Original Article | Journal of Biosystems Engineering

J. of Biosystems Eng. 39(4):377-388. (2014. 12)
http://dx.doi.org/10.5307/JBE.2014.39.4.377

eISSN : 2234-1862
pISSN : 1738-1266

Spatial Variability of Soil Properties using Nested Variograms at Multiple Scales

Sun-Ok Chung[1]*, Kenneth A. Sudduth[2], Scott T. Drummond[2], Newell R. Kitchen[2]

[1]*Dept. of Biosystems Machinery Engineering, Chungnam National University, Daejeon, Republic of Korea*
[2]*USDA-ARS Cropping Systems and Water Quality Research Unit, Columbia, Missouri, USA*

Received: August 5th, 2014; Revised: August 27th, 2014; Accepted: November 9th, 2014

Abstract

Purpose: Determining the spatial structure of data is important in understanding within-field variability for site-specific crop management. An understanding of the spatial structures present in the data may help illuminate interrelationships that are important in subsequent explanatory analyses, especially when site variables are correlated or are a combined response to multiple causative factors. **Methods:** In this study, correlation, principal component analysis, and single and nested variogram models were applied to soil electrical conductivity and chemical property data of two fields in central Missouri, USA. **Results:** Some variables that were highly correlated, or were strongly expressed in the same principal component, exhibited similar spatial ranges when fitted with a single variogram model. However, single variogram results were dependent on the active lag distance used, with short distances (30 m) required to fit short-range variability. Longer active lag distances only revealed long-range spatial components. Nested models generally yielded a better fit than single models for sensor-based conductivity data, where multiple scales of spatial structure were apparent. Gaussian-spherical nested models fit well to the data at both short (30 m) and long (300 m) active lag distances, generally capturing both short-range and long-range spatial components. As soil conductivity relates strongly to profile texture, we hypothesize that the short-range components may relate to the scale of erosion processes, while the long-range components are indicative of the scale of landscape morphology. **Conclusion:** In this study, we investigated the effect of changing active lag distance on the calculation of the range parameter. Future work investigating scale effects on other variogram parameters, including nugget and sill variances, may lead to better model selection and interpretation. Once this is achieved, separation of nested spatial components by factorial kriging may help to better define the correlations existing between spatial datasets.

Keywords: Nested variogram, Soil property, Spatial variability

Introduction

Precision agriculture addresses within-field variability and optimizes inputs spatially within fields to improve profitability and to reduce possible environmental damage. For successful implementation of precision agriculture, it is important to assess both (1) the spatial variability of variables such as soil properties, and (2) the relationships between these measured variables. The spatial structure and scale of variability in these properties determines the required spatial intensity of data collection, either by manual sample collection or by on-the-go sensors. Both the spatial variability and the relationships between variables are important considerations in data analyses and interpretation, and in the delineation of management zones.

Correlation analysis has often been used as a first step to understand the relationships between variables measured for precision agriculture (e.g., Chung et al., 2005). Since it is often the case that the measured spatial variables are not statistically independent, principal component analysis (PCA) is used to create a new, uncorrelated set of variables. This new dataset, the principal components (PCs), consists of linear combinations of the original variables and reflects

***Corresponding author:** Sun-Ok Chung
Tel: +82-42-821-6712; **Fax:** +82-42-823-6246
E-mail: sochung@cnu.ac.kr

the main features of the original multivariate information (Goovaerts, 1992). Standardized PCA, in which the correlation matrix is used, is superior to unstandardized PCA, as it forces each original variable to have equal weight in the derivation of the new components (Jensen, 1996).

Geostatistics, based on the theory of regionalized variables, is the primary tool of spatial variability analysis and has been used extensively to analyze precision agriculture data (e.g., Bakhsh et al., 2000; Borgelt et al., 1997; Chung et al., 2008; Webster and Nortcliff 1984). Unlike conventional statistics, geostatistics can be used to analyze spatially dependent data and to provide information on the ranges over which variables are spatially dependent. The most common geostatistical model, the single variogram, may not adequately fit field data in those cases where multiple spatial structures or scales are present due to interactions of variables and processes. In many cases, important short-range spatial variability structures are nested in long-range trends (e.g., Bakhsh et al., 2000; Oliver et al., 2000). To understand spatial structure of the data, the multiple variability structures need to be properly modeled. In such cases, a nested variogram, which is the linear combination of single variograms, allows the capture of more complex patterns of variability.

The nested variogram approach has been used in various applications including image filtering in remote sensing (Oliver et al., 2000; Rodgers and Oliver, 2007), determination of the optimum field-of-view for an optical sensor (Tsunemori et al., 2000), and investigation of the spatial structure of soil properties (Goovaerts and Webster, 1994; McBratney and Webster, 1986; Webster and Nortcliff, 1984). In spatial analysis of soil fertility for site-specific crop management, Cahn et al. (1994) found that soil organic carbon had a small-scale spatial variation nested within large-scale variation and used median polishing detrending to remove the nonstationarity effects. Van Meirvenne et al. (2003) used a nested model to express a 3-D variogram of soil nitrate-nitrogen, and Batista et al. (2002) used nested models of principal components of soil data. In general, if a response variable (e.g., soil electrical conductivity) contains multiple spatial structures due to causative predictor variables (e.g., soil erosion, landscape morphology), more information on the relationships among variables can be obtained by decomposing the spatial structure of the response variable into multiple components using nested variograms and kriging for subsequent analysis. In fact, Goovaerts (1998) stated that a nested structure was typically required to fit the shape of experimental variograms of soil properties.

In previous work, we used linear and non-parametric correlation, nonlinear regression, and neural network analysis to help explain relationships between soil and site variables and crop yield (Chung et al., 2005; Drummond et al., 2003; Sudduth et al., 1997). We used similar techniques to understand how apparent bulk soil electrical conductivity (EC_a) relates to point-measured soil properties (Sudduth et al., 2003). Similar to the analyses carried out by others, our work did not consider the spatial structure of the measured data. We hypothesized that additional information on the relationships between variables could be obtained by including the spatial component; for example, soil properties with similar ranges of spatial dependency might be related. Further, we assumed that dense, sensor-based measurements such as soil EC_a might exhibit multiple ranges, with each range corresponding to some particular causative agent either related to the measurement process or to other soil and site variables.

The objectives of this study were to (1) apply single and nested variograms to soil property data from multiple fields, and (2) use the information obtained to enhance our knowledge of the relationships between the variables when integrated with information from other statistical analyses such as correlation and PCA.

Materials and Methods

Data collection and processing

Data were collected on two fields (Field 1, 35 ha and Field 2, 13 ha) located within 3 km of each other near Centralia, in central Missouri, USA. The soils found at these sites are characterized as claypan soils of the Mexico-Putnam association (fine, smectitic, mesic Aeric Vertic Aqualfs; Stagnic Luvisols). Mexico-Putnam soils formed in moderately fine textured loess over a fine textured pedisediment. Surface textures range from silty loam to silty clay loam. The subsoil claypan horizon(s) are silty clay loam, silty clay, or clay, and commonly contain as much as 50 to 60% montmorillonitic clay. Within each study field, topsoil depth above the claypan ranged from less than 10 cm to greater than 100 cm, generally as a function of landscape position. Topsoil depth was generally intermediate at summit positions, low on eroded backslopes, and high on depositional footslope positions. Topsoil depth variations were also seen where rills and gullies caused

by past erosion had been refilled with topsoil due to tillage operations. Because of the high-clay subsurface horizons, topsoil depth above the claypan is often correlated to crop productivity (Kitchen et al., 1999).

Data used in this analysis included EC_a and soil chemical properties. Fields were grid soil-sampled to a 15-cm depth and analyzed for P, K, Ca, Mg, cation exchange capacity (CEC), organic matter, salt pH, and neutralizable acidity (NA) using standard University of Missouri procedures (Brown and Rodriguez, 1983). Grid spacing was 33 m for Field 1 and 25 m for Field 2. Additional small-scale soil samples were obtained as described by Birrell et al. (1996) to provide a more robust spatial analysis at short lag distances.

Soil EC_a was measured for each field in the fall of 1999 using two commercial sensor systems: the Geonics EM38[1)] (Geonics Limited, Mississauga, Ontario, Canada) and the Veris 3100 (Veris Technologies Inc., Salina, Kansas, USA). The EM38 operates on the principle of electromagnetic induction and, as operated in the vertical dipole mode, provides an effective measurement depth of approximately 1.5 m (EC_{em}). The EM38 was used with a GPS-enabled mobile system described by Sudduth et al. (2001) to collect data every 1 s on measurement transects spaced 10 m apart. The Veris 3100 is a complete commercial system that measures EC_a through coulter electrodes that penetrate the ground surface. This device provides effective measurement depths of approximately 0.3 m (EC_{sh}) and 1.0 m (EC_{dp}). It also collected data every 1 s on a 10-m transect spacing. At the operating speeds used, this time interval corresponded to a 4–6 m spacing between sample points for both EC_a systems. Lower-speed operation at some within-field locations and overlapped transects during turning at field edges provided data at sampling intervals less than 0.2 m, the smallest lag interval used in the variogram analyses. In previous research, we have found these two EC_a sensors to provide similar, but not identical, mapped information on claypan-soil fields (Sudduth et al., 2003). EM38 and Veris deep EC_a data have been reliable estimators of claypan-soil topsoil depth (Doolittle et al., 1994; Kitchen et al., 1999; Sudduth et al., 2001; Sudduth et al., 2003).

Analytical procedures

Analytical procedures included descriptive statistics, correlation, PCA, and geostatistical analysis. All analyses were completed using S-Plus 4.0 and S+SpatialStats 1.5 (MathSoft, Seattle, Washington, USA). Descriptive statistics and correlation coefficients were calculated to provide an overall evaluation of the relationships between variables. Standardized principal component analysis was applied to soil chemical property data to create new, uncorrelated variables that could be compared with subsequent geostatistical analysis. Because PCA groups the most highly correlated variables into each principal component (PC), we expected that those variables which grouped into a single PC might also exhibit similar ranges of spatial dependency.

Experimental semivariances were calculated for each soil variable using equation (1) (Webster and Oliver, 1990).

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \{z_i - z_{i+h}\}^2 \quad (1)$$

where, h is lag distance, g (h) is semivariance for interval distance class h, z_i is measured sample value at point i, z_{i+h} is measured sample value at point i+h, and N (h) is total number of pairs for the lag interval h.

Single variogram models were fitted to semivariance values of all data to compare ranges between variables and fields. Range in a variogram model is the spacing between samples, beyond which the samples are no longer spatially correlated. A large range, greater than the active lag distance (the maximum spacing between data points used in the analysis, a distance constrained by field dimensions) means that the variable continues to exhibit spatial dependency past the maximum distance. This indicates that a greater active lag distance should be used if possible to capture the long-range variation. A range near zero means that the variation is at a scale smaller than could be analyzed. In case it is important to analyze short-range variability, then sampling interval and/or the lag interval used in variogram analysis should be decreased. In general, the active lag distance and lag interval need to be selected through preliminary variogram analysis considering variability patterns, field geometry, data sampling interval, and number of data points within the lag intervals (Cahn et al., 1994; Chung et al., 2008).

Single variogram models were applied to each dataset with different active lag distances (i.e., 30, 50, 100, and 200 m) to investigate the effect of changing active lag on

1) Mention of trade names or commercial products is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the authors or their organizations.

the range calculation. A lag interval of 0.2 m was used for EC_a data, while a 2-m interval was used for the soil chemical property data, to insure a sufficient number of pairs in each interval. Gaussian, spherical, and exponential models were fitted to the data. For the exponential model, the range of an "equivalent" spherical model, calculated as 3 times the exponential range, was used to allow comparison between models (McBratney and Pringle, 1999). The same concept was applied to the Gaussian model, where the equivalent spherical model range was 1.73 times the Gaussian range.

Based on the likelihood that the spatial structure of a variable might not be completely modeled by a single variogram model, nested variogram models were also applied to the sensor-based soil EC_a data. A nested variogram is simply a linear combination of single variogram models, as in equation (2).

$$\gamma(h) = \gamma_1(h) + \gamma_2(h) + \cdots + \gamma_n(h) \quad (2)$$

To understand the robustness of both single and nested variogram ranges, two series of analyses using different active lag distances were completed. The first group of analyses consisted of varying the active lag interval for sensor-based EC_a measurement from 10 m to 30 m on a 2-m increment. In the second group of analyses, the active lag was varied from 20 m to 300 m on a 20-m increment. Variogram model parameters are affected by several factors such as active lag distance and lag interval, and the factors are generally determined by considering spatial data sampling interval (i.e., frequency). After preliminary exploration of the experimental variogram, the 30-m and 300-m active lag distances were selected for demonstration of scale-dependent nested variogram modeling. At each active lag, the calculated range (or ranges for a nested model) was tabulated. Several nested models were used based upon combinations of single models from the literature: Gaussian-spherical, Gaussian-exponential, double spherical, triple spherical, and Gaussian-spherical-exponential (Journel and Huijbregts, 1978; McBratney and Webster, 1986; Webster and Nortcliff, 1984). Nested variogram models were fit using the S+SpatialStats function "variogram.fit" which optimized parameters of the multiple models simultaneously using nonlinear least squares (MathSoft, 2000).

Results and Discussion

Descriptive statistics, correlation, and principal component analysis

Descriptive statistics of the variables and simple correlations between variables are summarized in Tables 1 and 2,

Table 1. Descriptive statistics of the data used in the study

	EC_{em} [a)] (mS/m)	EC_{sh} [a)] (mS/m)	EC_{dp} [a)] (mS/m)	pH	NA (cmol/kg)	OM (%)	P (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	CEC (cmol/kg)
Field 1											
Min	15.9	1.6	2.3	4.6	0.0	1.5	5	805	61	68	8
Max	77.7	34.7	53.8	7.3	7.0	3.8	114	3078	438	318	22
Mean	30.7	9.7	19.6	6.0	2.3	2.3	15	1056	198	103	12
C.V. %	12	32	43	10	65	13	72	29	32	28	19
Skewness	0.8	1.4	0.6	0.1	0.5	0.5	4.3	1	2	3	1.7
Kurtosis	4.2	4.1	0.1	-0.9	-0.3	0.8	26.3	2	5	14	3.8
n	10289	-------- 10995 --------		---------- 508 ----------							
Field 2											
Min	0.0	1.5	2.6	5.7	0.0	1.9	3	1478	72	92	6
Max	60.3	62.3	77.0	7.4	3.0	3.5	64	6650	683	277	39
Mean	34.8	15.2	23.7	6.9	0.4	2.6	21	2968	216	199	17
C.V. %	18	49	49	6	175	12	53	28	57	18	29
Skewness	0.1	1.8	0.5	1.1	1.9	0.3	1.4	2	2	1	1.5
Kurtosis	0.6	4.1	-0.1	0.7	2.9	-0.5	2.3	3	2	1	2.5
n	4372	---------- 5107 ----------		---------- 276 ----------							

[a)] EC_{em}, EC_{sh}, EC_{dp} are EM38, Veris shallow (0–0.3m) and Veris deep (0–1.0m) soil electrical conductivity measurements.

Table 2. Pearson correlation coefficients (r) between variables

	EC_{em}[a)] (mS/m)	EC_{sh}[a)] (mS/m)	EC_{dp}[a)] (mS/m)	pH	NA (cmol/kg)	OM (%)	P (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	CEC (cmol/kg)
Field 1											
EC_{em}											
EC_{sh}	**0.61**[a)]										
EC_{dp}	**0.69**	**0.78**									
pH	-0.08	-0.07	-0.23								
NA	0.19	0.19	0.31	**-0.95**							
OM	0.22	0.26	0.17	-0.25	0.34						
P	0.02	-0.18	-0.25	0.30	-0.23	0.27					
Ca	0.27	**0.51**	0.24	**0.51**	-0.37	0.37	0.28				
Mg	**0.52**	**0.63**	**0.53**	-0.35	0.47	**0.61**	0.02	0.44			
K	0.40	0.31	0.29	-0.13	0.26	**0.62**	**0.65**	0.42	**0.57**		
CEC	0.45	**0.64**	**0.51**	-0.40	**0.56**	**0.66**	0.06	**0.55**	**0.89**	**0.64**	
PC 1	0.44	**0.52**	0.44	-0.47	**0.61**	**0.80**	0.23	0.42	**0.88**	**0.78**	**0.94**
PC 2	0.07	0.09	-0.11	**0.85**	**-0.76**	0.15	**0.63**	**0.76**	-0.00	0.36	0.01
Field 2											
EC_{em}											
EC_{sh}	**0.74**										
EC_{dp}	**0.89**	**0.80**									
pH	0.22	0.22	0.12								
NA	-0.12	-0.04	-0.01	**-0.91**							
OM	0.32	0.57	0.37	-0.02	0.15						
P	-0.17	-0.36	-0.28	0.21	-0.30	-0.35					
Ca	**0.54**	**0.74**	**0.56**	0.40	-0.24	**0.63**	-0.23				
Mg	**0.61**	**0.79**	**0.67**	0.10	0.10	**0.71**	-0.38	**0.87**			
K	**0.54**	**0.72**	**0.58**	0.21	-0.07	**0.68**	-0.05	**0.80**	**0.82**		
CEC	**0.56**	**0.78**	**0.60**	0.23	-0.05	**0.69**	-0.30	**0.98**	**0.94**	**0.83**	
PC 1	**0.57**	**0.80**	**0.62**	0.27	-0.09	**0.79**	-0.33	**0.95**	**0.95**	**0.88**	**0.98**
PC 2	-0.08	0.03	0.05	**-0.92**	**0.95**	0.28	**-0.51**	-0.17	0.15	-0.04	0.01

[a)]Correlations >0.5 shown in bold.

respectively. A wide range in the amount of variation was observed for the different variables (Table 1). Soil EC_a was more variable for Field 2, while most soil chemical properties were more variable for Field 1. Soil EC_a measured with the Veris 3100 was more variable than EC_a measured with the EM38. Correlation analysis results showed that some variables had relatively higher correlation than other variables. The three types of soil EC_a measurements were relatively highly correlated ($r > 0.6$, shown in bold in Table 2) within each field. Relatively high correlations were seen between Ca, Mg, K, CEC, and OM for Field 2, although this pattern was not as strong for Field 1.

Table 3 shows the contribution of the soil fertility variables to the first and second PCs. These two PCs contained 77% of the total variation in the soil fertility variables for Field 1 and 81% of the variation found in Field 2. For both fields, OM, Mg, K and CEC showed relatively strong influence (loadings >0.35, shown in bold in Table 3) on the structure of PC 1. For both fields, pH, NA, and P showed relatively strong influence on PC 2. Only Ca exhibited differing behavior between the fields, most strongly influencing PC 2 for Field 1 and PC 1 for Field 2. For Field 2, PC 1 showed relatively greater correlation to EC_a. This PC included cations (Ca, Mg, K) as well as CEC and organic matter that are all variables expected to be correlated to EC_a. Similar behavior was seen for PC 1 on

Field 1, although correlations were not as strong (Table 2).

Single variogram models

First, each experimental variogram was examined visually. Variograms for soil EC_a exhibited potential multiple structures (Figure 1), while soil chemical properties showed a single structure (Figure 2). Chemical properties could be modeled appropriately with a single spherical or exponential model, possibly due to a relatively large amount of scatter in the data, compared to the more dense sensor-based measurements of soil EC_a. All three EC_a variables exhibited similar overall patterns in the variograms, with apparent multiple structures for both fields at different active lag distances (data not shown).

Next, single variogram models were fit to the measured data and range values were calculated using Gaussian, spherical, and exponential models. Generally, the Gaussian model fit the data better than spherical and exponential models, especially at smaller active lag distances, because most of the soil properties exhibited a parabolic trend in semivariance near the origin (e.g., Figure 1). Because of the superior fit, only the Gaussian model results are presented (Table 4). In a number of cases, similar ranges were found for related variables. For example, EC_{sh} and EC_{dp} ranges were similar for both 50-m and 100-m active lags on Field 1, and for 30-m and 50-m active lags on Field 2. As expected, NA and pH had similar ranges of app-

Table 3. Loadings of soil properties in principal components (PC) 1 and 2

Soil Properties	Field 1		Field 2	
	PC 1 (47%)[a)]	PC 2 (30%)	PC 1 (54%)	PC 2 (27%)
pH	-0.24	**0.54**[b)]	0.13	**-0.63**
NA	0.32	**-0.49**		**0.65**
OM	**0.41**		**0.38**	0.20
P	0.12	**0.41**	-0.16	**-0.35**
Ca	0.22	**0.49**	**0.46**	-0.12
Mg	**0.46**		**0.46**	0.10
K	**0.41**	0.23	**0.42**	
CEC	**0.49**		**0.47**	

[a)]Numbers in parentheses are the portion of the total variance explained by each PC.
[b)]PC loadings > 0.35 shown in bold.

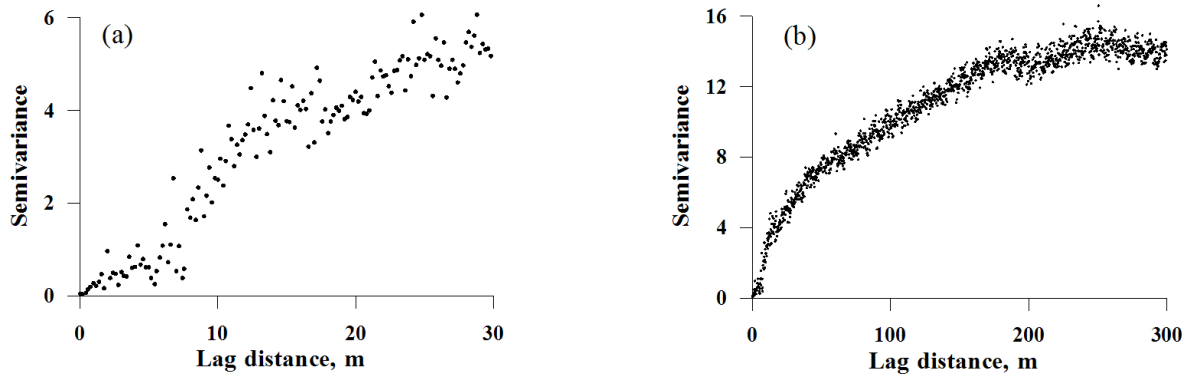

Figure 1. Variograms of EC_{em} for Field 1 with active lags of 30 m (a) and 300 m (b).

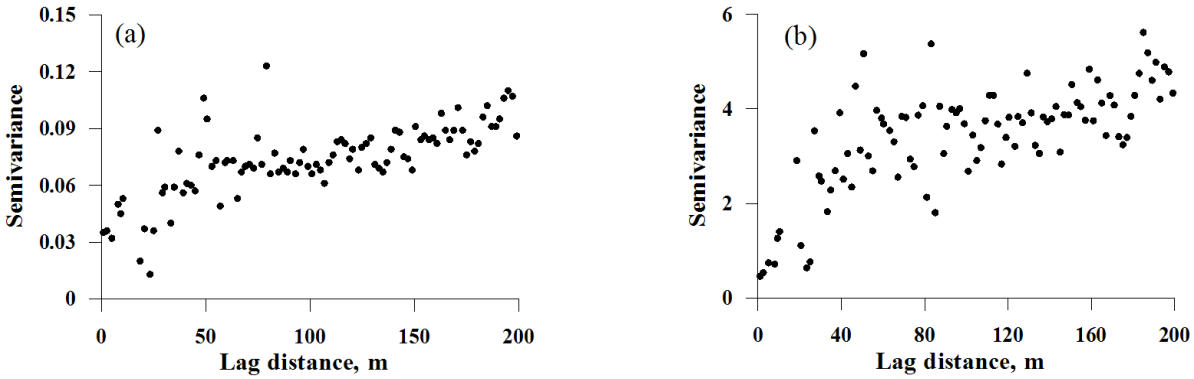

Figure 2. Variograms of OM (a) and CEC (b) for Field 1.

Table 4. Gaussian model ranges (m) for different active lag distances

Data	Field 1				Field 2			
	30 m	50 m	100 m	200 m	30 m	50 m	100 m	200 m
EC_{em}	**21.7**[a)]	**28.4**	75.2	179.9	33.8	44.5	0.0	171.6
EC_{sh}	**21.7**	47.9	**57.2**	**89.5**	24.7	42.2	90.3	352.9
EC_{dp}	30.6	43.8	**60.5**	151.3	27.4	36.8	**73.0**	253.3
pH	8.1[b)]	13.2	16.8	21.6	40.2	35.6	53.7	250.1
NA	10.5	16.0	20.0	142.8	396.3	36.9	42.6	47.1
OM	**19.3**	21.3	**59.8**	523.1	**15.3**	47.9	**66.6**	**293.7**
P	1285.2	1135.0	0.0	47.2	**10.8**	12.8	185.0	0.0
Ca	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0
Mg	0.0	38.2	**49.5**	0.0	0.0	52.8	87987.2	211.3
K	0.0	0.0	**45.6**	0.0	0.0	42.8	0.0	16124.5
CEC	**22.2**	**34.4**	**53.5**	64.0	**9.4**	11.2	13.0	**312.3**
PC1	**23.2**	**34.0**	**56.5**	**78.1**	**14.4**	23.7	**71.6**	**296.9**
PC2	10.5	12.1	15.8	19.3	30.2	37.9	52.7	63.2

[a)]Within a column, relatively similar range values for PC1 and soil properties are shown in bold.
[b)]Within a column, relatively similar range values for PC2 and soil properties are shown underlined.

roximately 10–20 m for Field 1 and 40–80 m for Field 2 with active lag distances of 100 m or less. Gaussian variograms were also fit to PC 1 and PC 2. In several cases, ranges of PC 1 were relatively more similar to those of EC_a, OM, and CEC (bold values in Table 4), while ranges of PC 2 were more similar to ranges found for pH and NA (underlined values in Table 4). Some, but not all variables, which were highly correlated, or which were captured by the same principal component, exhibited similar spatial ranges when fit with a single variogram model. For example, PC 1 and EC_a had similar ranges at multiple active lags for Field 1. Soil pH and NA ranges were similar to the range for PC 2 over a number of active lags for both fields (Table 4).

In some cases, range had a larger value than the active lag distance (e.g., OM), was zero or near zero (e.g., Ca, Mg, K), or fluctuated (e.g., P) as the active lag distance increased (Table 4). Ranges larger than the active lag indicate that the semivariance keeps increasing with the increase in lag and suggest the presence of a longer-range variability structure that may not be captured properly with a single model. Zero or near zero ranges indicate that the variance appears as a pure nugget with the particular model, lag interval, and active lag distance. Fluctuation in the range with different active lags indicates that spatial structure apparent at one scale (or active lag distance) may merely be regarded as noise at other scales. In all these cases, a single variogram model at a given active lag has not captured the complete structure of the experimental variogram and, because of this, does not give reliable estimates of variogram parameters. For instance, the Field 1 EC_{em} semivariance showed a distinct parabolic shape near the origin when the active lag distance was restricted to 30 m (Figure 1(a)). A Gaussian model with a short range would best model this parabolic shape. However, if a 300-m active lag was used, the pattern near the origin became insignificant compared to the overall variogram (Figure 1(b)). Although this loss of detail may not be important if the purpose of the semivariance calculation is spatial interpolation, it may be important if the purpose is to understand the underlying spatial structure of the data.

The Field 1 EC_{em} data (Figure 1) appeared to include multiple spatial scales. The variogram with a 30-m active lag (Figure 1(a)) showed the possible presence of three spatial structures at scales: (1) under 10 m, (2) between 10–20 m, and (3) larger than 20 m. The variogram with a 300-m active lag (Figure 1(b)) indicated possible components at other scales: (1) under 25 m, (2) between 25–200 m, and (3) greater than 200 m.

Nested variogram models

Nested variogram models were fit to the sensor-derived EC_a data set, and the range parameter and coefficient of determination (R^2) of each model was plotted versus active lag distance. Figure 3 shows a series of such plots for the EC_{em} data. Figure 3(a) and (b) show how range and

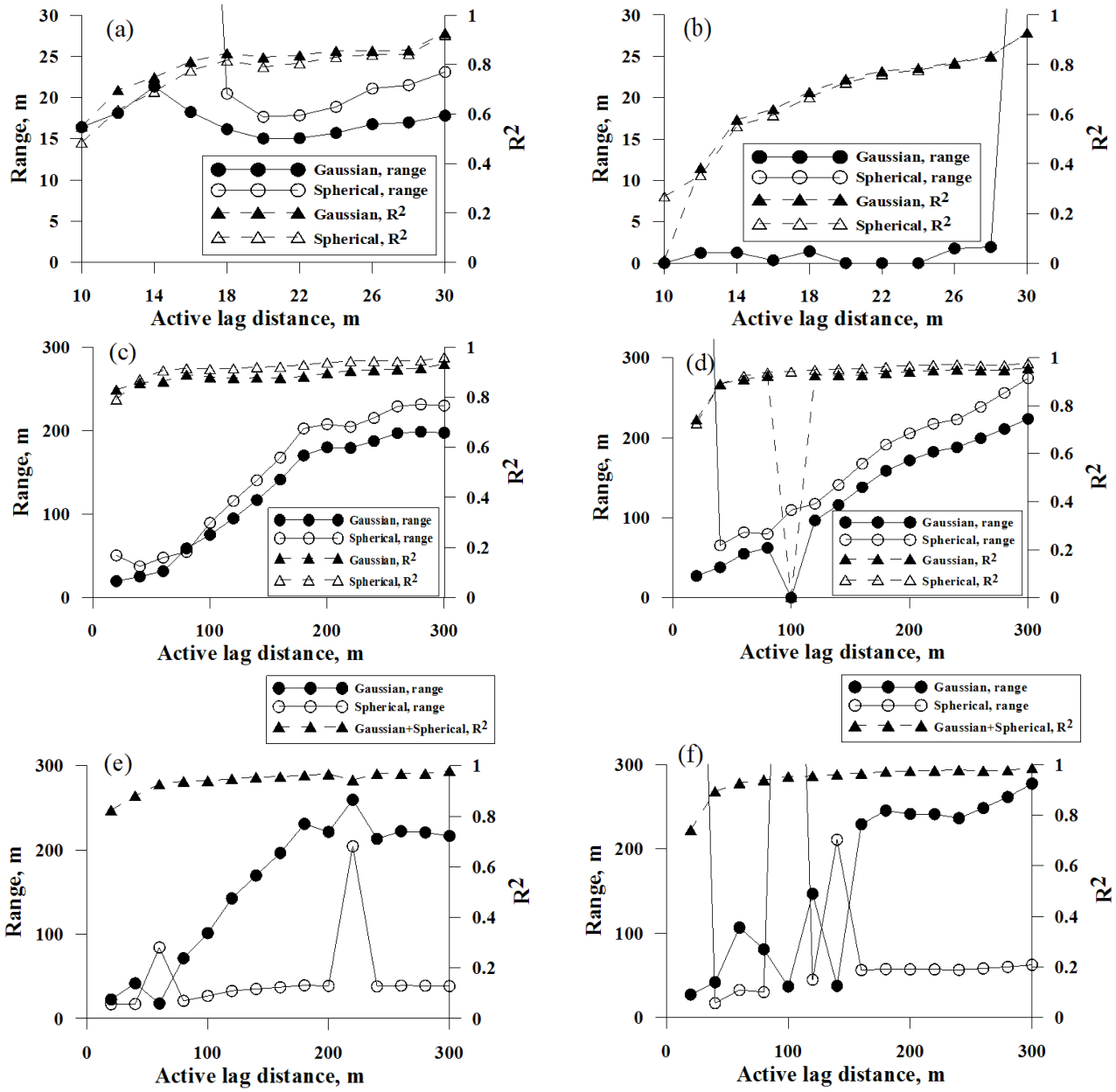

Figure 3. Ranges and R2 values of EC_{em} for Field 1 (a, c, e) and Field 2 (b, d, f), with short-lag (a, b) and long-lag (c, d) single variogram models, and nested variogram models (e, f).

R^2 changed as a function of active lag distance for the small-scale Gaussian and spherical single variogram models. The remainder of Figure 3 shows results at a larger scale with single variogram models and a nested Gaussian-spherical model. R^2 values generally increased with increasing active lag distance and were above 0.8 except at short active lags, confirming the generally good fits of the models. An exception was a near zero R^2, obtained when the range of the Gaussian model was near zero at 100-m active lag distance in Figure 3(d), assumedly due to an instability in the model-fitting procedure. Fluctuating range values were also seen in cases when models were fit with very large sill values (e.g., spherical models at 20- and 100-m active lag distances, Figure 3(f)).

For further examination, the regions where the range and R^2 remained stable over several lag intervals, indicating the possible existence of a significant structural component, were tabulated from each of these variograms plots. Table 5 shows the results for all EC_a variables with single (Gaussian or spherical) models with maximum 30-m and 300-m active lags and a nested (Gaussian-spherical) model with a maximum 300-m active lag. Gaussian and spherical

Table 5. Approximate ranges (m) of the single and nested variogram models determined at active lag distances where range was relatively stable

	Single model, 30 m active lag		Single model, 300 m active lag		Nested model, 300 m active lag	
	Gaussian	Spherical	Gaussian	Spherical	Gaussian	Spherical
Field 1						
EC_{em}	16 (18-28)[a)]	20 (18-28)	25 (20-60) 185 (180-300)	50 (20-80) 215(180-300)	220 (180-300)	40 (160-300)
EC_{sh}	13 (10-22)	17 (16-22)	60 (60-160)	70 (60-160)	60 (60-120)	13 (60-120) 60 (140-220)
EC_{dp}	Fluctuating	Fluctuating	60 (80-140) 160 (200-300)	70 (80-140) 200 (200-300)	80 (80-120)	40 (80-120) 55 (140-300)
Field 2						
EC_{em}	Near zero	Very large	Increasing	Increasing	235 (160-240)	60 (160-300)
EC_{sh}	Fluctuating Increasing	Fluctuating Increasing	Increasing	Increasing	Increasing	21 (60-100) 55 (140-260)
EC_{dp}	Fluctuating Increasing	Fluctuating Increasing	Increasing	Increasing	Fluctuating	45 (180-240)

[a)]Value in parenthesis is the active lag distance (m) that captures the corresponding range.

models were selected because they tended to fit variograms with relatively more stable range values than the exponential model. Double or triple spherical models produced similar ranges and structural components. For the Gaussian-spherical-exponential model, ranges of each model component varied widely and did not show a clear pattern. Although single variogram models provided stable range information in many cases for Field 1, they did not work well for the Field 2 EC_a data, in which stable ranges were not achieved (Figure 3, Table 5). The nested model approach did appear to work well for Field 2 and yielded ranges similar to those found for Field 1 as can be seen by comparing Figure 3(e) and (f).

From Figure 3(a), it is apparent that single variograms for EC_{em} on Field 1 included components with a range of about 16 to 20 m, possibly corresponding to transect spacing, at short active lags. When the active lag was increased to 300 m (Figure 3(c)), two regions were revealed where ranges tended to stabilize. At active lags up to 60 m, a range of about 40 m was indicated, and at active lags from 180–300 m, a range of about 200 m was indicated. However, the smaller-scale ranges seen at small active lags were no longer apparent. The significant increase in range values over active lag distances in the 80–180 m interval indicates that the single variogram models regarded the spatial structure as a trend in that region. The nested Gaussian-spherical model (Figure 3(e)) captured two components. One component remained stable over all active lags with a 40 m range, and the other component drifted until the active distance reached 180 m and then stabilized with a range of approximately 220 m. The data of Figure 3 point out potential problems in relying on a single semivariance analysis to understand spatial data structure. The ranges indicated by the analysis were, in most cases, variable with changes in the active lag distance. Nested models with relatively longer active lag distances seemed to be the best approach for obtaining the most complete information. Decomposition of the data based on the nested structure (e.g., through factorial kriging) could be an informative approach for subsequent analysis.

Results of varying the active lag distance in the semivariance analysis were tabulated for the three EC_a datasets (Table 5). In general, the Gaussian model indicated a somewhat lower range than the spherical model when both were applied separately. In the nested model, the spherical component fit to shorter ranges and the Gaussian component fit to the longer ranges. Figure 4 shows an example of the nested structure for the Field 1 EC_{em} data at both 30-m and 300-m maximum lag intervals combining spherical and Gaussian models.

Further examination of Table 5 showed several general regions of interest. For Field 1 and single variograms with up to a 30-m active lag, a stable range was apparent at 16–20 m. For single variograms up to a 300-m active lag, stable ranges were located in the intervals 25–50 m and 185–215 m. The nested Gaussian-spherical model captured both the small- and large-scale variability components. The ranges of the small-scale (40-m range spherical) and

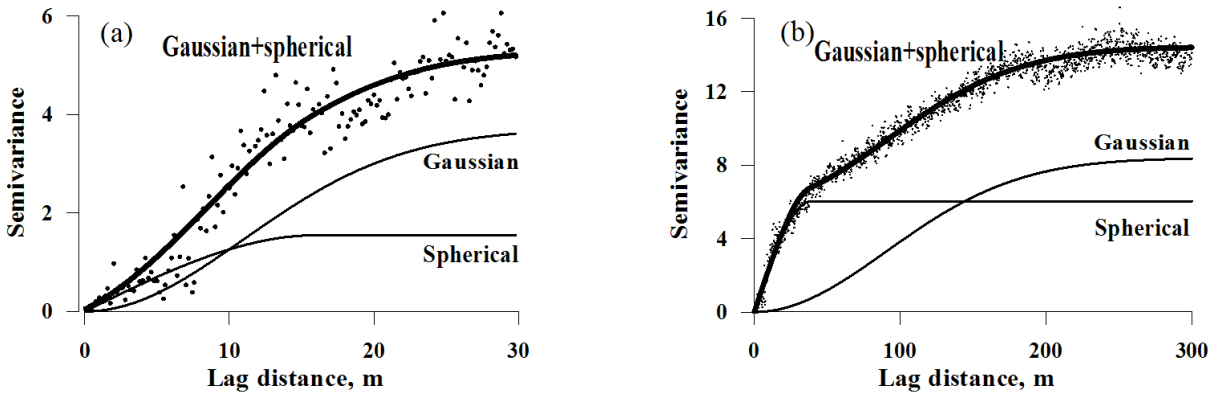

Figure 4. Gaussian-spherical nested models for EC_{em} on Field 1 with 30-m (a) and 300-m (b) active lag distances.

the large-scale (220-m Gaussian) components corresponded to the ranges found in single variograms for EC_{em} on Field 1. In other cases, a single model did not reliably determine the range of variability, while the nested model did capture the variability components present. For example, single EC_{em} variograms for Field 2 were not stable, while the nested model provided ranges similar to those found for Field 1 (Table 5).

For Field 1, the short-scale variation in EC_{em} was of similar range to Mg, K, and CEC (Table 4) that were all variables strongly loaded onto PC1 (Table 3). Cations and CEC are often strongly related to soil clay content, and spatial patterns of K and CEC were shown to be similar to EC_a on this field (Kitchen et al., 2005). Thus, we attributed these short-range (approximately 50 m) spatial structures in EC_a to variations in soil profile cation content and texture due to past erosion. Longer-range (approximately 200 m) spatial structures were likely indicative of profile soil texture variations related to landscape morphology.

Investigation of relationships between soil and site variables has often focused on conventional statistical approaches and not considered spatial structure. Geostatistical analysis of the relationships between soil variables can provide additional information useful for guiding sampling strategies and for the determination of management zones. Further, when variables (e.g., soil EC_a) are affected by multiple causative variables (e.g., soil erosion and landscape morphology), the data may contain multiple spatial structures, and separation into different components could enhance interpretation. In such cases, the nested variogram approach would be a feasible tool. In this study, we showed that the active lag distance used in variogram modeling could affect the value of the range parameter and thus the modeled spatial scale. Expanding this analysis to other variogram parameters such as nugget and sill variances might provide additional insight for determining the optimum nested variogram model. Additional analysis relating the separate components of nested datasets (e.g., soil erosion and landscape morphology effects in EC_a datasets) with other site variables (e.g., pH and cations) should be explored to provide better insights for site-specific crop and field management.

Conclusions

Geostatistical analysis, including single and nested variogram modeling along with correlation and principal component analysis, was applied to spatially referenced soil electrical conductivity and soil chemical property data to understand the relationships between the variables and the structure of their spatial variability. Specific findings include:

(1) Principal component analysis provided a reasonable way of dealing with multiple, highly collinear variables. With some exceptions, groupings produced by PCA were logical and repeatable across fields. More importantly, variables that grouped together generally showed similar spatial scales both with each other and with the PCs into which they were grouped.
(2) Nested variogram analysis of soil chemical properties was not useful as the semivariance over multiple lag intervals was very noisy for these variables. This is quite possibly due to the relatively small sampling density (~30 m) although other factors (such as laboratory analysis and/or sampling error) cannot be ruled out.
(3) More spatially dense data sets such as EC_a produced

considerably less noisy variograms, and the nested variogram approach was able to provide significantly better models of spatial structure than single models for these data sets. Further investigation regarding the effects of sampling density on variogram behavior could prove useful.

(4) The range of a variogram is not independent of the active lag distance. Although precise determination of range values may not be necessary if the goal of the analysis is to interpolate to unsampled locations, accurate range determination can be important for understanding the structure of spatial variability, particularly when considering multiple variables.

(5) Of the nested models investigated, the Gaussian-spherical model performed well over multiple active lag distances, with an ability to capture a variety of both short and long-range structures. For the example of EC_{em} on Field 1, the ranges of the small-scale (40 m range spherical structure) and the large-scale (220 m Gaussian) found by separate single variogram analyses were also captured by the nested variogram.

Conflict of Interest

The authors have no conflicting financial or other interests.

References

Bakhsh, A., D. B. Jaynes, T. S. Colvin and R. S. Kanwar. 2000. Spatio-temporal analysis of yield variability for a corn-soybean field in Iowa. Transactions of the ASAE 43(1):31-38.

Batista, A. C., E. A. Ferreira da Silva, M. C. C. Azevedo, A. J. Sousa and E. Cardoso Fonseca. 2002. Soil data analysis from central Portugal by principal component analysis and geostatistical techniques. Geochemistry, Exploration, Environment, Analysis 2:15-25.

Birrell, S. J., K. A. Sudduth and N. R. Kitchen. 1996. Nutrient mapping implications of short-range variability. In: *Proceedings of the 3rd International Conference on Precision Agriculture*, eds. P.C. Robert et al., pp. 207-216. Madison, WI: ASA, CSSA and SSSA.

Borgelt, S. C., R. E. Wieda and K. A. Sudduth. 1997. Geostatistical analysis of soil chemical properties from nested grids. Applied Engineering in Agriculture 13(4):477-483.

Brown, J. R. and R. R. Rodriguez. 1983. Soil testing in Missouri: A guide for conducting soil tests in Missouri, Univ. of Missouri Ext. Circ. 923. Columbia, Missouri: University of Missouri.

Cahn, M. D., J. W. Hummel and B. H. Brouer. 1994. Spatial analysis of soil fertility for site-specific crop management. Soil Science Society of America Journal 58:1240-1248.

Chung, S. O., I. K. Jung, J. H. Sung, K. A. Sudduth and S. T. Drummond. 2008. Analysis of spatial variability in a Korean paddy field using median polish detrending. Journal of Biosystems Engineering 33(5):362-369.

Chung, S. O., K. A. Sudduth and Y. C. Chang. 2005. Path analysis of factors limiting crop yield in rice paddy and upland corn fields. Journal of Biosystems Engineering 30(1):45-55.

Doolittle, J. A., K. A. Sudduth, N. R. Kitchen and S. J. Indorante. 1994. Estimating depths to claypans using electromagnetic induction methods. Journal of Soil and Water Conservation 49:572-575.

Drummond, S. T., K. A. Sudduth, A. Joshi, S. J. Birrell and N. R. Kitchen. 2003. Statistical and neural methods for site-specific yield prediction. Transactions of the ASAE 46(1):5-14.

Goovaerts, P. 1992. Factorial kriging analysis: a useful tool for exploring the structure of multivariate spatial soil information. Journal of Soil Science 43:597-619.

Goovaerts, P. 1998. Geostatistical tools for characterizing the spatial variability of microbiological and physio-chemical soil properties. Biology and Fertility of Soils 27:315-334.

Goovaerts, P. and R. Webster. 1994. Scale-dependent correlation between topsoil copper and cobalt concentrations in Scotland. European Journal of Soil Science 45:79-95.

Jensen, J. R. 1996. Image enhancement. In: *Introductory Digital Image Processing: A Remote Sensing Perspective*, eds. K.C. Clarke, pp. 172-179. Upper Saddle River, New Jersey: Prentice-Hall, Inc.

Journel, A. G. and C. J. Huijbregts. 1978. Mining Geostatistics. London, UK: Academic Press.

Kitchen, N. R., K. A. Sudduth and S. T. Drummond. 1999. Electrical conductivity as a crop productivity measure for claypan soils. Journal of Production Agriculture 12(4):607-617.

Kitchen, N. R., K. A. Sudduth, D. B. Myers, R. E. Massey, E. J. Sadler, R. N. Lerch, J. W. Hummel and H. L. Palm. 2005. Development of a conservation-oriented precision

agriculture system: Crop production assessment and plan implementation. Journal of Soil and Water Conservation 60(6):421-430.

Mathsoft. 2000. S+SpatialStats Version 1.5 Supplement. Seattle, Washington: MathSoft, Inc.

McBratney, A. B. and M. J. Pringle. 1999. Estimating average and proportional variograms of soil properties and their potential use in precision agriculture. Precision Agriculture 1(2):125-152.

McBratney, A. B. and R. Webster. 1986. Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates. Journal of Soil Science 37:617-639.

Oliver, M. A., R. Webster and K. Slocum. 2000. Filtering SPOT imagery by kriging analysis. International Journal of Remote Sensing 21(4):735-752.

Rodgers, S. E. and M. A. Oliver. 2007. A geostatistical analysis of soil, vegetation, and image data characterizing land surface variation. Geographical Analysis 39:195-216.

Sudduth, K. A., S. T. Drummond, S. J. Birrell and N. R. Kitchen. 1997. Spatial modeling of crop yield using soil and topographic data. *In Precision agriculture '97: Spatial variability in soil and crop*, pp. 439-447. Oxford, UK: BIOS Scientific Publishers Ltd.

Sudduth, K. A., S. T. Drummond and N. R. Kitchen. 2001. Accuracy issues in electromagnetic induction sensing of soil electrical conductivity for precision agriculture. Computers and Electronics in Agriculture 31:239-264.

Sudduth, K. A., N. R. Kitchen, G. A. Bollero, D. G. Bullock and W. J. Wiebold. 2003. Comparison of electromagnetic induction and direct sensing of soil electrical conductivity. Agronomy Journal 95(3):472-482.

Tsunemori, T. S., C. N. Washmon, M. L. Stone, J. B. Solie and W. R. Raun. 2000. Optimum field-of-view for mapping optically sensed data in winter wheat. ASAE paper No. 001007. St. Joseph, MI: ASAE.

Van Meirvenne, M., K. Maes and G. Hofman. 2003. Three dimensional variability of soil nitrate-nitrogen in an agricultural field. Biology and Fertility of Soils 37: 147-153.

Webster, R. and S. Nortcliff. 1984. Improved estimation of micro nutrients in hectare plots of the Sonning series. Journal of Soil Science 35:667-672.

Webster, R. and M.A. Oliver. 1990. Statistical methods in soil and land resource survey. New York, NY, USA: Oxford University Press.