Ecological Informatics 16 (2013) 1–9

Contents lists available at SciVerse ScienceDirect

Ecological Informatics

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

Assessing the spatial uncertainty in soil nitrogen mapping through stochastic simulations with categorical land use information

Mingkai Qu [a,b], Weidong Li [c,*], Chuanrong Zhang [c]

[a] Department of Resource and Environmental Information, Huazhong Agricultural University, Wuhan 430070, China
[b] Key Laboratory of Soil Environment and Pollution Remediation, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China
[c] Department of Geography and Center for Environmental Sciences and Engineering, University of Connecticut, Storrs, CT 06269, USA

ARTICLE INFO

Article history:
Received 16 July 2012
Received in revised form 30 March 2013
Accepted 3 April 2013
Available online 12 April 2013

Keywords:
Sequential Gaussian simulation
Land use type
Uncertainty assessment
Soil nitrogen
Predictive mapping

ABSTRACT

This study explores the capability of an extended sequential Gaussian simulation algorithm with incorporation of categorical land use information (SGS-CI) for simulating spatial variability of soil total nitrogen (TN) contents and assessing associated spatial uncertainty. 402 sampled data in soil TN contents in a county scale region and the categorical land use map data of the study area were used to perform sequential simulations for comparing the SGS-CI algorithm and the conventional SGS algorithm, and 135 validation samples were used to assess the improvement of SGS-CI over SGS in prediction accuracy and uncertainty reduction. Results showed that the validation data were more strongly correlated with the optimal prediction (i.e., E-type estimates) data of SGS-CI than with those of SGS, and the mean error and the root mean square error of the optimal prediction using SGS-CI were smaller than those using SGS. SGS-CI also performed slightly better than SGS in uncertainty modeling in terms of accuracy plots and goodness statistic *G*. In addition, because demands for soil total nitrogen by different crops are usually different in agricultural practice, we showed that SGS-CI could be used to assess spatial uncertainty of deficiency or abundance degrees of soil TN based on demands of different crops in different land use types. Therefore, SGS-CI may provide an effective method for improving prediction accuracy and reducing uncertainty in soil TN prediction.

1. Introduction

Soil nitrogen is an important nutrient for maintaining the earth's ecosystems. Besides the nitrogenous fertilizers widely applied to farmlands for improving crop production, atmospheric deposition also represents an important source (Galloway et al., 2008; Kaiser, 2001). Sometimes the content of soil nitrogen may exceed the requirement of plant growth. This generally results in low nitrogen use efficiency and high nitrogen loss (Li and Zhang, 1999). The nitrogen loss from soils may further lead to negative impacts on the environment. For example, nitrogen losses to water bodies often cause water eutrophication—a serious ecological issue facing the environment nowadays (Carpenter et al., 1998; Lu et al., 2007; Smith et al., 2001). Therefore, effectively mapping the spatial distribution of soil nitrogen contents and the associated uncertainties inherent in spatial prediction are crucial to agricultural management, environmental management, and ecological management.

Geostatistics comprises a set of spatial statistical techniques, which have been widely used to characterize spatial variability of soil properties (Burgess and Webster, 1980; Ferguson et al., 1998; Li and Heap, 2011; Tutmez and Hatipoglu, 2010). However, besides spatial prediction, geostatistics also concerns quantifying the uncertainty associated with a spatial prediction (Bourennane et al., 2007; Diodato and Ceccarelli, 2006; Qu et al., 2013). Currently uncertainty assessment is mainly conducted using stochastic simulation algorithms (Bourennane et al., 2007; Goovaerts, 2001; Zhao et al., 2005). Sequential Gaussian simulation (SGS) is one of the most frequently used stochastic simulation algorithms for continuous variables. The increasing utilization of stochastic simulation algorithms in modeling uncertainties is justified by the fact that interpolation algorithms, such as kriging, yield a unique response—the interpolated map, which usually smoothes out local details of spatial variability of the attribute being mapped (Goovaerts, 1997). This shortcoming of kriging results in overestimation of small values and underestimation of large values. For these reasons, stochastic simulations are generally preferred to interpolations for applications where the spatial variation of the measured field needs to be preserved and uncertainty assessment is required. Indeed, stochastic simulation techniques, that provide multiple possible realizations of an unknown spatial distribution, do not aim to minimize a local error variance. Fluctuations between realizations provide a quantitative measure of the uncertainty about the underlying phenomenon.

An important issue, which we should not ignore, is the effect of some categorical factors (e.g., geological formations, land use types, or soil types) on the spatial variability of the target environmental or ecological variable. Earlier studies suggested that soil type information could be used to improve the prediction accuracy of some soil properties. For

* Corresponding author. Tel.: +1 860 486 2728.
E-mail address: weidongwoody@gmail.com (W. Li).

1574-9541/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ecoinf.2013.04.001

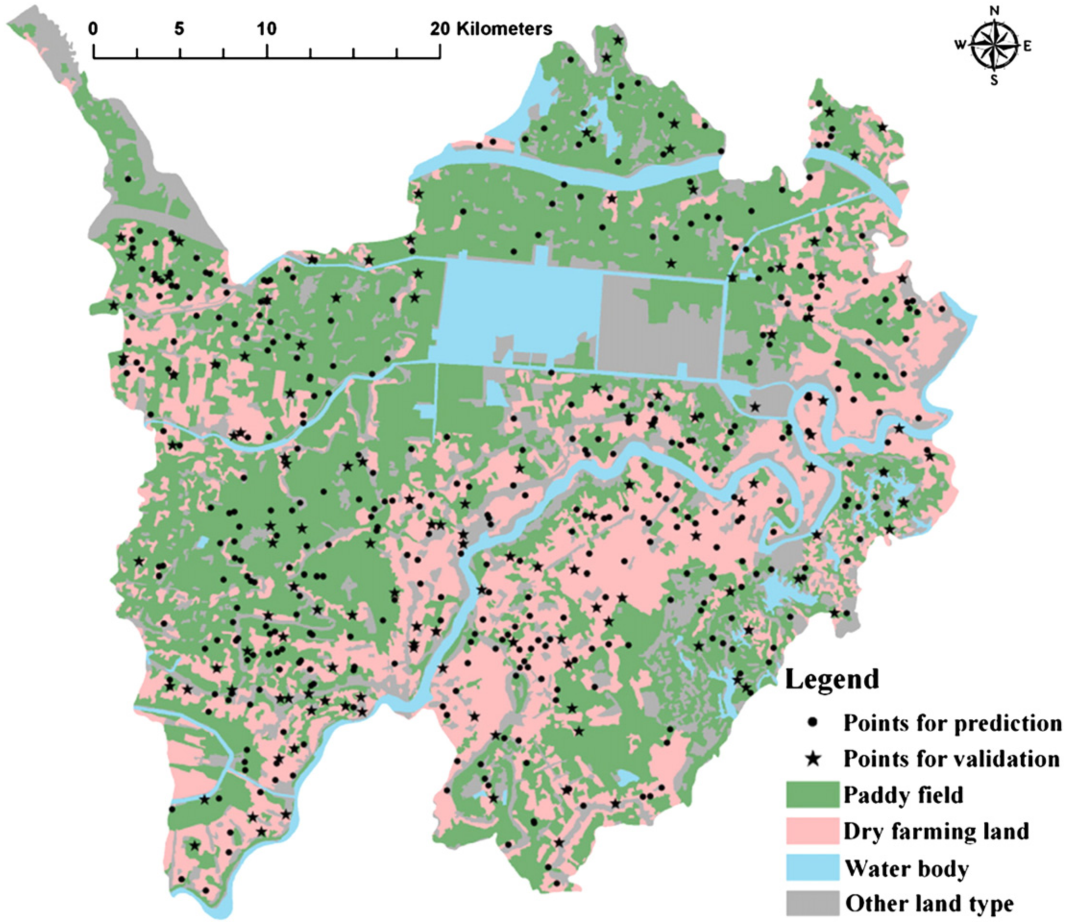

Fig. 1. Soil sample locations and land use type distribution.

example, Voltz and Webster (1990) and Van Meirvenne et al. (1994) used a method called stratified kriging (i.e., first stratify the survey area based on soil types and then perform kriging independently within each stratum) to improve the prediction accuracy of soil properties; Goovaerts and Journel (1995) used simple indicator kriging with varying means and indicator cokriging to incorporate the effect of soil types on interpolation of soil heavy metals, and found that the incorporation could improve the delineation of deficient areas; Liu et al. (2006) suggested a kriging combined with soil map-delineation (KSMD) method for incorporating the effect of soil types on several soil properties, and also took into account the contributions of both hard data and soil type data to the estimation variance. Recently, Goovaerts (2011) presented two approaches to incorporate both point and areal data in spatial interpolation of continuous soil attributes. Goovaerts (2010) also presented a general formulation of area-and-point kriging and demonstrated the effect of geological formations on soil heavy metals. Qu et al. (2012) recently investigated the effect of land use types on the spatial prediction of soil nitrogen using the area-and-point kriging method.

However, all of these studies focused on optimal interpolations rather than sequential simulations. As aforementioned, stochastic simulation algorithms such as SGS have the advantages in quantifying and visually displaying the uncertainty associated with spatial predictions. Given the influence of categorical factors such as land use types on local values of many soil properties, it is desirable to integrate the related categorical information into a geostatistical stochastic simulation algorithm such as SGS. Therefore, a satisfactory stochastic simulation for these variables should include two components—the spatial variation between different categories and the variability within each category. However, related studies in literature have been very rare so far. In this study, the conventional SGS algorithm was combined with categorical land use information for simulating the spatial distribution of soil total nitrogen (TN) contents in a study area and assessing the

Table 1
Soil TN content (g kg^{-1}) statistics for different land use types [a].

Land use type	Number	Range	Minimum	Maximum	Mean	SD	Skew	Kurt	CV
Total	402	2.73	0.31	3.04	1.46	0.54	0.47	−0.30	37.14
Paddy field	215	2.48	0.54	3.02	1.62	0.55	0.29	−0.56	33.82
Dry farmland	130	2.31	0.31	2.62	1.25	0.44	0.41	−0.09	35.18
Other land use type	57	2.46	0.58	3.04	1.38	0.58	0.75	0.12	41.80
Validation data	135	2.54	0.45	2.99	1.48	0.55	0.55	0.14	40.91

[a] SD—standard deviation; Skew—skewness; Kurt—kurtosis; CV—coefficient of variation (%).