ARTICLE IN PRESS

Spatial Statistics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Spatial Statistics

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

Spatio-temporal interpolation of soil water, temperature, and electrical conductivity in 3D + T: The Cook Agronomy Farm data set

Caley K. Gasch[a,*], Tomislav Hengl[b], Benedikt Gräler[c], Hanna Meyer[d], Troy S. Magney[e], David J. Brown[a]

[a] *Department of Crop and Soil Sciences, Washington State University, USA*
[b] *ISRIC — World Soil Information/Wageningen University and Research, The Netherlands*
[c] *Institute of Geoinformatics, University of Münster, Germany*
[d] *Department of Geography/Environmental Informatics, Philipps-Universität Marburg, Germany*
[e] *College of Natural Resources, University of Idaho, USA*

ARTICLE INFO

Article history:
Received 1 November 2014
Accepted 1 April 2015
Available online xxxx

Keywords:
Digital soil mapping
Random forests algorithm
Regression-kriging
Soil sensor network

ABSTRACT

The paper describes a framework for modeling dynamic soil properties in 3-dimensions and time (3D + T) using soil data collected with automated sensor networks as a case study. Two approaches to geostatistical modeling and spatio-temporal predictions are described: (1) 3D + T predictive modeling using random forests algorithms, and (2) 3D + T kriging model after detrending the observations for depth-dependent seasonal effects. All the analyses used data from the Cook Agronomy Farm (37 ha), which includes hourly measurements of soil volumetric water content, temperature, and bulk electrical conductivity at 42 stations and five depths (0.3, 0.6, 0.9, 1.2, and 1.5 m), collected over five years. This data set also includes 2- and 3-dimensional, temporal, and spatio-temporal covariates covering the same area. The results of (strict) leave-one-station-out cross-validation indicate that both models accurately predicted soil temperature, while predictive power was lower for water content, and lowest for electrical conductivity. The kriging model explained 37%, 96%, and 18% of the variability in water content, temperature, and electrical conductivity respectively versus 34%, 93%, and 5% explained by the random forests model. A less rigorous simple cross-validation of the random forests model

* Corresponding author.
E-mail address: caley.gasch@wsu.edu (C.K. Gasch).

http://dx.doi.org/10.1016/j.spasta.2015.04.001

Please cite this article in press as: Gasch, C.K., et al., Spatio-temporal interpolation of soil water, temperature, and electrical conductivity in 3D + T: The Cook Agronomy Farm data set. Spatial Statistics (2015), http://dx.doi.org/10.1016/j.spasta.2015.04.001

indicated improved predictive power when at least some data were available for each station, explaining 86%, 97%, and 88% of the variability in water content, temperature, and electrical conductivity respectively. The high difference between the strict and simple cross-validation indicates high temporal auto-correlation of values at measurement stations. Temporal model components (i.e. day of the year and seasonal trends) explained most of the variability in observations in both models for all three variables. The seamless predictions of 3D + T data produced from this analysis can assist in understanding soil processes and how they change through a season, under different land management scenarios, and how they relate to other environmental processes.

1. Introduction

Comprehension of dynamic soil properties at the field scale requires measurements with high spatial and temporal resolution. Distributed sensor networks provide frequent *in situ* measurements of environmental properties at fixed locations, providing data in 2- or 3-dimensions and through time (Porter et al., 2005; Pierce and Elliott, 2008). While sensor networks produce ample data for observing dynamic soil properties, data processing for inference and visualization become increasingly difficult as data dimensionality increases. Ideally, the end product should consist of seamless interpolations that accurately represent the spatial and temporal variability in the property of interest. These products can then be used for predictions at unobserved locations, they can be integrated into process models, and they can simply aid in visualization of soil properties through space and time.

Multiple approaches have been developed for spatial interpolation of soil properties and digital soil mapping, including:

(1) multiple regression models based on the soil forming factors, terrain attributes, spatial coordinates, or derived principal components (McKenzie and Ryan, 1999);
(2) smoothing (splines) and neighborhood-based functions (Mitas and Mitasova, 1999);
(3) geostatistics, or kriging, and variations thereof (see overviews by McBratney et al. (2003) and Hengl (2009)).

Of these, regression-kriging (Odeh et al., 1995; Hengl et al., 2007), which combines a multiple regression model (a trend) with a spatial correlation model (a variogram) for the residuals, produces unbiased, continuous prediction surfaces. Regression-kriging has been adapted for soil mapping with great success, in part because of the flexibility in defining the trend model as a linear, non-linear, or tree-based relationship between the response and predictors. Furthermore, regression-kriging relies on the incorporation of auxiliary data, providing mechanistic support for the soil property predictions.

The widest application of regression-kriging in soil science has likely been for producing 2-dimensional (2D) maps (Hengl, 2009). However, soil data is often also collected at multiple depths, and geostatistical interpolation techniques can be expanded to represent soil predictions across both vertical and horizontal space (Malone et al., 2009; Veronesi et al., 2012). Global predictions of multiple soil properties obtained from 3-dimensional (3D) regression models were recently showcased by Hengl et al. (2014a). Here, spline functions define the vertical trend (depth) within the regression model, while horizontal trends are defined by covariate grids. These approaches are sufficient for understanding static soil properties across 2- and 3D space; however, modeling dynamic soil properties requires expansion of the geostatistical model to incorporate correlation in data through time (Heuvelink and Webster, 2001; Kyriakidis and Journel, 1999). Addition of temporal and/or spatio-temporal predictors can assist in explaining temporal variation in a response variable, but fitting a variogram model in 2D and time (2D + T) poses additional challenges (summarized by Heuvelink and Webster, 2001). Specifically, time exists in only one dimension and has a directional component, while

Please cite this article in press as: Gasch, C.K., et al., Spatio-temporal interpolation of soil water, temperature, and electrical conductivity in 3D + T: The Cook Agronomy Farm data set. Spatial Statistics (2015), http://dx.doi.org/10.1016/j.spasta.2015.04.001