GIS-BASED ELABORATE SPATIAL PREDICTION OF SOIL NUTRIENT ELEMENTS USING ANCILLARY TERRAIN DATA IN CHONGQING TOBACCO PLANTING REGION, CHINA

Xuan Wang [1,2], Jiake Lv [1,2], Chaofu Wei [3], Deti Xie [3,*]

[1] *Chongqing key laboratory of digital agriculture, Chongqing, 400716,P.R.China*

[2] *College of computer and information science, Southwest University, Chongqing, 400 715,P.R.China*

[3] *College of resources and environment, Southwest University, Chongqing, 400716,P.R.China*

[*] *Corresponding author, Address: College of resources and environment, Southwest University, Chongqing 400716,P.R.China, Tel: +86-23-13883372436, Email: wangxuan @swu.edu.cn*

Abstract: The precision agriculture hopes to manage the variation in soil nutrient status continuously, which requires reliable predictions at places between sampling sites. For the long time, ordinary kriging has been used as one prediction method when the data are spatially dependent and a suitable variogram model exists. However, even if data are spatially correlated, there are often few soil sampling sites in relation to the area to be managed. Recently, Digital elevation models(DEMs) and remotely sensed data are becoming more readily available, these data are usually far more intensive than those from soil surveys. If these ancillary data are coregionalized with the sparse soil data, they might be used to increase the accuracy of predictions of the soil properties.

Under ArcGIS platform, this paper employed spatial predictions of the soil total N, P, K in Chongqing tobacco planting region, China, with cokriging and regression kriging respectively. For the both, intensive terrain data including elevation, slope and aspect were used with the soil data. Traditional ordinary kriging(OK) was investigated as comparison basis to determine which approach is appropriate for different soils properties mapping. And the results suggest that the use of intensive ancillary data can increase the accuracy of predictions of soil properties in arable fields provided that the variables are related spatially.

Please use the following format when citing this chapter:

Wang, X., Lv, J., Wei, C. and Xie, D., 2009, in IFIP International Federation for Information Processing, Volume 293, *Computer and Computing Technologies in Agriculture II, Volume 1*, eds. D. Li, Z. Chunjiang, (Boston: Springer), pp. 345–356.

Keywords: Ordinary kriging, Cokriging, Regression kriging, Ancillary terrain data, Soil nutrient elements, Elaborate spatial prediction

1. INTRODUCTION

One aim of regional precision agriculture is to use information about spatial distribution of soil properties to manage the field in a landscape-specific way, for example, by the application of nutrients at the places and in the amounts required(Sylvester-Bradley et al., 1999) . Sampling at discrete places is the traditional means of obtaining information about the soil properties. However, field surveys are often time-consuming, labor-intensive and costly although it is still the dominant way to obtain information about most soil properties. Predicting unsampled sites by available sampling points data with acceptable precision and cost hence has been the focus of pedometrics(Webster, 1985; Cambardella et al., 1994; Goovaerts, 1997; McBratney et al., 2000; Auemhammer, 2001) .Traditionally, this has been achieved by classification, but it has been known for some time that this approach does not describe adequately the variation that scientists and farmers are aware of intuitively(Webster and Oliver, 2001) .Ordinary kriging(OK) is quiet suited to this problem provided that the data are spatially dependent(Webster and Oliver, 2001) . However, the data from soil surveys, are often sparse even though they might be spatially autocorrelated and it can lead to considerable uncertainty in the kriging prediction(McBratney et al., 2000) . Also, the precision of the prediction was shown to be very sensitive to the density of observation points and the gains in precision were only sight for sites located near from the observation sites(Auemhammer, 2001) .

A way of refining these predictions could be to use easy-to-measure auxiliary data describing landscape features. Particularly, the use of attributes derived from Digital Elevation Models(DEMs) which becomes more readily available looks promising since some researchers have demonstrated the relationships between landforms features and soil properties(Chaplot et al., 2004; Lopez-Granados et al., 2005; Kay and Rainer, 2008; Wu et al., 2008) . Recently several predictive soil attributes techniques that use intensive auxiliary data also have been developed ,for example, cokriging and regression kriging.

Chongqing tobacco planting region, which locates in one of the poorest areas of southwest China and mostly is covered by hilly and mountains. Understanding how soil nutrients vary across landscape positions especially increasing the accuracy of spatial prediction of soil nutrient elements would be of huge economic and environmental benefit for local tobacco planting and management. This study predicted spatial distribution of total N, P, K of this region by cokriging and regression kriging respectively in which intensive terrain properties including elevation, slope and aspect were used as auxiliary data. Traditional ordinary kriging interpolation was investigated as comparison basis to determine which approach is the appropriate method for different soil nutrient elements spatial distribution.

This paper is organized as follows. Following the introduction, theory and methods used in this study are described in section 2.The study area and data are depicted in section 3.In section 4, results and discussion are illustrated. Conclusion is finally made in section 5.

2. THEORY AND METHODS

Ordinary kriging, cokriging and regression kriging were used in this study. The last two are described briefly below(for ordinary kriging, see (Webster and Oliver, 2001))

2.1 Cokriging

Cokriging is an extension of ordinary kriging that takes into account the spatial cross-correlation from two or more variables. The usual situation is one where the primary or target variable, $Z_u(x)$, has been measured at many fewer places, x, than the secondary one, $Z_v(x)$, with which it is coregionalized. We assume that they obey the intrinsic hypothesis. Both variables have an autovariograms, for variable u, this is

$$\gamma_{uu}(h) = \frac{1}{2}\left[\{z_u(x) - z_u(x+h)\}^2\right] \tag{1}$$

Where h is a vector, the lag.For v also, the expected differences are zero and its autovariogram is $\gamma_{vv}(h)$.The two variables have a cross-variogram, $\gamma_{uv}(h)$, defined as

$$\gamma_{uv}(h) = \frac{1}{2}E\left[\{Z_u(x) - Z_u(x+h)\}\{z_v(x) - z_v(x+h)\}\right] \tag{2}$$

Which describes the way in which u is spatially related to v.

To compute the usual cross-variogram, there must be sites where both u and v have been measured, i.e., collocated. The experimental cross-variogram, $\hat{r}_{uv}(h)$ can be estimated by

$$\hat{\gamma}_{uv}(h)=\frac{1}{2m(h)}\sum_{i=1}^{m(h)}\{z_u(x_i)-z_u(x_i+h)\}\{z_v(x_i)-z_v(x_i+h)\} \tag{3}$$

Where z_u and z_v have been measured at sites x_i and x_i+h ,and $m(h)$ is the number of pairs of data points separated by the particular lag vector h .

The cross-variogram can be modeled in the same way as the autovariogram, but there is an added condition. Any linear combination of the variables it itself a regionalized variable, and its variance must be positive or zero. This is ensured if we adopt the linear model of coregionalization. For any pair of variables u and v , the variogram is

$$\gamma_{uv}(h)=\sum_{k=1}^{k}b_{uv}^{k}g_k(h) \tag{4}$$

Where the b_{uv}^{k} are the variances, for example the nugget and sill variances. $g_k(h)$ is the spatial autocorrelation function which must be the same for both variables being analyzed.

The ordinary punctual cokriging prediction of the primary variable, Z_u is obtained from the linear sum

$$\hat{Z}_u(x_0)=\sum_{i=1}^{V}\sum_{i=1}^{n_l}\lambda_{il}z_l^{ok}(x_i) \tag{5}$$

Where there are V variables, $l=1,2,...,V$,of which u is the one to be predicted, and the subscript i refers to the sites of which there are n_l in the search neighborhood where the variable l has been measured. The λ_{il} are the weights, which in the case of 'classical 'cokriging (Goovaerts, 1997) satisfy

$$\sum_{i=1}^{n_l}\lambda_{il}=\begin{cases}1, & l=u\\ 0, & l\neq u\end{cases} \tag{6}$$

There are the non-bias conditions, and subject to them the weights, λ_{il} ,that minimize the estimation variance of $\hat{Z}_u$ for a point, x_0 ,are found by solving the kriging system for all $v=1,2,...,V$ and all $j=1,2,...n_v$ .The weights λ_i are inserted into Eq.(5) to estimate $\hat{Z}_u(x_0)$.

2.2 Regression kriging

Odeh et al.(Odeh et al., 1995) describe three types of regression kriging: model A, B and C which are developments of the general theme. For this study, we used model C, which we summarize here. The method is based on a linear regression between a target variable such as certain soil property (Z), and a secondary or third variable, such as elevation or slope (Y_i).The regression model so obtained is used to predict Z to the locations of the prediction grid at which Y_i is known. The residuals from the regression ε are kriged to the prediction grid using the variogram computed from the residuals. The predicted values $\hat{Z}_R$ and the kriged values of the residuals $\hat{\varepsilon}_{ok}$ are summed to give the predicted values of the target variable $\hat{Z}_{RK}$.

$$\hat{Z}_{RK}(x) = \hat{Z}_R(x) + \hat{\varepsilon}_{ok}(x) \tag{7}$$

2.3 Performance evaluation indicators

In order to evaluate the performance of different spatial prediction methods, mean squared error(MSE) and root mean squared standardized effect(RMSSE), in this paper are used as performance measure indicators.

The mean squared error(MSE) is expressed as

$$MSE = \frac{1}{l}\sum_{j=1}^{l}\left[z_1(x_i) - z_2(x_i)\right] \tag{8}$$

The root mean squared standardized effect(RMSSE) is expressed as

$$RMSSE = \sqrt{\frac{1}{1}\sum_{i=1}^{l}\left[z_1(x_i) - z_2(x_i)\right]^2} \tag{9}$$

Where $z_1(x_i)$ is standardized site true value and $z_2(x_i)$ represents standardized site prediction value, l is the number of validation sites. When MSE is more close to 0 and RMSSE more approaches 1, the accuracy of prediction is hold higher.

3. THE STUDY AREA AND DATA SOURCES

3.1 The study area

The study area is located in the east part of Chongqing between north latitudes 28° 09′ and 32° 12′ and east longitudes 106° 23′ and 110° 11′ (Fig.1).The climate is characterized by an average annual temperature between 10.1 and 18.2°C .Annual precipitation is about 1200mm. According to Chinese Soil Taxonomy, the soils are classified by yellow soil, yellow brown soil, limestone soil, purple soil and paddy soil.

The landform of study area is dominated by hills, low mountains and medium mountains. Generally, it is undulating with slopes ranging from 0 to 84 ° and altitudes within the range from 100 to 2750.92 m. Due to weak transportation and economy development, tobacco planting is one of local dominant crops productions.

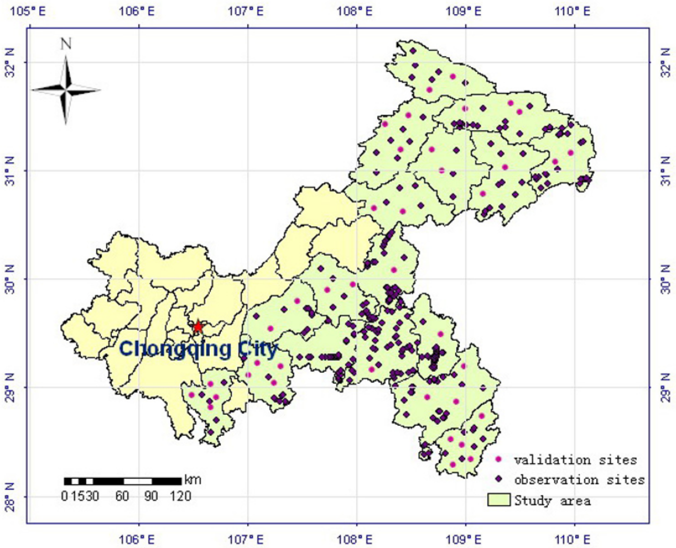


Fig.1: Location of the study area

3.2 Data sources

The soil was sampled in the spring of 2007 at 300 sites of 17 tobacco planting counties at a depth of 0-20cm. The position of each site was georeferenced using a DGPS(Differential Global Positioning System). From the observations, 40 sites, randomly selected, were kept aside for validation(the validation set).Therefore, the 260 remainder sites formed the prediction set.

The field soil was analyzed for total Nitrogen(N),Phosphorus(P) and K_2O as soon after sampling as possible. Kjeldahl method was utilized to measure the soil total N , and the total P and K_2O were determined by x-ray fluorescence(XRF) analysis(Lu and Yang, 1993) .

A DEM was constructed using ArcMAP Version 9.2(ESRI). 50×50m DEM data were derived from automated image matching of scanned panchromatic aerial photograph based on AUNDEM(Auemhammer, 2001) . Primary terrain attributes including elevation, slope and aspect were extracted by ArcMAP spatial analyst module.

4. RESULTS AND DISCUSSION

4.1 Exploratory data analysis

Table 1 list the descriptive statistics of total N,P and K_2O in the study area. The K-S test indicates that the whole dataset all followed a normal distribution.

Table 1. Descriptive statistics for Total N,P, K_2O

Item	Mean	Min	Max	Skewness	S.D	K-S
Total N(ug/g)	1534.40	798.64	2933.	0.89	356.70	0.21
Total P(ug/g)	717.50	299.10	1448.90	0.78	202.34	0.17
Total K_2O(%)	2.41	0.68	5.66	0.65	0.74	0.15

The Pearson product moment correlation coefficients were calculated between the total N,P, K_2O and elevation, slope and aspect from the co-located data to determine whether it was feasible to use these terrain data to improve their prediction. Table 2 gives the correlation coefficients of N,P and K_2O with terrain data. For all three soil nutrients, they all related to elevation and ranged from weak for K_2O,to moderate for P and strong for N. For slope, both N and P were inversely related but K_2O is likely to be indirect. Besides K_2O, the other soil elements have no obvious links with aspect.

Table 2. Pearson product moment correlation coefficients between soil nutrients and terrain data

Item	Elevation	Slope	Aspect
Total N	0.295*	0.173*	-0.026
Total P	0.179*	0.116*	-0.051
Total K_2O	0.113*	0.075	.-132*

4.2 Cokriging

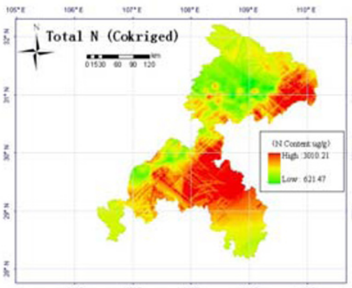
Experimental cross- and auto-variograms were obtained by applying Eq.(4) using the DEM predictions of elevation, slope and aspect at the observation sites. We fitted the models of coregionalization to soil total N,P, K_2O and three terrain attributes. It is shown as in table 3.

It was indicated that the two sets variograms of elevation, slope for N and P were both bounded and were fitted by exponential, spherical ,Gaussian, spherical function respectively. The variograms of aspect for N and P were unbounded, whereas it it was bounded for K_2O with the range set to

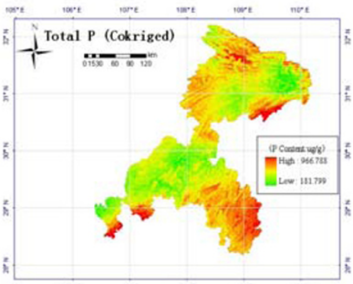
121.04.Under the ArcGIS platform, elevation, slope were associated as covariances for N,P spatial prediction and elevation, aspect act covariances for K_2O spatial prediction. The whole prediction maps were shown in Fig.2.

Table3 Varioram model parameters with elevation, slope and aspect of soil nutrients

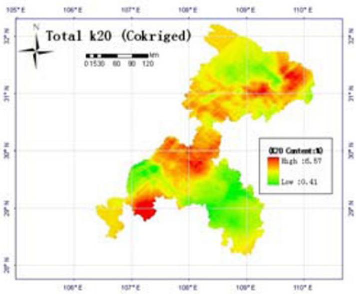
Item	Variable	Model	Nugget	Sill	Range(Km)
N	Elevation	Exponential	0.69	0.90	141.163
	Slope	Spherical	1.13	1.21	320.606
	Aspect	Linear	1.21		
P	Elevation	Gaussian	0.71	0.88	118.198
	Slope	Spherical	1.13	1.21	320.606
	Aspect	Linear	1.21		
K_2O	Elevation	Exponential	0.61	0.89	117.379
	Slope	Linear	0.71		
	Aspect	Spherical	0.62	1.01	121.04



(a) Total N-Cokriged



(b) Total P-Cokriged



(c) Total K_2O-Cokriged

Fig.2 Spatial distribution for total N, P,K_2O with cokriging

4.3 Regression kriging

Linear regressions of total N,P, K_2O were done on the DEM data at the observation sites. The equations were depicted as follows:

$$Y_1 = 0.482h + 1.8\alpha - 0.151\beta + 128.067 \; (r^2 = 0.635) \tag{10}$$

$$Y_2 = 0.32h + 0.751\alpha + 0.205\beta - 263.189 \; (r^2 = 0.517) \tag{11}$$

$$Y_3 = 0.001h + 0.003\alpha - 0.001\beta + 2.254\,(r^2 = 0.611) \tag{12}$$

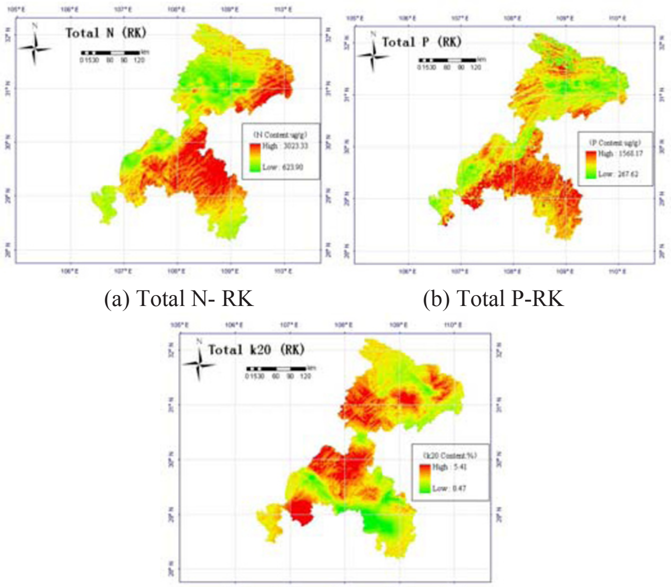
where Y_1, Y_2 and Y_3 represents total N, P, K_2O respectively, h, α, β represents elevation, slope and aspect.

An experimental variogram was also computed on the residuals of total N, P and K_2O from the regression at each site. It is shown as table 4.

Table 4 Parameters for the fittest residuals theoretical models of total N,P,K_2O

Item	Model	Nugget	Sill	Range(km)	C/C_0+C	R^2
N residuals	Spherical	7.94	10.99	211.40	0.72	0.872
	Exponential	6.05	10.61	114.601	0.57	0.958
	Gaussian	7.11	10.41	75.820	0.68	0.811
P residuals	Spherical	11.77	29.49	81.259	0.40	0.672
	Exponential	4.94	29.68	78.022	0.17	0.713
	Gaussian	15.23	29.53	72.210	0.52	0.966
K_2O residuals	Spherical	0.15	0.23	72.651	0.65	0.753
	Exponential	0.12	0.24	57.718	0.50	0.821
	Gaussian	0.16	0.23	54.479	0.70	0.934

It was found that the highest determined coefficient(R^2) existed in Exponential, Gaussian, Gaussian Model for total N,P, K_2O residuals respectively. The whole spatial prediction map is displayed in Fig.3.



(a) Total N- RK (b) Total P-RK

(c) Total K_2O-RK

Fig.3 Spatial distribution for total N, P,K_2O with regression kriging

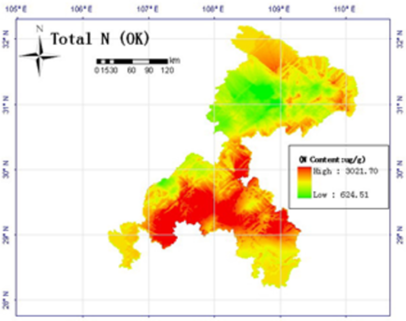
4.4 Ordinary kriging

Different semivariogram models for models for the soil N,P,K_2O were presented in Table 5.

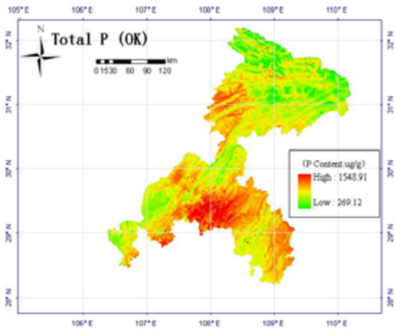
Table 5 Parameters for the fittest residuals theoretical models of total N,P,K_2O

Item	Model	Nugget	Sill	Range(km)	C/C_0+C	R^2
	Spherical	0.94	1.29	11.811	0.73	0.731
N	Exponential	0.77	1.31	10.235	0.59	0.742
	Gaussian	1.02	3.85	10.496	0.26	0.867
	Spherical	2.39	4.50	10.983	0.53	0.821
sP	Exponential	1.89	4.53	11.378	0.42	0.910
	Gaussian	2.74	4.51	9.437	0.61	0.877
	Spherical	4.91	5.68	121.766	0.86	0.726
K_2O	Exponential	4.11	5.63	63.034	0.73	0.891
	Gaussian	4.73	5.62	58.767	0.84	0.677

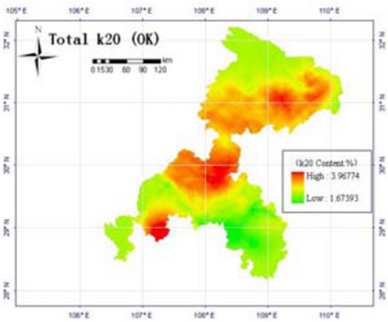
It was found that the highest determined coefficient(R^2) existed in Gaussian, Exponential, Exponential Model for total N,P, K_2O respectively. The whole spatial prediction map is displayed in Fig.4.



(a) Total N-OK



(b) Total P-OK



(c) Total K_2O-OK

Fig.4 Spatial distribution for total N, ,K_2O with ordinary kriging

4.5 Discussion

The MSEs and RMMSEs for total N,P, K_2O are given in table 6 for each method of spatial prediction.

Table 6 MSE and RMSSE between the measured and predicted values at the validation sites

Item		*CoK*	*RK*	*OK*
N	MSE	0.065	0.047	0.091
	RMSSE	0.872	0.911	0.811
P	MSE	0.071	0.082	0.102
	RMME	0.946	0.901	0.777
K_2O	MSE	0.034	0.027	0.061
	RMSE	0.911	0.973	0.897

For total N, the MSE for regressing kriging was the smallest, followed by that for cokriging, and last was the ordinary kriging. The RMSSE of being close to 1 is ranked by RK,CoK and OK. These results show that that some benefit aroused from using the more intensive terrain data to predict the sparser soil properties.Fig.2,3,4 shows the spatial prediction maps with different methods. Generally, the major patterns of variation are evident but the detail is different. The most variable map is the one from the OK prediction, and the smoothest is from regression kriging.

Table 6 shows that CoK was the most accurate method of prediction for total P, followed by regression kriging, and the ordinary kriging was the worst in this case. The less accurate predictions from regression kriging probably reflect regression model less good determined coefficient .

For total K_2O ,the prediction accuracy rank shows similar pattern with total N.

5. CONCLUSION

Understanding the spatial variation of soil properties to mange the field is one aim of precision agriculture. Increasing the accuracy of the spatial predictions of soil nutrient elements with the aid of available ancillary data is quiet economic and environmental typically in complex terrain hilly areas ,because soil nutrient elements are measured sparsely compared with ancillary data such as elevation, slope and aspect. This study predicted the three soil nutrient elements spatial distribution with CoK, RK and OK in Chongqing tobacco planting area, and the methods of incorporating ancillary terrain data both show good advantage.

With the intensive environmental data like DEM and remotely sensed data from satellites and ground-based systems become increasingly available, they are likely to confer benefit in the context of general environmental management where sampling to record the variable and limits the accuracy of predictions. However, up to now there is no single best method for all variables. The coregionalization and the relations between the deterministic components of the variation should still be examined carefully before deciding on the most appropriate method of prediction.

ACKNOWLEDGEMENTS

This paper was supported by the grants from National Science &Technology Pillar program in the Eleventh Five-year Plan Period (No.2006BAD10A01-02) and Chongqing Tobacco Corporation Development Foundation(NO.2006016).

REFERENCES

Auemhammer, H. (2001). "Precision farming-the environment challenge." Computers and Electronics in Agriculture(30): 31-43.

Cambardella, C. A. and T. B. Moorman, et al. (1994). "Field-scale variability of soil properties in central Iowa soils." Soil science society of American Journal(58): 1501-1511.

Chaplot, V. and C. Walter, et al. (2004). "Using the topography of the saprolite upper boundary to improve the spatial prediction of the soil hydromorphic index." Geoderma(123): 343-354.

Goovaerts, P. (1997). Geostatistics for natural resources evaluation, Oxford university press.

Kay, S. and D. Rainer (2008). "Prediction of soil property distribution in paddy soil landscape using terrain data and satellite information as indicators." Ecological indicators(8): 485-501.

Lopez-Granados, F. and M. Jurado-Exposito, et al. (2005). "Using geostatistical and remote sensing approaches for mapping soil properties." Eur.J.Agron(23): 279-289.

Lu, M. and Y. Yang (1993). "Study on the combined digestion of total N,P and K in soil." Acta Pedologica Sinica(3): 334-340.

McBratney, A. B. and I. O. A. Odeh, et al. (2000). "An overview of pedometric techniques for use in soil survey." Geoderma(97): 293-327.

Odeh, I. O. A. and A. B. McBratney, et al. (1995). "Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression kriging." Geoderma(67): 215-226.

Sylvester-Bradley and L. R, et al. (1999). "An analysis of the potential of precision farming in northern Europe." Soil use and management(15): 1-8.

Webster, R. (1985). "Quantitative spatial analysis of soil in the field." Advance in soil science(3): 2-16.

Wu, W. and Y. Fan, et al. (2008). "Assessing effects of digital elevation model resolutions on soil-landscape correlations in a hilly area." Agriculture, Ecosystems and Environment(126): 209-216.