Geoderma 262 (2016) 243–253

Contents lists available at ScienceDirect

Geoderma

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

Comparing regression-based digital soil mapping and multiple-point geostatistics for the spatial extrapolation of soil data

Brendan P. Malone [a,]*, Sanjeev K. Jha [b], Budiman Minasny [a], Alex B. McBratney [a]

[a] Department of Environmental Sciences, Faculty of Agriculture and Environment, C81 Biomedical Building, The University of Sydney, New South Wales 2006, Australia
[b] School of Civil and Environmental Engineering, University of New South Wales, Sydney, New South Wales 2052, Australia

ARTICLE INFO

Article history:
Received 12 April 2015
Received in revised form 24 August 2015
Accepted 25 August 2015
Available online 8 September 2015

Keywords:
Digital soil mapping
Gamma radiometrics
Multiple-point geostatistics
Regional soil mapping

ABSTRACT

In this study, two approaches for spatial data extrapolation are investigated. The intention here is to predict at fine spatial resolution, total gamma radiometric counts across a large mapping extent (recipient site) on the basis of finely resolved information collected from a nearby donor site. The extrapolation methods used were a digital soil mapping (DSM) regression model approach and a multivariate multiple-point statistical (MPS) approach. Qualitative interpretation of the results from both extrapolation approaches across the recipient site in the Lower Hunter Valley, Australia (area ≈ 220 km^2) shows promise in terms of highlighting known geochemical and physical variations of soils in this area. The extrapolated map was evaluated in a small portion of the study area (area ≈ 4 km^2) where similar high-resolution gamma radiometric data were available. Results show comparable performance of both approaches where a root-mean-square error of 87 ppm was found. A concordance correlation coefficient value of 0.04 was found for the DSM approach, but higher for the MPS approach (0.16). Under the Homosoil framework, where soil point data and mapping are sparse, either method investigated in this study would be suitable as a 'first-cut' approach for developing a comprehensive soil information system in those areas.

1. Introduction

One of the issues in developing high-resolution global and national spatial soil information systems of consistent coverage is reconciling some of the disparity between those areas that have well developed soil information resources with those that are comparatively underdeveloped (Minasny and McBratney, 2010). To address this disparity, most soil scientists would advocate a rebirth of soil survey and mapping programmes to rival the efforts made internationally during the early to mid-20th century (Brevik and Hartemink, 2010) in the areas where information is currently sparse. While appealing, we need to permit ourselves to consider alternative and possibly less costly approaches; with one being model extrapolation, to which is the focus of this investigation.

The concept of Homosoil (Mallavan et al., 2010) has particular relevance in that regard, because it aims, through similarity assessment, the evaluation of which soils (unknown) are similar to other soils (known). For example, if one specified area has very detailed soil mapping (donor site), and has similar soil forming factors to another area that has little to no soil mapping, then it may be possible to extrapolate the information or model from the detailed area to the sparse area (recipient site). These ideas have been around for a while; for example, Lagacherie et al. (1995) implemented an extrapolation concept in France where soil pattern rules were acquired from a reference area or donor site and applied across a wider area where a lower intensity of survey had been achieved. The extrapolation of data is a general concept, and one that can be applied for other variables that are not exclusively soil attributes or classes. For example, proximal soil sensing instruments are able to collect very detailed information about the geochemical and geophysical properties of soils (with gamma radiometrics and electromagnetic induction as a few common examples).

Such proximally sensed information has been demonstrated to be invaluable for soil studies in terms of digital mapping and precision agriculture (Viscarra Rossel et al., 2010). However, their application is commonly restricted to farm and field spatial extents. Using them at regional and larger extents is rare because it is difficult and costly to maintain the same sampling frequency at these scales as for field and farm extents. This issue of practicality has prompted a few recent studies to use proximal soil sensing instruments for regional scale studies. For example, both Viscarra Rossel et al. (2014) and Stockmann et al. (2015) developed efficient methods of traversing a landscape that dually attempt to minimise the time spent in the field yet maximise the potential to capture the spatial soil variation at their scale of investigation. In a similar context, Podgorski et al. (2015) demonstrated the value of integrating proximal sensed geophysical data – that was collected at limited

---

* Corresponding author.
E-mail addresses: brendan.malone@sydney.edu.au (B.P. Malone), s.jha@unsw.edu.au (S.K. Jha), budiman.minasny@sydney.edu.au (B. Minasny), alex.mcbratney@sydney.edu.au (A.B. McBratney).

http://dx.doi.org/10.1016/j.geoderma.2015.08.037
0016-7061/

sites – with airborne sensed data for constraining and delivering a more detailed hydrological and geological model across a large spatial extent of Botswana (Okavango Delta).

In this study we approach the problem of delivering detailed mapping differently by investigating the efficacy of model extrapolation through the use and subsequent comparison of two contrasting (extrapolation) approaches. The first is using a digital soil mapping approach (McBratney et al., 2003) as suggested in Mallavan et al. (2010). The second is via multiple-point statistics, in particular the Direct Sampling algorithm as described in Mariethoz et al. (2010).

The first extrapolation method hereafter referred to as the DSM approach, entails the following steps. From the area with detailed information, first the target variable of interest is decided upon. Using existing point observations (for which there should be many), or sampling directly from an available raster of the property of interest, these data are then intersected with a portfolio of spatially exhaustive environmental covariate data. This information could be retrieved from an available digital elevation model, remote sensing data platform or some other similar source (Mulder et al., 2011). A DSM model is then constructed, which is essentially a numerical model that relates the information on the variable of interest to the environmental factors. The constructed model is then applied to the recipient site. Grinand et al. (2008) used a DSM approach in France for mapping soil types to investigate the extent to which a model yields a valid prediction. The accuracy of predictions made for the extrapolated area (recipient site) was found to be lower than that made in the training or donor area. Intuitively, this type of result is expected because of the complexity of spatial soil variation, and the impossibility of matching soil forming factors between donor and recipient sites. The results from Grinand et al. (2008) are encouraging from the perspective that such an extrapolation approach would be useful to fill the gaps in present soil map coverage and to increase efficiency of ongoing soil survey to target areas of greatest uncertainty.

Multiple-point statistics (MPS) (Guardiano and Srivastava, 1993) has not before been used in the context of Homosoil. In fact, there have only been a limited number of soil science studies that have explored MPS, with Meerschman et al. (2013a) and Meerschman et al. (2014) being a few examples. Originally developed in the field of geological reservoir modelling, MPS represents an alternative to two-point statistics such as that of variogram modelling and subsequent kriging, and even DSM modelling, with recent applications in hydrogeology (Chugunova and Hu, 2008; Jha et al., 2014), geophysics (Liu et al., 2004; Comunian et al., 2014), and remote sensing (Ge and Bai, 2010; Mariethoz et al., 2012). A stated advantage of MPS is its ability to capture complex patterns and connectivity in data, which is difficult to do with two-point statistics (Mariethoz et al., 2010). In statistical literature, Markov Random Fields serve as the statistical construct that underpins MPS, e.g. Besag (1986) and Emery and Lantuéjoul (2014). Central to MPS, is the training image, which is a conceptual image of the expected spatial structure of the variable to be predicted. The idea of training images is that there may exist another site – a soil analogue in this case (i.e. the training image) – where large amounts of information are available, and from which it is possible to learn spatial or textural information. This idea is very much in line with the concept of Homosoil, making MPS an interesting candidate technique in this context. Spatial patterns learnt from a training image were particularly relevant for Meerschman et al. (2014) in processing proximal soil sensor data given a repeating polygonal fossil ice-wedge soil pattern. Extending MPS to include multivariate training images (Jha et al., 2013a, 2013b, 2015) provides an opportunity to explore its broader application for digital soil mapping efforts, and consequently for Homosoil. The hypothesis here is that environmental covariates together with detailed (soil) mapping from the donor site can be used as training image to inform the spatial pattern of mapping at the recipient site.

The subsequent investigation is a scoping study and details the use of the above-described methods of extrapolation for mapping the total count gamma-ray emission from soils across the Lower Hunter Valley, NSW (recipient site), given some existing detailed survey from the same area (albeit at a much smaller spatial extent). We firstly describe the study area and data used in this study. Secondly the theoretical underpinnings of DSM and MPS are described, followed by description of the procedures for implementing each of the approaches. Lastly, subsequent results and outputs are presented together with a broader discussion of their significance.

2. Materials and methods

2.1. Study area

The study area is located in the Lower Hunter Valley, NSW, Australia (32.83°S 151.35°E), approximately 140 km north of Sydney, NSW, Australia, and covers an area of approximately 220 km^2 (Fig. 1). This area is referred to as the Hunter Wine Country Private Irrigation District (HWCPID). This area is situated in a temperate climatic zone, and experiences warm humid summers, and relatively cooler yet also humid winters. Rainfall is mostly uniformly distributed throughout the year. The area receives on average just over 750 mm of rainfall annually (Australian Government Bureau of Meteorology, 2014). Topographically, this area consists mostly of undulating hills that ascend to low mountains to the south-west. The underlying geology includes predominantly Early Permian siltstones, marl, and some minor sandstone (Hawley et al., 1995). Other parent materials include Late Permian siltstones, and Middle Permian conglomerates, sandstones and siltstones. Soils are quite variable, but in general terms are weathered mixed kaolinitic–smectitic type soils.

2.2. The data

The recipient site for this study is the entire HWCPID. In 2013 an area of 15 km^2 was surveyed using a ground-based gamma-ray detector (Stockmann et al., 2015) to produce raster maps of the radiometric ROIs (regions of interest) with a raster cell size of 25 by 25 m (shown in yellow in Fig. 1). Specifically, that work entailed driving across the landscape following a network of pre-determined transects. A gamma-ray spectrometer was attached to the vehicle which recorded on-the-go radiometric signals being emitted from the soil surface. On average, the ‘sampling’ density of the on-the-go proximal sensing was 45 points per hectare. For the work of Stockmann et al. (2015), the data was collected for total gamma-ray count and the ROIs that corresponded to Potassium, Thorium, and Uranium. All data were mapped in the units of counts-per-second (cps). The mapped outputs from Stockmann et al. (2015) represent the donor site in this study — they are detailed data that need to be extrapolated to the entire HWCPID. It is possible that this extrapolated information could be used in the future for updating existing soil mapping, and more generally for digital soil mapping studies in this region such as the refinement of soil and landscape regions or terrons as described in Malone et al. (2014a). This study focuses specifically on the mapping of the total gamma-ray counts rather than each of the individual ROIs.

Both extrapolation methods (DSM and MPS) make use of spatially exhaustive covariate information derived principally from a digital elevation model (25 m × 25 m spatial resolution). In total 7 environmental covariates were used in this study: elevation, altitude above channel network, incoming solar insolation, mid-slope position, multi-resolution valley bottom flatness, terrain wetness index, and slope. The processing of the digital elevation model (DEM) to derive these additional terrain-based variables was performed using SAGA-GIS (System for Automated Geoscientific Analyses, http://www.saga-gis.org). Maps of each of the covariates are shown in the supplementary material associated with this manuscript.