EVALUATING COKRIGING FOR IMPROVING SOIL NUTRIENT SAMPLING EFFICIENCY

S. Han, S. M. Schneider, R. G. Evans

ABSTRACT. *The spatial variability of soil texture and soil nitrate–N, P, and K was studied in two center–pivot irrigated fields (89 ha total). Two soil texture components (clay and silt) were found to be correlated with soil nitrate–N, P, and K, and were used as auxiliary variables in the cokriging procedure to estimate soil nitrate–N, P, and K at unsampled locations. With a sampling density of 2.7 sites/ha (61 × 61 m grid) as the baseline, removal of 50% of the sampling sites resulted in a normalized mean absolute error (NMAE) of 34.5%, 22.9%, and 15.3% for soil nitrate–N, P, and K, respectively. These numbers reduced to 32.8%, 20.7%, and 12.0% when only 25% of the sampling sites were removed. The study showed that the cokriging technique provided slightly better estimates than the ordinary kriging method for soil P and K at a higher sampling density (>2.1 sites/ha). However, when a variable has a large random variation, such as the soil nitrate–N, cokriging did not provide better estimates than ordinary kriging. The results of this study provide guidelines on the selection of kriging or cokriging in improving the soil nutrient sampling efficiency.*

Keywords. *Cokriging, Kriging, Spatial variability, Variogram.*

Understanding the spatial distributions of soil properties, particularly soil nitrate–N, P, and K, within a field is important for the development of site–specific fertilizer management strategies. Because of limited resources, intensive soil sampling is impractical. Typically, soil samples are collected from only a small number of sites in the field. The data at the sampling locations are then used to predict values at unsampled locations and thus to generate the spatial distributions.

Spatial interpolation techniques are needed to study the spatial distributions of soil properties. If a soil property is autocorrelated in space, kriging methods, based on Matheron's regionalized variable theory (Matheron, 1963), can be utilized to provide the best linear unbiased estimates (BLUE). Many researchers have applied the kriging techniques to evaluate variability of soil physical and chemical properties at the field level (Burgess and Webster, 1980a, 1980b; Meirvenne and Hofman, 1989; Paz et al., 1996). Studies have shown that better precision of estimation can be achieved and smaller sampling size is needed when using kriging techniques than when using the conventional statistical approach (McBratney and Webster, 1983; Di et al., 1989; Chung et al., 1995).

Kriging depends on computing an accurate semivariogram model from which estimates of variance can be calculated. A sufficient number of samples are needed to develop an accurate semivariogram that can represent the autocorrelation of the soil property under consideration. On the other hand, soil properties may be interdependent. The property of interdependency suggests that we can estimate a property whose values are difficult to measure from other properties whose values are easier to determine. An extended technique of kriging, called cokriging, can be applied in this case. As examples, Yates and Warrick (1987) estimated the gravimetric moisture content using a cokriging procedure in which the bare soil surface temperature and the percent sand content were used as auxiliary variables. Vaughan et al. (1995) predicted soil salinity at unsampled points by cokriging soil–paste–measured electrical conductivity and the apparent electrical conductivity at the surface. If accurate cross–correlation functions can be determined, then cokriging can reduce the data requirement for the primary variable and improve the precision of estimation.

A number of studies have been done on applying the kriging and cokriging techniques to estimate soil nutrient distributions. As examples, Meirvenne and Hofman (1989) studied the spatial structure of soil nitrate–N in a 1 ha polder field and determined the number of samples required to estimate the mean nitrate–N content by using the kriging technique. Zhang et al. (1999) used cokriging with nonsymmetric pseudo–cross–variograms to estimate soil nitrate–N distributions at deeper layers by including soil nitrate–N sample data at shallower layers. They found that cokriging, compared with kriging, reduced the mean square error by 60% and used less than half the data for the estimation of nitrate–N distributions. The main objective of this study was to investigate the application of the cokriging technique for estimating soil nitrate–N, P, and K by using soil texture components as auxiliary variables. The potential application of this technique is to reduce the cost of soil nutrient analysis.

Article was submitted for review in January 2002; approved for publication by the Information & Electrical Technologies Division of ASAE in March 2003.

The authors are **Shufeng Han, ASAE Member Engineer,** Engineering Scientist, Deere and Company, Urbandale, Iowa; **Sally M. Schneider,** Research Plant Pathologist, USDA–ARS, Fresno, California; and **Robert G. Evans, ASAE Member Engineer,** Supervisory AgriculturalEngineer, USDA–ARS, Sidney, Montana. **Corresponding author:** Shufeng Han, Deere and Company, 4140 114th Street, Urbandale, IA 50322; phone: 515–331–4875; e–mail: hanshufeng@johndeere.com.

Materials and Methods

Two adjacent center–pivot irrigated fields (89 ha total) located in south central Washington State were selected for the study. The soil was classified as Quincy loamy sand (Xeric Torripsamments, mixed, mesic) with 2% to 15% slopes. Typical crops in the area include potato, wheat, and buckwheat in rotation. The cropping history and management practices of the two fields were the same for the last several years.

In March 1995, prior to the pre–planting fertilizer application, the experimental fields were sampled at 246 locations on a roughly 61 × 61 m grid. At each location, three borings were done within a 2 m radius. The borings were segmented into three depth increments (0 to 30 cm, 30 to 60 cm, and 60 to 90 cm), and the soil from all three borings at each depth increment was aggregated. A total of 738 soil samples was collected.

Soil samples were analyzed in a commercial laboratory for texture (sand, silt, clay), nitrate–N, P, K, S, and pH. Soil texture, nitrate–N, and pH were analyzed for each of the three soil layers, and soil P, K, and S were analyzed only for the top soil layer. Soil texture, nitrate–N, P, and K from the 0 to 30 cm soil layer were used in this study. Because of the close similarity in cropping history and management of the two fields, the data were not separated by individual field for the analysis.

A complete data set for each of the primary variables (nitrate–N, P, K) was partitioned into two subsets: a calibration data set, and a validation data set. The calibration data set was used in the cokriging analysis, and the validation data set was used to evaluate the accuracy of the cokriging procedure. Two soil texture components (clay and silt) were used as auxiliary variables. Another soil texture component (sand) was found to be less correlated with soil nutrients and was not used in the study. A normality test was performed for each data set. When appropriate, the data set was log–transformed to create a better normal distribution. Each data set was also normalized to a zero mean and unit variance. Finally, a geostatistical technique, cokriging, was applied to estimate values at the validation sites by using the calibration data set, with soil texture components as auxiliary variables. The equations defining the cokriging algorithm can be found in many references (Isaaks and Srivastava, 1989; Yates and Warrick, 1987; Journel and Huijbregts, 1978). The generalized cokriging estimator is defined as a linear combination of all the available sample data (Bogaert et al., 1995):

$$Z_{v0}^{k}(x_0) = \sum_{p_v=1}^{n_{p_v}} \sum_{v=1}^{n_v} \lambda_{p_v v} Z_v(x_{p_v}) \tag{1}$$

where

$Z_{v0}^{k}(x_0)$ = cokriged value of the primary variable at point x_0

$\lambda_{p_v v}$ = weight of the p_vth value of the vth regionalized random variable

$Z_v(x_{p_v})$ = vth variable at point x_{p_v}

n_{p_v} = number of known values of the vth variable

n_v = number of studied variables.

The weights, $\lambda_{p_v v}$ ($p_v = 1, ..., n_{p_v}$, $v = 1, ..., n_v$), are chosen in such a way that the estimation is not biased, and the variance of the estimation is minimized. The weights can be determined by solving a set of linear equations defined by the theoretical semivariogram and cross–semivariogram between any two variables.

Finally, the estimated values were compared to the actual sample values at the validation sites under different partitions of the original data sets.

The accuracy of the cokriging procedure was evaluated by two criteria: the normalized mean absolute error (NMAE), and the normalized root mean square error (NRMSE). The NMAE and NRMSE are defined by:

$$\text{NMAE} = 100\% \times \frac{\sum_{i=1}^{M} |\hat{z}_i - z_i|}{\sum_{i=1}^{M} z_i} \tag{2}$$

$$\text{NRMSE} = 100\% \times \frac{\sqrt{\frac{1}{M}\sum_{i=1}^{M} (\hat{z}_i - z_i)^2}}{\frac{1}{M}\sum_{i=1}^{M} z_i} \tag{3}$$

where

$\hat{z}_i$ = estimate of property Z at location i

z_i = true value of property Z at location i

i = 1, 2, ..., M

M = total number of points in the comparison.

Both the NMAE and NRMSE are measures of the global average deviation between the estimated values and the true values. The NRMSE is the most commonly used criterion, although the NMAE was mentioned to be more robust than the NRMSE (Journel, 1984).

Results and Discussion

The descriptive statistics for the complete data set are given in table 1 and include the total number of samples (n), the mean, the standard deviation (SD), the maximum, the minimum, and the coefficient of variation (CV). Soil nitrate–N and texture components displayed large CVs, indicating the large spatial variability for these soil properties. Both P and K exhibited a moderate spatial variability. To investigate the degree of linear correlation between different variables, Pearson correlation coefficients (table 2) were calculated by the CORR procedure of the SAS software (SAS, 1988). Nitrate–N, P, and K were all highly correlated with the two texture components ($P < 0.01$), with the exception of P with clay content ($P > 0.05$). The results justified the selection of the texture components as the auxiliary variables for estimating the nitrate–N, P, and K.

Table 1. Descriptive statistics for each complete data set.

Variable	n	Mean	SD[a]	Max.	Min.	Unit	CV[b]
Nitrate–N	246	2.7	1.84	23.0	0.8	mg/kg	66.8
P	246	24.2	6.44	68.0	7.0	mg/kg	26.7
K	246	184	40.7	355	100	mg/kg	22.1
Clay	243	1.6	0.93	4.8	0.0	%	58.0
Silt	243	12.3	5.80	32.0	1.0	%	47.2

[a] SD = standard deviation

[b] CV = coefficient of variation (%).

Table 2. Pearson correlation coefficients for each complete data set.[a]

	Nitrate–N	P	K	Clay	Silt
Nitrate–N	1.000				
P	0.098	1.000			
K	0.186	0.179	1.000		
Clay	0.222	0.113	0.205	1.000	
Silt	0.202	0.254	0.306	0.538	1.000

[a] Underlined coefficients are statistically significant at the 0.01 probability level.

Table 3. Normal distribution test for each complete data set.

	Before Transformation		After Log–Transformation		
Variable	W[a]	Normal Dist.	W[a]	Normal Dist.	P[b]
Nitrate–N	0.662	Rejected	0.962	Rejected	
P	0.943	Rejected	0.977	Not rejected	0.13
K	0.937	Rejected	0.978	Not rejected	0.17
Clay	0.960	Rejected	0.966	Rejected	
Silt	0.903	Rejected	0.976	Not rejected	0.09

[a] W = Shapiro–Wilk statistic (Shapiro and Wilk, 1965).
[b] P = significance level; normal distribution is not rejected at P > 0.05.

The Shapiro–Wilk statistic (W) was computed for each complete data set to test the normality of the distributions (Shapiro and Wilk, 1965). All variables were not normally distributed (table 3). Early studies showed that soil chemical properties are more closely log–normally distributed (Zhang et al., 1992; Bahri et al., 1993). The log–transformation of the data improved the normality of the distributions, as indicated by the increased W values in table 3. To eliminate the scale effect, all variables were normalized to a zero mean and unit variance after the log–transformation. When the normalized value was greater than 3, it was considered a probable outlier and was removed from the data set. One probable outlier was identified for each of the silt and nitrate–N data sets, and two probable outliers were identified for each of the P and K data sets.

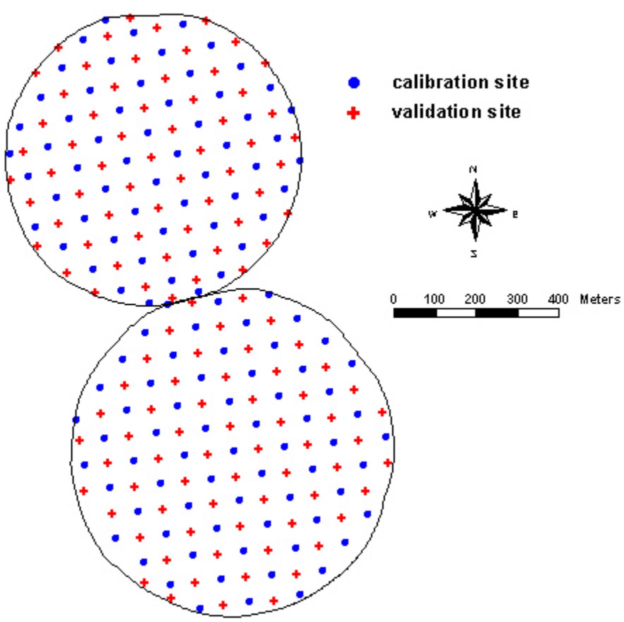

Figure 1. The 124 calibration sites and 122 validation sites. About 50% of the sampling sites are validation sites.

The 246 sampling sites were divided into two groups: one group of calibration sites, and another group of validation sites. About 50% of the total number of sites were chosen as validation sites (fig. 1). The validation sites were selected to have a systematical distribution across the field to minimize the average distance (McBratney et al., 1981).

A data detrending method, based on a polynomial regression of data against the coordinates, was applied to each data set prior to the calculation of the semivariances. By visual inspection of the semivariograms in four directions (0°N, 90°N, 180°N, and 270°N), no obvious anisotropic spatial dependence was found for the data sets. Therefore, only the isotropic semivariogram models were used in this study. In calculating the experimental semivariances, 17 distance (lag) classes were chosen to cover the range from 46 to 552 m. The distance spacing was not selected as a constant in order to ensure that there were about the same number of distance pairs within each class. The experimental semivariograms and cross–semivariograms are shown in figure 2. The experimental semivariances were then fitted to four basic models — a nugget model and three spherical models:

$$\gamma(h) = c_0 + c_1\left[\frac{3}{2}\left(\frac{h}{a_1}\right) - \frac{1}{2}\left(\frac{h}{a_1}\right)^3\right] + c_2\left[\frac{3}{2}\left(\frac{h}{a_2}\right) - \frac{1}{2}\left(\frac{h}{a_2}\right)^3\right] + c_3\left[\frac{3}{2}\left(\frac{h}{a_3}\right) - \frac{1}{2}\left(\frac{h}{a_3}\right)^3\right] \quad (4)$$

where c_0 is the nugget, and a_1, a_2, and a_3 are the ranges for each spherical model. The estimation of model coefficients was based on the linear model of coregionalization (Bogaert et al., 1995), and all the model coefficients are given in table 4. The fitted semivariogram and cross–semivariogram functions are shown in figure 2. All semivariogram models exhibited a strong nugget effect, indicating the large short–range variation.

A cokriging program (Bogaert et al., 1995) was used to estimate soil nitrate–N, P, K values at each validation site, which were compared with the sampling values at that site. Plots of the estimated data against the measurement data at the validation sites consistently showed under–estimation at several sites with high nitrate–N, P, K values. The normalized mean absolute error (NMAE) and the normalized root mean square error (NRMSE) for soil nitrate–N, P, K estimation are given in table 5.

The NMAE values were 34.5%, 22.9%, and 15.3% for soil nitrate–N, P, and K, respectively. Soil nitrate–N displayed the highest cokriging error, and soil K displayed the lowest error. This result is expected from the difference in CV values for these variables (table 1). In general, if a variable has a larger random variation, a higher cokriging error is anticipated.

To investigate possible improvement in cokriging accuracy, 62 more sampling sites were added to the previous calibration sites, which resulted in 75% validation sites (fig. 3). Based on the expanded data set, semivariogram and cross–semivariogram functions were recalculated, and the cokriging procedure was performed. The NMAE and the NRMSE for soil nitrate–N, P, K estimation are given in table 5. The results indicated a 4.9%, 9.6%, and 21.6% reduction in the estimation errors for soil nitrate–N, P, and K, respectively. It seems that the random variation in soil

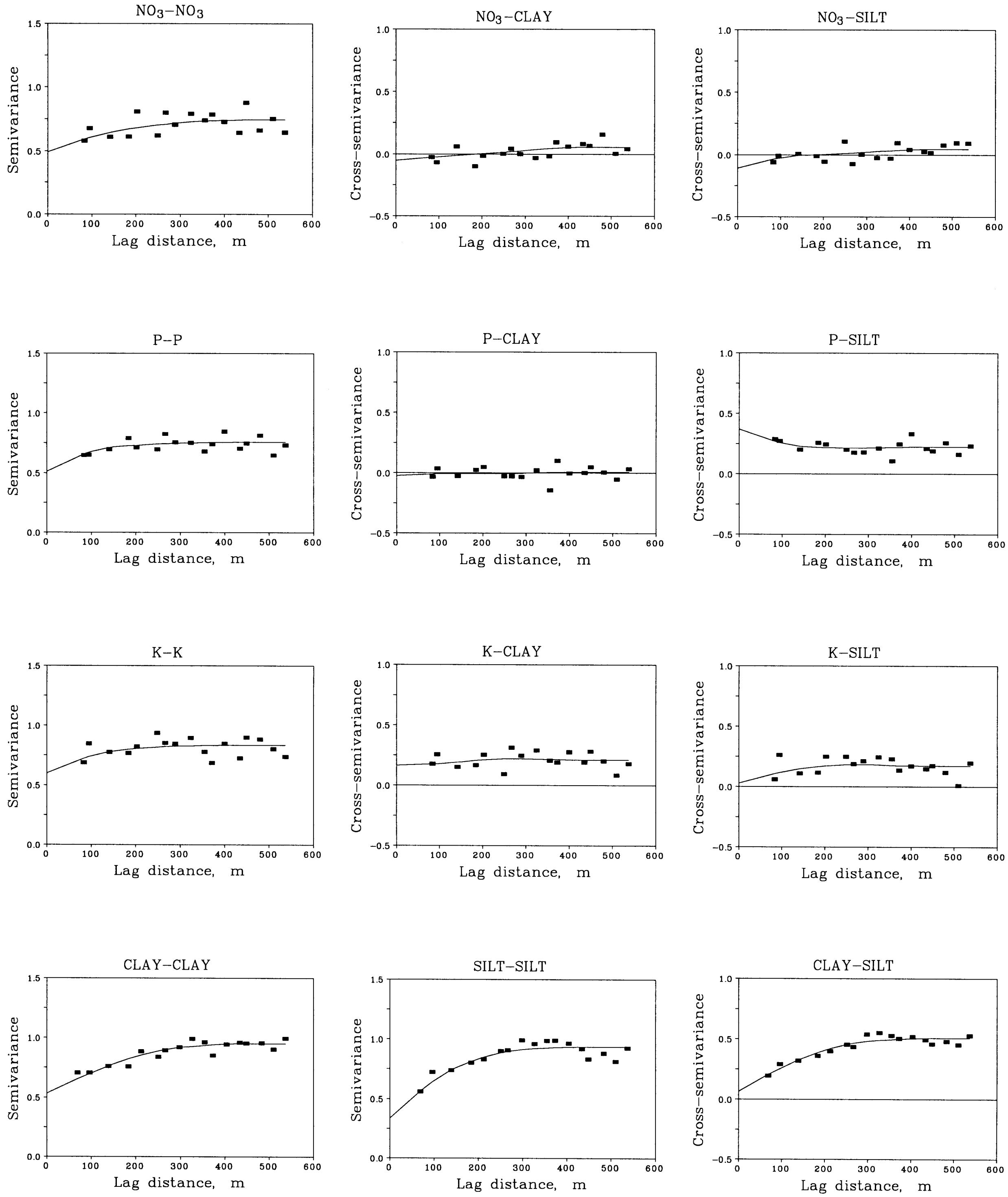

Figure 2. Semivariogram and cross–semivariogram functions for soil nitrate–N, P, K, clay content, and silt content, using 50% of the sampling sites.

nitrate–N dominates, and improvement in cokriging accuracy was impossible without using a higher sampling density.

As the final comparison with the cokriging accuracy, the ordinary kriging procedure was performed using the same data sets without soil texture components. The NMAE and the NRMSE are also shown in table 5. In general, ordinary kriging performed about the same as cokriging for the test data sets. However, cokriging showed some improvements over ordinary kriging with 75% of the sampling data (n = 186) for soil P and K.

Conclusions

Correlation between soil texture and soil nutrients seems to justify the selection of the texture components as auxiliary variables for estimating soil nitrate–N, P, and K using the cokriging technique. It is of practical importance since the analysis of soil nitrate–N, P, and K is more costly than that of soil texture. The results of this study suggest that the cokriging technique provides better estimates than the ordinary kriging method, if the cross–semivariogram func–

Table 4. Coefficients of the semivariogram and cross–semivariogram models, using 50% of the sampling sites.

	Nugget Model C_0	Spherical Model I (a_1 = 152.4 m) C_1	Spherical Model II (a_2 = 304.8 m) C_2	Spherical Model III (a_3 = 457.2 m) C_3
Nitrate–nitrate	0.491	0.057	0.044	0.151
Nitrate–clay	–0.053	0.016	–0.080	0.173
Nitrate–silt	–0.110	0.090	–0.078	0.140
P–P	0.510	0.158	0.028	0.060
P–clay	–0.025	0.033	–0.072	0.071
P–silt	0.369	–0.107	–0.089	0.050
K–K	0.596	0.114	0.074	0.048
K–clay	0.163	–0.045	0.153	–0.060
K–silt	0.027	0.040	0.172	–0.069
Clay–clay	0.534	0.004	0.213	0.197
Silt–silt	0.335	0.142	0.326	0.131
Clay–silt	0.065	0.025	0.254	0.160

Table 5. The normalized mean absolute error (NMAE) and the normalized root mean square error (NRMSE) for soil nitrate–N, P, K estimation.

	50% of Data Remaining (n = 124)		75% of Data Remaining (n = 186)	
Variable	NMAE	NRMSE	NMAE	NRMSE
Using cokriging				
Nitrate–N	34.5	49.9	32.8	43.9
P	22.9	28.8	20.7	26.4
K	15.3	21.4	12.0	19.0
Using ordinary kriging				
Nitrate–N	33.6	48.3	32.5	44.3
P	21.2	26.8	21.5	27.1
K	15.4	21.5	13.1	19.6

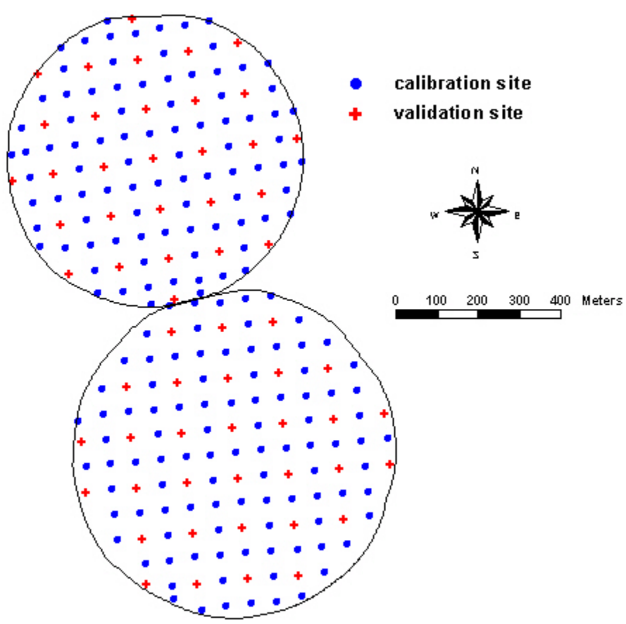

Figure 3. The 186 calibration sites and 60 validation sites. About 75% of the sampling sites are validation sites.

tions can be accurately defined with a sufficient amount of data. At a sampling density of 2.1 sites/ha (corresponding to 186 sites in the study), the NMAE values were reduced by 9.6% and 21.6% for soil P and K, respectively, as compared to the ordinary kriging method. A better improvement in the estimation accuracy is expected when a higher sampling density is used. However, the cokriging technique did not improve the estimation accuracy at the lower sampling density (1.4 sites/ha in the study). When a variable has a large random variation, such as the soil nitrate–N in the study, cokriging does not provide better estimates than ordinary kriging.

References

Bahri, A., R. Berndtsson, and K. Jinno. 1993. Spatial dependence of geochemical elements in a semiarid agricultural field: I. Scale properties. *SSSA J*. 57(5): 1316–1322.

Bogaert, P., P. Mahau, and F. Beckers. 1995. The spatial interpolation of agro–climatic data. FAO Agrometeorology Series Working Paper No. 12. Rome, Italy: FAO.

Burgess, T. M., and R. Webster. 1980a. Optimal interpolation and isarithmic mapping of soil properties: I. The semi–variogram and punctual kriging. *J. Soil Science* 31(2): 315–331.

_____. 1980b. Optimal interpolation and isarithmic mapping of soil properties: II. Block kriging. *J. Soil Science* 31(2): 333–341.

Chung, C. K., S. K. Chong, and E. C. Varsa. 1995. Sampling strategies for fertility on a Stoy silt loam soil. *Comm. Soil Science and Plant Analysis* 26(5/6): 741–763.

Di, H. J., B. B. Trangmar, and R. A. Kemp. 1989. Use of geostatistics in designing sampling strategies for soil survey. *SSSA J*. 53(4): 1163–1167.

Isaaks, E. H., and R. M. Srivastava. 1989. *An Introduction to Applied Geostatistics*. New York, N.Y.: Oxford University Press.

Journel, A. G. 1984. MAD and conditional quantiles estimators. In *Geostatistics for Natural Resource Characterization*, 261–270. G. Verly et al., eds. NATO ASI Series C, v. 122. Dordrecht, The Netherlands: Reidel Publishing.

Journel, A. G., and C. J. Huijbregts. 1978. *Mining Geostatistics*. New York, N.Y.: Academic Press.

Matheron, G. 1963. Principles of geostatistics. *Econ. Geol.* 58: 1246–1266.

McBratney, A. B., and R. Webster. 1983. How many observations are needed for regional estimation of soil properties? *Soil Science* 135(3): 177–183.

McBratney, A. B., R. Webster, and T. M. Burgess. 1981. The design of optimal sampling schemes for local estimation and mapping of regionalized variables: I. Theory and methods. *Comput. Geosci.* 7: 331–334.

Meirvenne, M. V., and G. Hofman. 1989. Spatial variability of soil nitrate–nitrogen after potatoes and its change during winter. *Plant and Soil* 120: 103–110.

Paz, A., M. T. Taboada, and M. J. Gomez. 1996. Spatial variability in topsoil micronutrients in a one–hectare cropland plot. *Comm. Soil Science and Plant Analysis* 27(3/4): 479–503.

SAS. 1988. *SAS Procedure Guide*. Release 6.03 ed. Cary, N.C.: SAS Institute, Inc.

Shapiro, S. S., and M. B. Wilk. 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52: 591–611.

Vaughan, P. J., S. M. Lesch, D. L. Corwin, and D. G. Cone. 1995. Water content effect on soil salinity prediction: A geostatistical study using cokriging. *SSSA J*. 59(4): 1146–1156.

Yates, S. R., and A. W. Warrick. 1987. Estimating soil water content using cokriging. *SSSA J*. 51(1): 23–30.

Zhang, R., D. E. Myers, and A. W. Warrick. 1992. Estimation of the spatial distribution of soil chemicals using pseudo–cross–variograms. *SSSA J*. 56(5): 1444–1452.

Zhang, R., P. Shouse, and S. Yates. 1999. Estimations of soil nitrate distribution using cokriging with pseudo–cross–variograms. *J. Environ. Qual.* 28(2): 424–428.