Geoderma 239–240 (2015) 135–142

Contents lists available at ScienceDirect

Geoderma

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

Characterising soil quality clusters in relation to land use and soil order in New Zealand: An application of the phenoform concept

B.A. Stevenson [a,*], S. McNeill [b], A.E. Hewitt [b]

[a] Landcare Research, Private Bag 3127 Hamilton 3240, New Zealand
[b] Landcare Research, PO Box 69040 Lincoln 7640, New Zealand

ARTICLE INFO

Article history:
Received 27 June 2014
Received in revised form 29 September 2014
Accepted 1 October 2014
Available online 16 October 2014

Keywords:
Dynamic soil properties
Fuzzy-c means clustering
Genoform
Soil natural capital
Soil classification

ABSTRACT

The multivariate character of seven dynamic soil properties from a national soil quality data set was explored to determine if generalizations can be made about the status of the properties from land use and soil order. The genoform–phenoform concept (where soil phenoforms arise from a genoform due to modification of dynamic soil properties through specific land use history) was used to frame three hypotheses. Hypothesis one proposed that managed sites were distinct from native sites. This was supported by discriminant analysis and permutational multivariate analysis of variance. Hypothesis two proposed that managed sites were clustered into statistically significant distinct classes. This was supported by principal components fuzzy-c means clustering, with recognition of five to seven statistically significant clusters. Hypotheses three proposed that the clusters had functional meaning. This was supported by inspecting the clusters for rational relationships between land use, soil order and soil quality status as estimated by indicator mean values for each cluster. While organic status (e.g., soil C and N) appeared to be the primary driver of clustering, other soil quality indicators (such as macroporosity) were also important in differentiating the effects of land use and soil type on cluster patterns. The results indicate that a taxonomy of phenoforms is possible, but would require input of both inherent and dynamic soil properties. Such a phenoform clustering approach would provide a more quantitative framework for defining intergrades and uncertainty in mapping. Used in conjunction with spatial inherent-property-based databases, the phenoform clustering approach could also be beneficial to assess soil natural capital and to predict susceptibility of specific soils to land-use intensification.

1. Introduction

Soil survey, soil classification and land evaluation programmes have largely focused on inherent soil properties that change very slowly over time and are largely insensitive to land use. Conventional soil maps are made with reference to these relatively stable soil properties. This focus has traditionally been the domain of pedology. Soils are continually evolving and transforming within anthropogenic timescales (Richter et al., 2011), and the study of soil dynamic properties that are sensitive to land use has generally been the domain of agronomy, soil fertility, soil biology, soil quality, and soil ecology.

The state factor approach of Jenny (1941) firmly placed soil formation in an ecosystem context. A new emphasis in soil science is emerging with the recognition of soil as a crucial component of the earth's natural capital, that highlights not only the ecological integrity of soils but also the economic and social services that underpin the earth's ecosystems and economies (Robinson et al., 2013; Dominati et al., 2010). This growing area of research will force the integration of traditional sub-disciplines of soil science because the quantification and valuation of soil natural capital (Hewitt et al., 2012) and of soil services require the integration of data on both inherent and dynamic soil properties.

Droogers and Bouma (1997) have provided a useful conceptual framework to integrate inherent and dynamic soil properties that may bridge this gap. 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. McBratney et al. (2014), discuss the genoform/phenoform concept in the context of soil capability and condition and suggest that the genoform represents a reference state that encompasses the inherent capability of the soil and condition under a specific long-term circumstance (e.g. natural vegetation). The phenoform reflects the condition due to specific management, but they also note that in a soil that passes a critical threshold, a phenoform may also become a new reference state.

Soil classifications and spatial soil databases generally exclude dynamic soil properties and are limited in their ability to support realistic spatial analyses of land use issues involving dynamic soil properties. Although there has been some movement toward linking knowledge of dynamic soil characteristics into soil survey, soil classification and land evaluation (see for instance Pennock and Veldkamp, 2006), progress

* Corresponding author.
E-mail address: stevensonb@landcareresearch.co.nz (B.A. Stevenson).

http://dx.doi.org/10.1016/j.geoderma.2014.10.003
0016-7061/

has been slow. In contrast, soil quality is a subdiscipline that has focused almost exclusively on the variation of dynamic soil properties where indicators are deliberately chosen to represent key dynamic responses of soil natural capital to the impacts of human land use management. The soil quality literature provides many examples of relationships of individual soil quality indicators with land use and soil type (for example, Brejda et al., 2000; Sparling and Schipper, 2004; Cotching and Kidd, 2010). Sparling and Schipper (2002) examined an initial New Zealand soil quality data set using principal components.

Here, we examine in more detail the multivariate clustering of an expanded New Zealand soil quality data set that contains only dynamic soil properties. Cluster centroid classification has been proposed for soil classification systems of inherent soil properties (see McBratney and De Gruijter (1992) and Minasny et al. (2010)) and would also allow development of classifications that incorporate both dynamic and inherent soil properties — a feature particularly useful in the characterisation of soil natural capital. The motivation for our research was to utilise the genoform–phenoform concept to indicate where land-use change may significantly affect soil quality indicators for specific soils or groups of soils.

This paper considers three hypotheses:

1. Managed sites (sites that are managed for a particular purpose and/or have had significant anthropogenic alteration) are statistically distinct from sites under native vegetation
2. Managed sites are clustered into statistically distinct classes of related soil quality states
3. Clusters have functional meaning as assessed by relationships to land use and soil type and their impress on soil quality states.

2. Material and methods

2.1. Soil quality database description

The New Zealand soil quality data set for regional-scale monitoring currently holds data for in excess of 700 sites, over 12 geographical regions within the country. The core set of soil quality indicators (pH, total C, total N, anaerobically mineralizable N, Olsen P, bulk density, and macroporosity) and their symbols, are listed in Table 1. There are very few soils containing carbonates in New Zealand (and of those that do, carbonates do not occur in the A horizon), so that total C is equivalent to organic C.

Schipper and Sparling (2000) and Sparling and Schipper (2002) discussed the rationale for choice of these indicators, sampling strategy and analytical methods. Sampling sites were stratified by soil order (Hewitt, 2010) and by land use. The land use categories were: cropping, horticulture, dairy pasture, dry stock pasture (sheep and beef cattle), exotic forestry (predominantly *Pinus radiata*), native forest, and native tussock grassland. Sparling et al. (2004) found that land use and soil order together explained 50–68% of total variance for the different indicators (12–49% for land use alone and 21–39% for soil type alone). Cotching and Kidd (2010) also reported that land use and soil order explained a high proportion of the total variance, for 6 indicators (4 of which are common to the New Zealand set of indicators) from a Tasmanian data set of 271 sites.

Table 1
Soil indicators and their transformations used for analysis.

Indicator	Symbol	Transformation
pH	pH	pH
Total carbon	TC	TC
Total nitrogen	TN	TN
C:N ratio	CN	Log(CN)
Olsen P	OLSEN	Log(OLSEN)
Anaerobically mineralizable N	AMN	Sqrt(AMN)
Bulk density	BD	BD
Macroporosity	MP	Sqrt(MP)

Principal component analysis by Sparling and Schipper (2002) related the variance of the seven indicators to acidity (pH), organic resources (total C + N, anaerobically mineralizable N), physical resources (bulk density and macroporosity) and fertility (Olsen P). Consequently the SINDI (Soil INDIcators) system for interpretation of soil quality indicator values was designed. The SINDI system used a panel of soil scientists to develop target ranges for the indicators by soil type and land use to establish soil quality ratings for the different indicators (Sparling et al., 2003). The development and operation of the monitoring system were described by Lilburne et al. (2002), Sparling et al. (2004), Sparling and Schipper (2004), and Lilburne et al. (2004). Giltrap and Hewitt (2004) described the short range variability of soil quality indicators which declined in a sequence: native forest > exotic forest > pasture > cropping.

2.2. Soil sampling

Sampling strategy and protocols, and lab analysis methods were described by Sparling and Schipper (2002). In brief, at each site approximately 25 soil cores (25 mm diameter to a depth of 10 cm) were collected along a 50-m transect and bulked for chemical analysis. Three soil cores (100 mm diameter and 75 mm depth) were also collected equidistant along the transect for soil physical analyses. The sites were sampled between 1978 and 2009. If a site was revisited for repeat sampling over time, only data from the most recent sampling date were used.

Land-use history is not available at most of the sites. An assumption in this study is that sites have been predominantly under the same land use for sufficient time that the observed soil quality reflects the effects of impacts from the current land use at time of sampling.

2.3. Data analysis

The soil quality data set was analysed in R (R Development Core Team, 2013) unless otherwise noted. The soil C:N ratio was included along with the seven indicators previously listed above as it is easily derived from the existing data, and generally much less correlated to total C than is total N. Previous records for a particular site were excluded, as noted before, if they had been resampled at a later date, and records were only retained for analysis if all seven indicators were available. This resulted in a total of 720 records in the final data set. Initially the data was explored without considering relationships to land use and soil factors. Some of the distributions of the different indicators were strongly skewed to the right, in part because the values are all non-negative. The set of soil quality indicators analysed, abbreviations for the indicators, and any transformations before analysis are shown in Table 1.

We anticipated a degree of correlation between some of the soil quality indicators (such as C and N), so principal components (PC) analysis was used to produce independent components, and provide some opportunity to reduce the dimensionality of the data set and thus to re-examine the initial findings of Sparling and Schipper (2002). In order to investigate whether the soil records exhibited natural grouping, we used fuzzy *c*-means clustering (Bezdek, 1981), which is very similar to the *k*-means clustering algorithm (Hastie et al., 2009), but has better convergence properties (to clarify further, the algorithms for fuzzy c-means clustering and fuzzy k-means clustering are identical). 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. No differential weighting of indicators was used so that each individual indicator was considered equal to all others.

The fuzzy *c*-means method requires an initial estimate of the number of clusters, begins with a random assignment of points to clusters,