratio)	sqrt(AMN)	log(OlsenP)	BD	sqrt(MP)
Anthropic	1.005	9.71	0.638	0.1120	0.752	0.603	0.155	0.345
Brown	0.500	17.39	1.767	0.2728	3.298	1.131	0.176	1.198
Melanic	0.571	12.59	1.364	0.1527	2.743	0.982	0.182	1.114
Gley	0.449	21.90	1.499	0.1591	2.605	0.826	0.229	0.979
Allophanic	0.489	16.79	1.696	0.1897	2.638	1.058	0.165	1.190
Pumice	0.493	14.38	1.392	0.2698	2.242	1.146	0.163	1.120
Granular	0.651	19.85	1.515	0.2215	3.081	1.213	0.165	1.153
Organic	0.723	35.09	2.139	0.3537	2.169	0.665	0.191	1.357
Pallic	0.499	13.41	1.262	0.1814	2.998	0.773	0.194	0.965
Recent	0.483	16.48	1.400	0.2103	2.750	1.072	0.224	1.142
Ultic	0.575	11.56	1.354	0.3039	3.022	0.967	0.148	0.739
Raw	0.127	1.04	0.017	0.0674	1.821	0.531	0.297	0.638
Podzol	0.504	19.10	1.345	0.2203	1.808	0.984	0.149	1.376

Coefficient of variation for soil quality indicators by soil order

Soilorder	pH	Cconc	Nconc	log(CNratio)	sqrt(AMN)	log(OlsenP)	BD	sqrt(MP)
Anthropic	0.1473	0.446	0.3340	0.0465	0.109	0.203	0.106	0.129
Brown	0.0891	0.324	0.4179	0.1058	0.323	0.413	0.183	0.316
Melanic	0.0895	0.252	0.3082	0.0628	0.275	0.274	0.171	0.389
Gley	0.0755	0.419	0.3461	0.0642	0.241	0.224	0.231	0.314
Allophanic	0.0831	0.258	0.2973	0.0773	0.233	0.367	0.221	0.350
Pumice	0.0884	0.282	0.3645	0.1026	0.237	0.368	0.228	0.260
Granular	0.1090	0.316	0.3025	0.0879	0.274	0.376	0.175	0.335
Organic	0.1267	0.302	0.3183	0.1228	0.225	0.201	0.368	0.422
Pallic	0.0835	0.281	0.2900	0.0755	0.280	0.235	0.165	0.315
Recent	0.0811	0.380	0.3940	0.0838	0.273	0.325	0.199	0.336
Ultic	0.1054	0.198	0.3368	0.1121	0.293	0.376	0.157	0.222
Raw	0.0213	0.120	0.0404	0.0221	1.059	0.243	0.241	0.105
Podzol	0.0985	0.442	0.5301	0.0774	0.220	0.800	0.300	0.280