Geoderma 309 (2018) 60–74

Contents lists available at ScienceDirect

Geoderma

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

Spatial variability of Australian soil texture: A multiscale analysis

Stacey Paterson*, Budiman Minasny, Alex McBratney

Sydney University, 1 central avenue, Eveleigh, NSW 2015, Australia

A R T I C L E I N F O

Handling Editor: Morgan Cristine L.S.

Keywords:
Variogram
Soil texture
Scale
Fractal

A B S T R A C T

Understanding how soil variability changes with spatial scale is critical to our ability to understand and model soil processes at scales relevant to decision makers. The compilation of large legacy data sets has opened up new possibilities to model spatial variability at the continental or even global scale. Using the National Soil Site Collation (NSSC) dataset of Australia we created empirical variograms for sand and clay fraction at extents from 1 km to continental. The NSSC dataset is highly spatially clustered; a typical feature of legacy datasets. This leads to lumpy artefacts in the variograms. To reduce this lumpiness we employed grid based declustering. We used the declustered empirical variograms to calculate the Hausdorff Besicovitch Dimension – a unitless measure of spatial roughness. We first fit a power model to each declustered variogram and calculated the Hausdorff Besicovitch dimension at each modelled scale. This allowed us to assess the roughness or variability at each modelled extent, however this assessment was somewhat arbitrary and showed that roughness depends on the extent. We have proposed a new model that allows us to calculate the Hausdorff Besicovitch dimension continuously across all extents. The conceptual basis of this model moves away from a multi-fractal framework typically used by soil scientists. It allows us to describe spatial variability or stochasticity as a continuous function of spatial separation. Both our new model and the continental scale variograms of texture emphasise the high degree of short range variability in soil texture. Empirical variograms indicate that around 50% of spatial variability occurs at < 10 km, and 30% at < 1 km. Spatial variability of soil texture increases with depth consistently across all modelled extents. Beyond extents of around 100 km, the Hausdorff Besicovitch Dimension remains relatively stable. Soil spatial variability is highly stochastic at fine scales however it changes gradually with extent and scale rather than abruptly.

1. Introduction

Our ability to understand and manage the soil resource is dependent on the scale at which we can observe and model soil characteristics and processes. As soil scientists, one of our key challenges is to produce information about soil quality and processes at a resolution and extent useful for decision makers (Lark, 2005; Malone et al., 2013). It may often be necessary to do this without the collection of additional data (Malone et al., 2013; Pongpattananurak et al., 2012). Modelling soil properties is challenging because a soil property at any given location is the result of a complex interplay of environmental and management factors over time. While these are in theory deterministic processes, the outcomes of these complex soil forming processes are often so unpredictable that they appear random (Heuvelink and Webster, 2001; Webster, 2000). The relative dominance and interactions of these different factors will vary with location and with the scale of observation (Heuvelink and Webster, 2001; Lark, 2011). This applies to both the deterministic and the 'random' component of soil variability. Capturing variability at relevant spatial scales is critical to production of useful models and maps, but is not a simple task. Without a priori knowledge of patterns in soil spatial variability it is easy to design a soil survey that misses important spatial variation either by sampling with spacing that is too broad or an extent that is too narrow. The importance of this issue has led to much work on the efficient design of soil surveys across multiple scales (Lark, 2005, 2011; Pettitt and McBratney, 1993; Webster et al., 2006). Even with these efficient methods, multiscale sampling strategies tend to be both time consuming and expensive and not always possible. It may be possible for the soil scientist to shape their expectations about the likely variability of soil at different scales from existing literature, but generally speaking our understanding of the variability of the majority of soil properties at different spatial scales is still limited.

The availability of continental-scale soil data allows new avenues for approaching the question of how soil variability changes with scale. In this paper:

* Corresponding author.
E-mail addresses: sepaterson87@gmail.com, stacey.paterson@sydney.edu.au (S. Paterson).

http://dx.doi.org/10.1016/j.geoderma.2017.09.005
Received 30 April 2017; Received in revised form 22 August 2017; Accepted 3 September 2017

- We use legacy data to calculate empirical variograms at spatial extents ranging from continental to 1 km by varying bin size and extent. This illustrates the utility of the empirical variogram as a tool for exploiting large legacy datasets in investigation of scaling patterns (Sections 2.2, 2.3, 2.4 and 3.1).
- We describe two methods for calculating the spatial roughness for the quantitative assessment of spatial variability (Sections 2.5, 2.6, 2.7 and 2.8) and apply these methods to the empirical variograms we have calculated (Sections 3.2 and 3.3). This allows us to expand on the inferences we have drawn about changes in variability across scales (Sections 3.4, 3.5, 3.6, and 3.7).

2. Methods

2.1. Conceptual overview

Collaborative efforts to build large scale digital soil maps such as *GlobalSoilMap* have led to the creation of consolidated databases of soil information. These databases represent a significant resource for empirical characterization of soil spatial variability. Our idea was to take advantage of the inherent flexibility in the experimental variogram to create soil variograms at a range of spatial scales using compiled legacy data. By adjusting the bin size and extent of each variogram we adjusted the spatial scale so that each variogram captures a different magnitude of spatial variability. Creation of variograms across a range of spatial scales allowed the characterization of patterns of spatial variability with scale. We fit power curves to the empirical variograms across a range of modelled scales. The exponent parameter from the fitted power curve was used to calculate the Hausdorff Besicovitch Dimension or D value, a unitless measure for the roughness of an object. Burrough (1983) used this dimension to compare spatial variability between environmental properties. The 'variogram method' used by Burrough, 1983 is rooted in the concept of Multifractals (regions of similar variability separated by 'zones of transition'). We introduce a differentiation-based method for estimating this dimension continuously across changing extents. Because the underlying conceptual framework for our model is distinct from the Multifractal framework we replace the term Hausdorff Besicovitch Dimension with the more general 'roughness index'. Because the roughness index is dependent upon the shape of the variogram but not the units, it provides a simple but useful quantitative tool for assessing spatial variability between properties and between scales. We calculated the 'roughness index' across different spatial extents and at several different depths using both methods.

2.2. NSSC soil texture data

The soil texture data used in this analysis was compiled to support the Australian contribution to the *GlobalSoilMap* (Grundy et al., 2015). A collaboration of state and national government agencies and some universities worked together to produce the National Soil Site Collation or NSSC (Searle, 2014). The database includes geo-located soil observations collected by research and government agencies from the 1930s onwards. The NSSC is a composite of data from a variety of sources therefore it does not have a unified sampling design and the NSSC dataset reflects the research priorities of the different data collecting institutions at different times. The dataset is heavily focused in agricultural regions and includes areas of high density sampling and sparse sampling (Fig. 1). The complete database contains information on several soil properties including percentage clay and percentage sand fraction from almost 16,000 soil profiles. Percentage sand and clay fractions from this database are used in this study.

Observations in the NSSC database were not taken at consistent depths. Data was normalized using the generalized equal area spline depth function (Malone et al., 2009). Soil depth intervals were selected in line with the *GlobalSoilMap* depth intervals (0–5 cm, 5–15 cm, 15–30 cm, 30–60 cm, 60–200 cm, Arrouays et al., 2014). Prior to applying the spline depth function, locations with no top soil measurement were discarded and locations with multiple observations or overlapping depths were deleted. The number of observations that were used at each interval are shown in Tables 1 and 2 below. The spline function does not return values at depths below the available data. The NSSC database contains more observations for the topsoil than the subsoil. This results in fewer data points observations available at lower depth intervals.

Summary statistics for percentage Clay and Sand fraction are presented in Tables 1 and 2 below.

2.3. Experimental variograms — modelling at multiple scales

We calculated experimental variograms using Matheron's (1963) method-of-moments estimator (Eq. (1)).

$$\hat{\gamma}(h) = \frac{1}{2M(h)} \sum_{j=1}^{M(h)} \{z(x_j) - z(x_j + h)\}^2 \tag{1}$$

In Eq. (1) (above) the theoretical relationship between separation distance (lag or h) and semivariance, γ, is estimated by the function $\hat{\gamma}(h)$. $M(h)$ is the number of paired comparisons at a particular lag (h). $z(x_j)$ and $z(x_j + h)$ are the values of the property Z at places x_j and $x_j + h$ separated by lag h. It is common practice for the lag h to cover a specified distance interval.

Use of method of moments to estimate semivariograms has been criticized for bias and for subjectivity (Lark, 2000). However, bias decreases as sample size increases (Oliver and Webster, 2014). Variograms are typically estimated based on tens to hundreds of data points, while this study uses several thousand. This significantly reduces problems of bias. Another reason to favour the use of Method of Moments in this context is the difficulty associated with using either REML (Restricted Maximum Likelihood) or MCMC (Markov Chain Monte Carlo) methods on very large datasets. In addition model-based geostatistics require a priori establishment of a variogram model.

When using method of moments, the practitioner is required to select both bin size and extent. In relation to Eq. (1), the bin sizes determine the interval over which the term h spans. The intrinsic subjectivity of this method provided a convenient method for modelling variograms at different scales. Fixing the number of bins at 1000, the maximum extent of the experimental variogram was gradually reduced. As the extent decreased, the bin size decreased proportionally. Combinations of bin size and extent are displayed in Table 3.

2.4. Experimental variograms — improving fit using spatial declustering

It has been established that empirical variograms calculated from spatially clustered data can be biased or lumpy (Emery, 2007; Marchant et al., 2013; Richmond, 2002). This makes them less suitable for modelling variograms and for kriging because in clustered situations the variability at different lags is unequally characterised. As discussed above, the NSSC dataset used in this paper has been compiled from a variety of government agencies and research bodies, and reflects the priorities of those bodies at the time of data collection. As such, the dataset is heavily clustered. Empirical variograms calculated with the method-of-moments exhibit strong lumpiness at spatial extents > 40 km (you can see this lumpiness illustrated in Fig. 7). This is consistent with the pattern noted by Marchant et al. (2013) when using a similar dataset.

Methods for reducing this bias have been suggested by Emery (2007), Richmond (2002) and Marchant et al. (2013). We favour the last method as, unlike the first two, it is dependent only upon the spatial location of the data and not upon the values of the data themselves. It is also easily computed and has intuitive appeal.

We use Marchant et al. (2013) modified declustering method of moments estimator (Eq. (2)) to recalculate the empirical variograms at