Geoderma 324 (2018) 56–66

Contents lists available at ScienceDirect

Geoderma

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

Robust variogram estimation combined with isometric log-ratio transformation for improved accuracy of soil particle-size fraction mapping

Zong Wang[a,b,c], Wenjiao Shi[a,c,*]

[a] Key Laboratory of Land Surface Pattern and Simulation, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

[b] State Key Laboratory of Resources and Environment Information System, Institute of Geographical Science and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

[c] College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China

ARTICLE INFO

Editor: A.B. McBratney

Keywords:
Soil particle-size fractions
Spatial interpolation
Log-ratio transformation
Log-ratio cokriging
Robust estimator

ABSTRACT

Mapping soil particle-size fractions (psf) plays an important role in regional hydrological, ecological, geological, agricultural and environmental studies. To map soil compositional data like soil psf, interpolators such as compositional kriging and the combination of log-ratio transformations with ordinary kriging or cokriging were developed. In addition, robust estimators were proposed for these interpolators to improve the variogram models. However, few studies have focused on how to choose log-ratio transformation, kriging, cokriging, or robust variogram estimation methods based on data characteristics to achieve optimal performance when mapping soil psf by comprehensive comparative analysis. Here, we selected different compositional kriging, log-ratio kriging, log-ratio cokriging and log-ratio cokriging methods combined with a robust variogram estimator to improve the accuracy of spatial predictions of soil psf when using 262 soil samples from the upper reaches of the Heihe River in China. In this study, a comprehensive comparative analysis of soil psf maps generated by using different interpolators is presented, and appropriate methods for mapping psf based on the characteristics of the available data are explored. The results show that using isometric log-ratio (ILR) transformation with different interpolators can achieve relatively better performance than the other log-ratio transformation methods. In addition, combining the interpolators with robust variogram estimators significantly improve the prediction accuracy compared with using standard estimators, which presented reasonable and smooth transitions when mapping soil psf. Combining ILR cokriging with a robust variogram estimator had the best accuracy, with the lowest root mean squared error (sand, 10.50%; silt, 11.24%; clay, 7.32%), an Aitchison's distance of 0.76, a standardized residual sum of squares of 0.70 and a relatively higher rate of correctly predicting soil texture types 90.04%. In the future, guideline for using log-ratio transformation methods with linear regression, a generalized linear model or random forest should be developed and combined with ancillary variables to improve the interpolators.

1. Introduction

The distribution of soil particle-size fractions (psf), including sand, silt and clay, is a key parameter in many process models and pedotransfer functions, which are important for modeling most soil physical and chemical processes and different aspects of hydrology, such as water movement and solute transport, water, heat and nutrient fluxes, and soil erosion (Haverkamp and Parlange, 1986; Rawls et al., 1982). As compositional data, statistically analyzing and spatially predicting soil psf have the problem of spurious spatial correlation and can result in a closure effect (Odeh et al., 2003; Woronow and Love, 1990). Four requirements for the spatial prediction of soil psf need to be honoured: unbiasedness, non-negativity, constant sum, and a minimum sum of prediction error variances (Walvoort and de Gruijter, 2001).

Abbreviations: psf, soil particle-size fractions; ALR, additive log-ratio transformation; CLR, centered log-ratio transformation; ILR, isometric log-ratio transformation; SLR, symmetry log-ratio transformation; ALR_OK, CLR_OK, ILR_OK and SLR_OK, ALR, CLR, ILR and SLR, combined with ordinary kriging, respectively; ALR_COK, CLR_COK, ILR_COK and SLR_COK, ALR, CLR, ILR and SLR, combined with cokriging, respectively; ALR_RCOK, CLR_RCOK, ILR_RCOK and SLR_RCOK, ALR_COK, CLR_COK, ILR_COK and SLR_COK, combined with robust variogram estimator, respectively; log-ratio CK, log-ratio cokriging; log-ratio RCK, log-ratio CK combined with robust variogram estimator; MAD, median absolute deviation; AD, Aitchison distance; STRESS, standardized residual sum of squares

* Corresponding author at: Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, 11A, Datun Road, Chaoyang District, Beijing 100101, China.
E-mail address: shiwj@lreis.ac.cn (W. Shi).

https://doi.org/10.1016/j.geoderma.2018.03.007
Received 23 August 2017; Received in revised form 4 March 2018; Accepted 10 March 2018
Available online 21 March 2018
0016-7061/

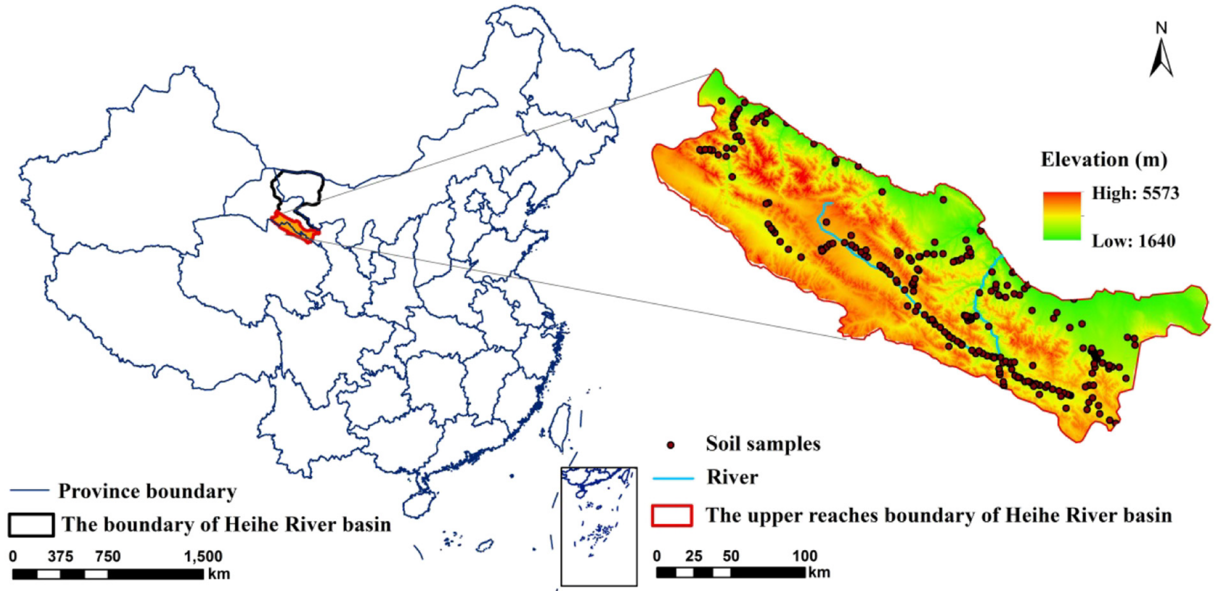

Fig. 1. The geographical location of the upper researches of Heihe River basin, China.

Traditional methods such as ordinary kriging (OK) and cokriging (COK) may not satisfy these requirements and do not guarantee that the sum of components is 1 (or 100%) (Odeh et al., 2003; Pawlowsky et al., 1995).

Several methods for mapping soil compositional data, such as log-ratio kriging and compositional kriging (CK) have been developed, and these methods have been compared in many previous studies (Huang et al., 2014; Odeh et al., 2003; Prasad et al., 2006; Saavedra et al., 2010; Salazar et al., 2015; Sun et al., 2014; Walvoort and de Gruijter, 2001; Zhang et al., 2013). In our previous study (Wang and Shi, 2017), we compared the performances of log-ratio cokriging (log-ratio CK) and log-ratio transformations (the additive log-ratio transformation (ALR) (Aitchison, 1986), centered log-ratio transformation (CLR) (Aitchison, 1986), isometric log-ratio transformation (ILR) (Egozcue et al., 2003; Filzmoser et al., 2009) and symmetry log-ratio transformation (SLR) (Odeh et al., 2003; Zhang et al., 2013)) combined with kriging, namely, ALR_OK, CLR_OK, ILR_OK and SLR_OK. The results showed that log-ratio CK had better mapping accuracy than the log-ratio kriging methods, and the maps generated by log-ratio CK were more suitable for the variable environmental patterns. When sampling is sparse, COK is particularly advantageous for mapping soil properties (Bishop and McBratney, 2001). Thus, cokriging in conjunction with Aitchison transformation (Aitchison, 1986) (log-ratio cokriging) has been proposed (Pawlowsky-Glahn and Olea, 2004; Pawlowsky et al., 1995), which consists of ALR_COK, CLR_COK, ILR_COK and SLR_COK. For example, Lark and Bishop (2007) found that ALR_COK had a similar performance to COK combined with untransformed data in terms of the standardized residual sum of squares (STRESS). Sun et al. (2014) compared ALR_COK, CK, COK and log-ratio CK in two case studies. Log-ratio CK had the best performance according to the integrated accuracy (IA) index, followed by ALR_COK, COK and CK in the first case, and the reverse results were found in the second case. Niang et al. (2014) had compared OK and COK combined with ILR and observed that ILR_COK performed better in terms of overall accuracy when mapping soil texture.

In addition, the standard estimator of the variogram in log-ratio cokriging is sensitive to the outlying values, which may result in the overestimation of the variogram and eventually lead to errors in the interpolators (Lark, 2000). Robust variogram estimators based on location and scale estimations were proposed by Lark (2003) to improve the variogram estimator. Log-ratio cokriging combined with robust variogram estimators (log-ratio RCOK) has also been applied by Lark and Bishop (2007), Sun et al. (2014) and Lark et al. (2012).

However, the performances of these methods were relatively uncertain, and only a few methods were used to compare their performance in previous studies. Moreover, previous studies did not consider the specific data characteristics when mapping soil psf. Thus, the conclusions drawn from the previously used methods may not guide researchers to choose the appropriate interpolator according to the data characteristics.

Although log-ratio transformations combined with OK, COK, RCOK and CK have achieved the goal of mapping soil psf, few comprehensive comparisons of different log-ratio transformation methods (ALR, CLR, ILR and SLR), different estimators of the variogram in COK and RCOK, and different CK methods (log-ratio CK and log-ratio RCK) have been made. Thus, the objectives of this study were (i) to comprehensively compare four log-ratio transformation methods, two estimators for variograms, and two CK methods, and (ii) to provide a reference for choosing a suitable interpolator according to the data characteristics by comparing the performances of different interpolators.

2. Materials and methods

2.1. The study area

The study area is mainly distributed in the Qilian mountain region in the northern area of the Tibetan Plateau and covers an area of 32,000 km^2 (Fig. 1). The elevation of the study area varies from 1640 m to 5573 m (Fig. 1), and the climate is cold and damp, with an annual mean temperature of < 4 °C and an average annual rainfall of approximately 350 mm. The main vegetation types are meadow and steppe (Fig. 2a), and the main forest types are Qinghai spruce and Qilian juniper. The main land use types are grassland and forest land (Fig. 2b). The aspect and slope are also shown in Fig. 2c and Fig. 2d. The main soil types in the soil genetic classification of China (Shi et al., 2004) are frigid felty soil, cold desert soil, gray-brown desert soil and castanozems (Fig. 2e).

2.2. Soil sampling

A total of 262 soil samples were collected to represent a range of different vegetation types, soil types, land use types, etc. in the upper