Front. Earth Sci. 2011, 5(3): 239–244
DOI 10.1007/s11707-011-0174-1

RESEARCH ARTICLE

Regression-kriging for characterizing soils with remote-sensing data

Yufeng GE[1], J. Alex THOMASSON[1], Ruixiu SUI (✉)[2], James WOOTEN[3]

1 Department of Biological and Agriculture Engineering, Texas A&M University, College Station, TX 77843-2117, USA
2 USDA-ARS, Crop Production Systems Research Unit, Stoneville, MS 38776, USA
3 Department of Agricultural and Biological Engineering, Mississippi State University, Mississippi State, MS 39762, USA

Abstract In precision agriculture regression has been used widely to quantify the relationship between soil attributes and other environmental variables. However, spatial correlation existing in soil samples usually violates a basic assumption of regression: sample independence. In this study, a regression-kriging method was attempted in relating soil properties to the remote sensing image of a cotton field near Vance, Mississippi, USA. The regression-kriging model was developed and tested by using 273 soil samples collected from the field. The result showed that by properly incorporating the spatial correlation information of regression residuals, the regression-kriging model generally achieved higher prediction accuracy than the stepwise multiple linear regression model. Most strikingly, a 50% increase in prediction accuracy was shown in soil sodium concentration. Potential usages of regression-kriging in future precision agriculture applications include real-time soil sensor development and digital soil mapping.

Keywords precision agriculture, regression-kriging, remote sensing, soil sensors

1 Introduction

Evaluating in-field variability of soil properties is an essential part of much precision-agriculture research involving management-zone delineation (Barnes and Baker, 2000), real-time soil-sensor development (Hummel et al., 2001; Christy et al., 2003), variable-rate applications (Wollenhaupt et al., 1994; Bajwa and Tian, 2001), etc. In many of these studies, regression has been applied extensively to quantify the relationships among soil attributes (such as moisture content, texture, pH, organic matter, nitrogen, and micronutrients) and between these soil attributes and related environmental variables including elevation, crop bio-physical parameters, remotely sensed data, etc.

One of the potential problems overlooked in using regression is that regression assumes sample (or residual) independence, which is violated when spatial correlation exists among soil samples collected from a field. Numerous studies have shown different degrees of spatial correlation in virtually all soil attributes under investigation (Cambardella and Karlen, 1999; Geypens et al., 1999; Solie et al., 1999; Iqbal et al., 2005). Researchers have started to address the problem by incorporating kriging into regression. Depending on how the two techniques are coupled, the hybrid method is named by different scientists as kriging with external drift, regression-kriging, and universal kriging. Advantages of using the hybrid method for quantitative soil characterization have been successfully demonstrated in the literature (Odeh et al., 1994; Knotters et al., 1995; Odeh et al., 1995; Bishop and McBratney, 2001; Hengl et al., 2004; Sullivan et al., 2005). Kitanidis (1993), Hengl et al. (2003), and McBratney et al. (2003) have documented the statistical foundation of this approach.

Although applied successfully in other disciplines, the combination of kriging and regression for evaluating in-field variability of soil properties is new in the agricultural engineering community. The main objective of this study was to apply regression-kriging for characterizing soil properties with remotely sensed images of an agricultural field and compare the results to those with simple-linear regression.

2 Materials and methods

The study site was a cotton field (referred to as field 1) in Quitman County, Mississippi, USA. A bare-soil

Received April 10, 2011; accepted May 11, 2011

E-mail: ruixiu.sui@ars.usda.gov

LANDSAT ETM (Enhanced Thematic Mapper) image of the field was acquired. A total of 273 soil samples were collected from an equilateral grid (each grid cell representing roughly 0.4 ha) in the field (Fig. 1). Soil samples were taken to the laboratory and assayed with the hydrometer method for textural composition (clay and sand percentage) and the Mississippi Soil Test and Mehlich 3 methods for a series of micronutrient concentrations (Ca, Na, Mg, K, P, Zn). A detailed description of soil sample collection and analysis can be found in Thomasson et al. (2001). Soil samples were split into two sets: 191 samples (or 70%) were randomly selected for calibration and the remaining 82 (or 30%) were used for validation. Figure 1 shows a color infrared (CIR) image of the study field, overlaid with the field boundary and the systematic grid of soil sampling locations.

Stepwise Multiple Linear Regression (SMLR) was used to relate soil properties to the multi-spectral bands of the LANDSAT image. The thermal band (Band 6) was not included because of its different pixel size, so six independent variables were included as regressors. *P*-values for variable inclusion and removal were 0.05 and 0.1, respectively.

Semivariogram analysis was applied to the regression residuals (the difference between the predicted and actual values) to quantify the spatial structure of each variable. Empirical semivariogram was first computed with Metheron's (Matheron, 1965) Method of Moments estimator. The empirical semivariograms were then fitted (based on a weighted least squares approximation) with theoretical models that provide three parameters: c_0, the nugget variance, $c_0 + c_1$, the sill variance, and a, the range of spatial dependence. The spherical model (Eq. (1)), which fitted the empirical semivariograms better than other models, was chosen for all variables. The spatial structure (or the covariance matrix) of residuals was then used to define the regression-kriging model.

$$\gamma(h) = \begin{cases} 0 & h = 0 \\ c_0 + c_1\left[\dfrac{3h}{2a} - \dfrac{1}{2}\left(\dfrac{h}{a}\right)^3\right] & h < a \\ c_0 + c_1 & h \geqslant a \end{cases}, \tag{1}$$

where, $\gamma(h)$: semivariance.

As its name indicates, regression-kriging is consisted of a regression part and a kriging part as shown in the following equation:

$$z(s_0) = \underbrace{\sum_{k=0}^{p} \beta_k \cdot l_k(s_0)}_{\text{regression}} + \underbrace{\sum_{i=1}^{n} \omega_i(s_0) \cdot \varepsilon(s_i)}_{\text{kriging}}, \tag{2}$$

where l represents multiple (p) independent variables; β is the regression coefficients ($\beta_0 \cdot l_0$ is the intercept term of regression); ε represents the regression residuals of n known samples surrounding the unknown sample $z(s_0)$; and ω represents the kriging coefficients.

Both SMLR and regression-kriging models were evaluated with the validation data. A model's goodness-of-fit was determined mainly by the root mean squared error (RMSE) as shown in Eq. (3).

$$\text{RMSE} = \sqrt{\frac{1}{N} \times \sum_i \left[z(i)_{\text{predicted}} - z(i)_{\text{measured}}\right]^2}, \tag{3}$$

where $z(i)_{\text{predicted}}$ and $z(i)_{\text{measured}}$ are predicted and measured values of a soil sample, and N is the number of soil sample in the validation set.

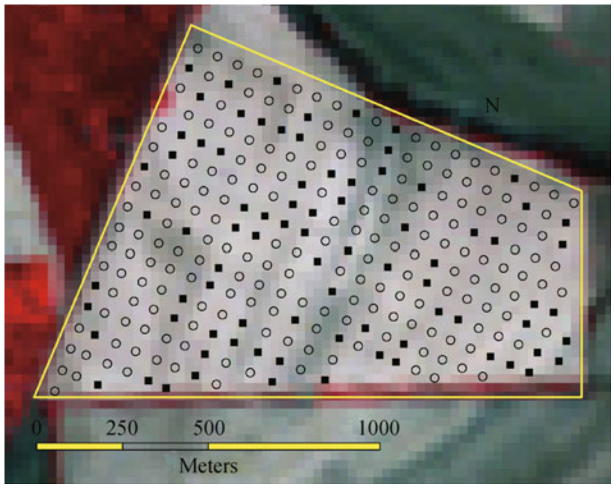


Fig. 1 Field boundary and soil samples (○ – calibration set, ▪ – validation set) superimposed on color-infrared composite image of field from LANDSAT data

3 Results and discussion

Particular attention was paid to four soil properties: clay, sand, Ca, and Na. Their summary statistics including mean, standard deviation (STD), and coefficient of variation (CV) are given in Table 1. As can be seen, the soil properties are quite variable in the field, with CVs ranging from 21% for Ca to nearly 50% for Na.

Table 2 presents the SMLR model for each soil property with selected LANDSAT spectral bands. The Ca and clay models had R^2 (coefficient of determination) values of 0.51 and 0.32, demonstrating moderate prediction power of the SMLR model for these soil properties. The sand and Na models had R^2 values of 0.17 and 0.09, indicating poor prediction power.

The results of semivariogram analysis of prediction residuals are shown in Fig. 2. To evaluate the degree of spatial correlation on a relatively equal basis, percent nugget, which is equal to nugget/sill × 100%, is calculated. A strong spatial correlation is indicated by a low percent

Table 1 Summary statistics of selected soil properties

Soil property	Mean	STD	CV/%
Clay/%	37.4	15.3	41
Sand/%	32.5	9.0	28
Ca/($mg \cdot kg^{-1}$)	2116.4	442.2	21
Na/($mg \cdot kg^{-1}$)	92.1	44.4	48

Table 2 Results of the SMLR model, relating soil properties to LANDSAT spectral bands

Soil property	R^2	Model
Clay/%	0.32	Clay = 2.10 − 0.00813 × Band 5
Sand/%	0.17	Sand = −0.0377 − 0.00587 × Band 2 + 0.00571 × Band 7
Ca/($mg \cdot kg^{-1}$)	0.51	Ca = 8149 + 44.2 × Band 2 − 54.8 × Band 4 − 15.5 × Band 5
Na/($mg \cdot kg^{-1}$)	0.09	Na = 158.7 − 5.41 × Band 4 + 2.3 × Band 5

nugget and a large range. Based on this criterion, residuals of Na show strong spatial correlation with a percent nugget of 20% and a range of 770 m (The study field has a diagonal length of about 1600 m). Moderate spatial correlation is shown in residuals of sand (with the percent nugget of 65% and range of 445 m) and Ca (with the percent nugget of 54% and range of 390 m). Residuals of clay are interpreted as primarily a nugget effect, with a high percent nugget (80%) and a very short range (150 m).

Table 3 compares RMSE of the SMLR and regression-kriging models as tested on the validation data. It is clear that both Na and Ca have sizeable decreases in RMSE, indicating superiority of regression-kriging over SMLR in these cases. Sand has only a slight improvement in RMSE.

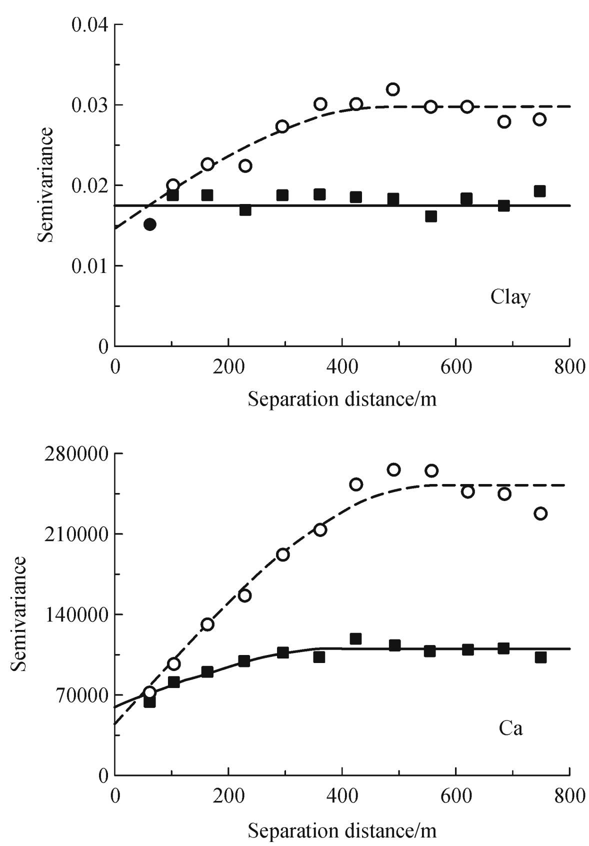


Fig. 2 Sample semivariogram (■) and fitted spherical model (–) of prediction residuals of each soil property. For comparison, the sample semivariogram (○) and fitted spherical model (– –) of the original variables are also shown

Table 3 RMSE of SMLR and regression-kriging model tested on validation set

Method	Clay/%	Sand/%	Ca/(mg·kg^{-1})	Na/(mg·kg^{-1})
SMLR	11.8	8.2	317	44.3
Regression-kriging	12.1$^{a)}$	8.0	276	29.7

Note: a) The regression-kriging model for clay was calculated with residuals estimates from inverse-distance-weighted interpolation

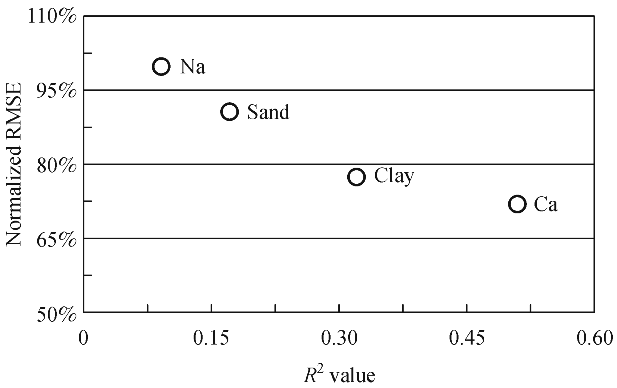


Fig. 3 Relationship between SMLR model's R^2 value and the normalized RMSE from the validation data

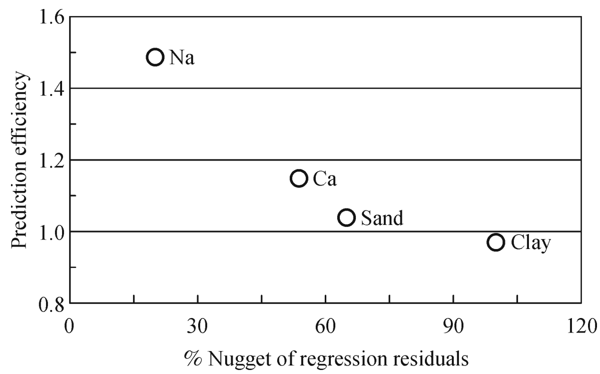
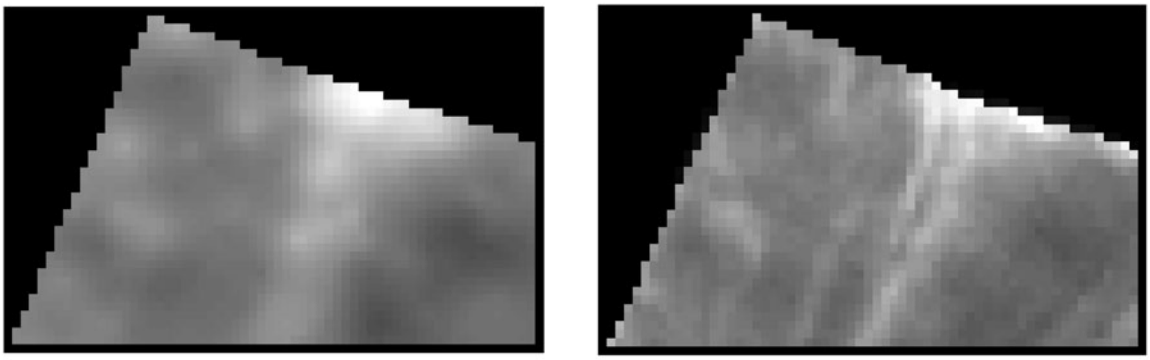


Fig. 4 Relationship between prediction efficiency of regression-kriging over SMLR and residual percent nugget

Since residuals for clay indicate only a nugget effect, the regression-kriging model for clay cannot be specified. To make the comparison possible, an inverse-distance-weighting (IDW) technique with a power of two was applied on its regression residuals, and the result is reported in the regression-kriging row of Table 3. An increase of RMSE in clay could be explained as follows: because there is no spatial structure in clay residuals, applying spatial interpolation techniques (such as kriging or IDW) to regression residuals of neighboring points and adding them to the regression is not helpful in improving prediction accuracy.

The R^2 value of SMLR models in the calibration data was plotted against the normalized RMSE in the validation data for all soil-property values (Fig. 3). Normalization was by ratioing RMSE to the overall STD for a given soil property and reported as a percentage. It is clear from Fig. 3 that higher R^2 values in calibration led to lower RMSEs in validation. To evaluate how integrating the kriging term (regression-kriging) improved prediction accuracy over SMLR, prediction efficiency of regression-kriging was plotted against residual percent nugget (Fig. 4). Prediction efficiency was calculated as RMSE of SMLR (the second row of Table 3) divided by RMSE of regression-kriging (the third row of Table 3). Higher prediction efficiency indicates a greater improvement in regression-kriging. Clearly, a strong spatial correction exhibited in the regression residuals (small percent nugget) was related to high prediction efficiency. Considering Figs. 3 and 4 together, the following conclusions may be drawn: 1) if weak spatial correlation exists in regression residuals, simple regression is adequate (such as with clay in this study); 2) if strong spatial correlation exists, regression-kriging substantially improves prediction accuracy (such as with Na); 3) if moderate spatial correlation exists, regression-kriging is still beneficial in improving prediction accuracy (such as with sand and Ca).

The method of regression-kriging has important practical implications in precision agriculture. Real-time soil sensors developed for precision agriculture usually require a specific calibration to improve accuracy. This procedure involves primarily collecting soil samples from various locations in the field, measuring target soil properties in the laboratory, and constructing a regression line to relate sensor outputs (such as reflectance readings of an optical sensor) to the actual value. Conceivably, regression-kriging can also be used in this calibration procedure: where not only a regression line, but also the spatial structure of the residuals, will be modeled based on the location information of the known samples. These models could be further used jointly (as in Eq. (2)) if a sample at a new location is encountered. It is noteworthy that no extra physical effort would be incurred for regression-kriging, because using GPS receivers to record the location information of soil samples has already been a routine practice with real-time sensors. Another possible usage of regression-kriging is to map soil properties based on other baseline data (such as a digital elevation model, remote-sensing imagery, and electrical-conductivity maps). Figure 5 shows Ca-concentration maps predicted from the LANDSAT image based on SMLR and regression-kriging. Differences between the maps are discernible; the



(a) (b)

Fig. 5 Ca maps developed with SMLR (a) and regression-kriging (b); brighter areas are associated with higher concentrations

map developed with SMLR is smoother than the one developed with regression-kriging. Indeed, the preceding analysis already showed that regression-kriging is more accurate for Ca than SMLR. It is anticipated that regression-kriging can be included into commercial GIS software (such as ArcGIS) for digital soil mapping purposes.

4 Conclusions

A bare-soil remote-sensing image was used with two different estimation methods to characterize several soil properties of a cotton field in Mississippi, USA and the following conclusions were drawn:

1) Using regression-kriging to account for spatial correlation among sample residuals generally increases prediction accuracy over simple regression.

2) The degree of prediction improvement with regression-kriging is proportional to the degree of spatial correlation shown in the residuals; the higher the spatial correlation, the larger the prediction improvement. In the extreme case, in which residuals represented mainly a nugget effect (such as clay), no improvement was obtained.

References

Bajwa S G, Tian L (2001). Aerial CIR remote sensing for weed density mapping in a soybean field. Trans ASAE, 44(6): 1965–1974

Barnes E M, Baker M G (2000). Multispectral data for mapping soil texture: Possibilities and limitations. Appl Eng Agric, 16: 731–741

Bishop T F A, McBratney A B (2001). A comparison of prediction methods for the creation of field-extent soil property maps. Geoderma, 103(1–2): 149–160

Cambardella C A, Karlen D L (1999). Spatial analysis of soil fertility parameters. Precis Agric, 1(1): 5–14

Christy C, Drummond P, Laird D (2003). An On-the-go Spectral Reflectance Sensor for Soil. Paper No. 031044, ASAE, Mich: St. Joseph

Geypens M, Vanongenval L, Vogels N, Meykens J (1999). Spatial variability of agricultural soil fertility parameters in a gleyic podzol of Belgium. Precis Agric, 1(3): 319–326

Hengl T, Heuvelink G B M, Stein A (2003). Comparison of kriging with external drift and regression kriging. Technical Report. International Institute for Geo-information Science and Earth Observation. Available at: www.itc.nl/library/academic_output. Accessed 11 Jun 2007

Hengl T, Heuvelink G B M, Stein A (2004). A generic framework for spatial prediction of soil variables based on regression-kriging. Geoderma, 120(1–2): 75–93

Hummel J W, Sudduth K A, Hollinger S E (2001). Soil moisture and organic matter prediction of surface and subsurface soils using an NIR soil sensors. Comput Electron Agric, 32(2): 149–165

Iqbal J, Thomasson J A, Jenkins J N, Owens P R, Whisler F D (2005). Spatial variability analysis of soil physical properties of alluvial soils. SSSAJ, 69(4): 1338–1350

Kitanidis P K (1993). Generalized covariance function in estimation. Math Geol, 25(5): 525–540

Knotters M, Brus D J, Voshaar J H O (1995). A comparison of kriging, co-kriging, and kriging combined with regression for spatial interpolation of horizon depth with censored observations. Geoderma, 67(3–4): 227–246

Matheron G (1965). Les Variables Re'gionalise'es et leur Estimation: Une Application de la the'orie de Fonctions Ale'atoires aux Sciences de la Nature. Paris: Masson et Cie

McBratney A B, Santosm M L M, Minasney B (2003). On digital soil mapping. Geoderma, 117(1–2): 3–52

Odeh I O A, McBratney A B, Chittleborough D J (1994). Spatial prediction of soil properties from landform attributes derived from a digital elevation model. Geoderma, 63(3–4): 197–214

Odeh I O A, McBratney A B, Chittleborough D J (1995). Further results on prediction of soil properties from terrain attributes: Heterotopic co-kriging and regression-kriging. Geoderma, 67(3–4): 215–226

Solie J B, Raun W R, Stone M L (1999). Submeter spatial variability of selected soil and bermudagrass production variables. SSSAJ, 63(6): 1724–1733

Sullivan D G, Shaw J N, Rickman D (2005). IKONOS imagery to

estimate surface soil property variability in two Alabama physiographies. SSSAJ, 69: 1789–1798
Thomasson J A, Sui R, Cox M S, Al-Rajehy A (2001). Soil reflectance sensing for determining soil properties in precision agriculture. Trans ASAE, 44(6): 1445–1453
Wollenhaupt N C, Wolkowski R P, Clayton M K (1994). Mapping soil test phosphorous and potassium for variable-rate fertilize application. J Prod Agric, 7(4): 441–447