Geoderma Regional 7 (2016) 279–292

Contents lists available at ScienceDirect

Geoderma Regional

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

Prediction of soil texture using descriptive statistics and area-to-point kriging in Region Centre (France)

Mercedes Román Dobarco [a,1], Thomas G. Orton [b,1], Dominique Arrouays [a], Blandine Lemercier [c,d,e], Jean-Baptiste Paroissien [a], Christian Walter [c,d,e], Nicolas P.A. Saby [a,*]

[a] INRA, Centre de Recherche d'Orléans, US1106 Unité InfoSol, F-45075 Orléans, France
[b] Faculty of Agriculture and Environment, The University of Sydney, 1 Central Avenue, Australia Technology Park, Eveleigh, NSW 2015, Australia
[c] INRA, UMR1069, Sol Agro et hydrosystème Spatialisation, F-35042 Rennes, France
[d] Agrocampus Ouest, UMR1069, Sol Agro et hydrosystème Spatialisation, F-35042 Rennes, France
[e] Université Européenne de Bretagne, France

ARTICLE INFO

Article history:
Received 14 December 2015
Received in revised form 29 March 2016
Accepted 30 March 2016
Available online 6 April 2016

Keywords:
Area-to-point kriging
Area-to-point cokriging
Area-to-point regression cokriging
Topsoil texture
Disaggregation
French soil test database
REML

ABSTRACT

The French soil-test database (Base de Données d'Analyses de Terre: BDAT) is populated with analytical results of agricultural topsoil samples requested by farmers for fertilization planning. The coordinates of the farms are unknown due to data confidentiality policies, and the best available georeference is at level of municipality. We compared four approaches for mapping soil texture of agricultural land in Region Centre (France) using BDAT data: 1) a reference approach of mapping the mean of the aggregated data by municipality, 2) a boosted regression tree (BRT) model fitted with the municipality-averaged data, 3) area-to-point cokriging (AToP CK), and 4) a regression kriging version of this (AToP RCK, for which the BRT predictions were used to give the trend). Specifically, parameters for these last two approaches were fitted through the summary statistics approach to AToP kriging, which accounts for the full set of municipality summary statistics data (i.e. the mean, variance and number of measurements from each municipality). We could thus determine whether more complex and statistically-challenging approaches improve our knowledge on the spatial distribution of soil texture compared with maps of data aggregated by municipality. Texture data from 105 sites form the French soil monitoring network (Réseau de Mesures de la Qualité des Sols: RMQS) were used for independent validation. In general, the R^2 was greater for sand (average $R^2 = 0.69$) and silt (average $R^2 = 0.72$) than for clay (average $R^2 = 0.40$). The three methods for disaggregating the summary statistics data (BRT, AToP CK, and AToP RCK) showed similar prediction accuracies—although BRT predictions showed the greatest bias—and were better than the BDAT reference approach. AToP RCK was able to give similar prediction accuracy to BRT modelling alone, reduced the bias considerably, and gave a reasonable (although slightly conservative) assessment of prediction uncertainty. The results indicate that geostatistical methods for change of support expand the utility of aggregated data from soil-test databases.

1. Introduction

Soil texture is a key controlling factor of soil properties and functions like water and nutrient holding capacity, retention of pollutants, root development, soil biodiversity, and biogeochemical cycling (Silva et al., 2012). The effects of land-use change and agricultural practices on soil properties often depend on soil texture (Khan et al., 2012; van Capelle et al., 2012; Cotching et al., 2013). Therefore, detailed information on the spatial distribution of soil texture is essential for sustainable agricultural management, environmental protection, and hydrological planning (Adhikari et al., 2013; Akpa et al., 2014; Zhao et al., 2009). In addition, understanding and describing the relationships between environmental factors and soil texture fractions at fine scale is needed for site-specific planning (Greve et al., 2012). The availability of spatially continuous covariates, the incorporation of remote sensing technologies, the development of advanced quantitative techniques, and increased computing capacity have multiplied the possibilities for digital soil mapping, making feasible the production of high-resolution digital soil maps at national and even global scale (Grunwald, 2009; Sanchez et al., 2009; Arrouays et al., 2014).

The French soil-test database, or Base de Données Analyse des Terres (BDAT), is populated with analytical results of agricultural topsoil

* Corresponding author at: INRA, Centre de recherche d'Orléans, US1106 Unité InfoSol, 2163 avenue de la Pomme de Pin, CS 40001 Ardon, F45075 Orléans CEDEX 2, France.
E-mail addresses: mercedes.romandobarco@orleans.inra.fr (M. Román Dobarco), Thomas.orton@sydney.edu.au (T.G. Orton), Dominique.Arrouays@orleans.inra.fr (D. Arrouays), Blandine.Lemercier@agrocampus-ouest.fr (B. Lemercier), jeanbaptiste.paroissien@gmail.com (J.-B. Paroissien), christian.walter@agrocampus-ouest.fr (C. Walter), nicolas.saby@orleans.inra.fr (N.P.A. Saby).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.geodrs.2016.03.006

samples collected across France since 1990 (Lemercier et al., 2008). Farmers and landowners interested in the fertility status and physico-chemical properties of their soils request these analyses from commercial laboratories certified by the French Ministry of Agriculture, usually to optimize the application of fertilizer (Saby et al., 2008). The sampling depth is usually of 0–25 cm for crops and the 5–15 cm for pasture (Schvartz et al., 1997). Several soil cores sampled across the totality of the plot, or in a representative subplot of 10–20 m of diameter, are combined into a composite sample and sent for analysis (Schvartz et al., 1997). The locations of the farms are unknown to protect their anonymity, hence the best spatial approximation is at the level of municipality. The simplest way to analyze this type of data is by calculating summary statistics for administrative units and mapping the aggregated data. This approach ignores the variability of soil properties within the units due to changes in soil forming factors and inherent spatial variability. However, soil-test databases of aggregated data still provide valuable information for digital soil mapping and soil monitoring (Brus et al., 2014; Lemercier et al., 2008). The potential of the BDAT has been demonstrated for estimating and mapping phosphorus bioavailability (Follain et al., 2009), temporal changes in soil organic carbon concentration (Saby et al., 2008; Orton et al., 2012b), and potential for soil organic carbon sequestration of agricultural soils at national scale (Angers et al., 2011). The density of observations in the BDAT for the Region Centre is higher than in other French regions (Saby et al., 2014), which makes it appropriate for spatial analyses and for testing novel methodologies dealing with aggregated data prior to its application at national level.

The use of data summarized by areal units for spatial prediction effectively entails a change of support. Sophisticated geostatistical methods for change of spatial support developed during the last decade enable the prediction of point values from areal data and also provide an estimate of the prediction uncertainty (Kyriakidis, 2004). This particular case of disaggregation — where the data are of true areal averages, and the predictions are on a point support — is known as area-to-point kriging (AToP kriging). Orton et al. (2012a,b) developed a related area-to-point kriging (and cokriging) approach for use when data are in the form of summary statistics (mean, variance and number of observations) for areal units. Covariance and cross-covariance parameters were estimated by restricted maximum likelihood method (REML), and the method was applied to predict soil organic carbon concentration (and its temporal change) in the Franche-Comté region, France. One of the advantages of this method is that it takes into account the number of observations by unit and the unit size, estimating the sampling error of the areal mean while fitting a covariance model to the data (Brus et al., 2014).

The scorpan model provides a very general formulation for modelling soil properties: a soil property is a function of other soil properties (s), climate (c), organisms (o), relief (r), parent material (p), age (a), and spatial position (n) (McBratney et al., 2003). There are a multitude of regression techniques that have been used to quantify relationships for such a model. Techniques range from multiple linear regression (Meersmans et al., 2008) to data mining and tree-based methods such as rule-based algorithms (Adhikari et al., 2013), regression trees (Greve et al., 2012), boosted regression trees (Martin et al., 2014; Ciampalini et al., 2014), or random forests (Akpa et al., 2014). In particular, the regression tree methods can accommodate non-linear relationships between soil forming factors and soil properties, are robust against outliers, handle non-normal data distributions, and provide measures of the influence of each predictor variable.

Whatever technique is used to model the relationships there will be errors, and if these are spatially correlated then a kriging method can be applied to the residuals to improve predictions. This regression kriging (RK) approach has been demonstrated in a large number of studies when data are on point support (Odeh et al., 1994; Hengl et al., 2007; Sumfleth and Duttmann, 2008). For areal-support data, Kerry et al. (2012) used a function of relief and parent material to give the trend and applied AToP to the residuals (i.e. AToP RK) for mapping soil organic carbon concentration in Northern Ireland, obtaining more accurate results than with just AToP kriging. In the case of areal summary statistics data, Orton et al. (2012a,b) used the elevation as an auxiliary covariate for mapping soil organic carbon concentration.

The objective of this study was to produce and compare four approaches for mapping soil texture of agricultural land in the Region Centre using data from the BDAT: 1) a reference approach of mapping the mean of the aggregated data by municipality, 2) a boosted regression tree (BRT) model (Martin et al., 2014; Ciampalini et al., 2014; Friedman, 2001) fitted with the municipality-averaged data, 3) AToP cokriging, and 4) AToP regression cokriging (for which the BRT predictions are used to give the trend). Specifically, parameters for these last two approaches are fitted through the summary statistics approach to AToP kriging, which accounts for the full set of municipality summary statistics data (i.e. the mean, variance and number of measurements from each municipality). We could thus determine whether more complex and statistically-challenging approaches improve our knowledge on the spatial distribution of soil texture compared with maps of data aggregated by municipality.

2. Methods

2.1. Study region

The study area is the Region Centre (Fig. 1), which covers 34,151 km^2 and occupies the Middle Loire basin. The relatively flat

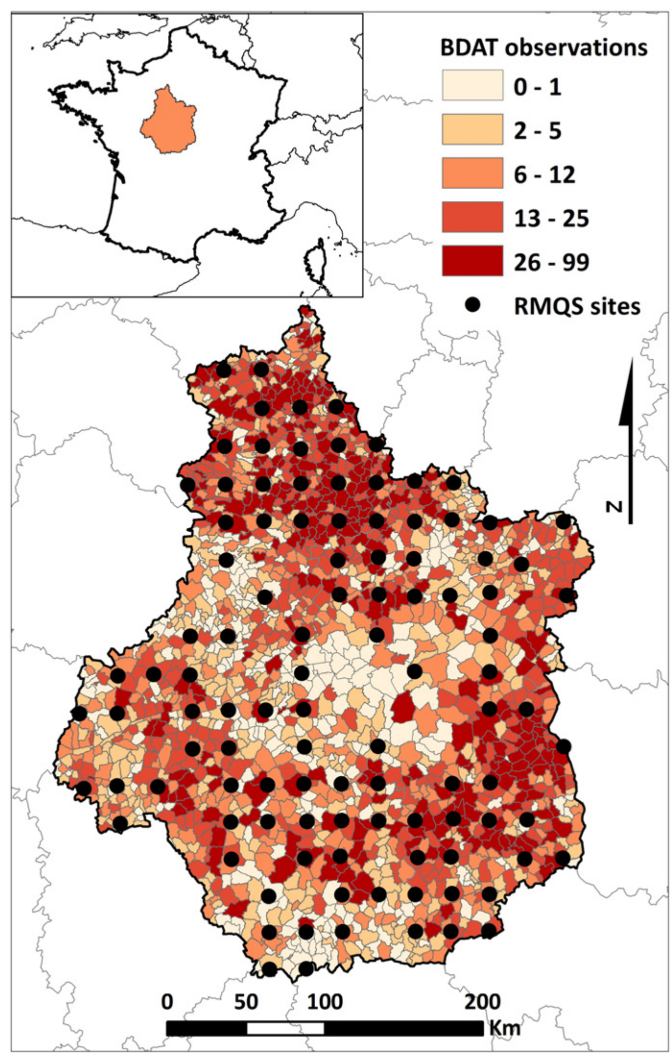

Fig. 1. General location of Region Centre (France), number of BDAT texture observations per municipality and location of RMQS sites used for independent validation.