

# Spatial Variability of Soybean Quality Data as a Function of Field Topography: II. A Proposed Technique for Calculating the Size of the Area for Differential Soybean Harvest

A. N. Kravchenko and D. G. Bullock\*

## ABSTRACT

A growing demand for soybeans [*Glycine max* (L.) Merr.] with high protein and oil concentrations has raised interest in obtaining high quality soybeans using differential harvesting practice. Differential harvesting from the locations with soybeans with higher protein or oil concentrations would be beneficial for producers and for the public who would receive a higher quality product. However, sizes and locations of the areas with high protein or oil concentrations need to be determined prior to harvesting. In this study, we develop a procedure for calculating the size of the area for differential soybean harvesting based on the relationships between protein or oil concentrations and field topography. Range of significant spatial cross-correlation between protein or oil concentrations and elevation as determined by the experimental cross-correlogram is used to calculate effective correlation distance. The effective correlation distance is proposed as an estimate of the size of the area for potential differential harvesting. The effective correlation distance is calculated based on the experimental cross-correlograms between protein or oil concentrations and elevation. Comparison of the experimental cross-correlograms and elevation variograms has shown that the range of the significant spatial cross-correlation between protein or oil concentrations and elevation and, hence, the effective correlation distance, are related to the changes in shape of the elevation variograms. Therefore, the effective correlation distance or the size of the area for differential soybean harvesting can be approximated based on the shape of the elevation variogram.

DEVELOPMENT OF PRECISION AGRICULTURE CONCEPTS strongly increased the demand for site-specific management of agricultural fields based on soil or topographical land attributes specific to certain parts of the fields. Topography is of a particular interest for site-specific management, since (i) it is frequently highly related to yield and (ii) topographical data are easy to obtain compared with time- and labor-consuming measurements of soil properties. A growing body of evidence is being collected regarding utility of topographical information for predicting soil properties and creating landscape models (Gessler et al., 2000; Moore et al., 1993) as well as regarding the importance of topography as a yield affecting factor on a field scale basis (Kravchenko and Bullock, 2000; Timlin et al., 1998). To use topographic land attributes for site-specific management it is necessary to develop guidelines for choosing size and location of management areas based on available topographical information. The objective of this study is to develop a procedure for determining the size of the area for site-specific management based on the

field topography and the spatial correlation between the variable of interest (e.g., crop yield) and topography. It is understood that the usefulness of such a procedure in each particular case will depend on the strength of the relationship between the variable of interest and topography, which may vary from field to field and from year to year.

We derived the procedure using as an example the information on spatial variability of soybean protein and oil concentrations and their relationship with topography (Kravchenko and Bullock, 2002). Kravchenko and Bullock (2002) analyzed within field spatial variability of the soybean protein and oil concentrations, and demonstrated that both soybean protein and oil concentrations were distributed within a field with well defined spatial structure and relatively large ranges of spatial correlation. Large ranges of spatial correlation indicate that there are relatively large areas within the fields with similar protein or oil concentration values. Such areas, hence, could be harvested differentially, resulting in portions of soybean grain with protein or oil concentrations greater or lesser than the field mean. However, the sizes and locations of such areas have to be defined prior to harvesting, preferably with minimal soybean seed sampling, which is costly and time-consuming.

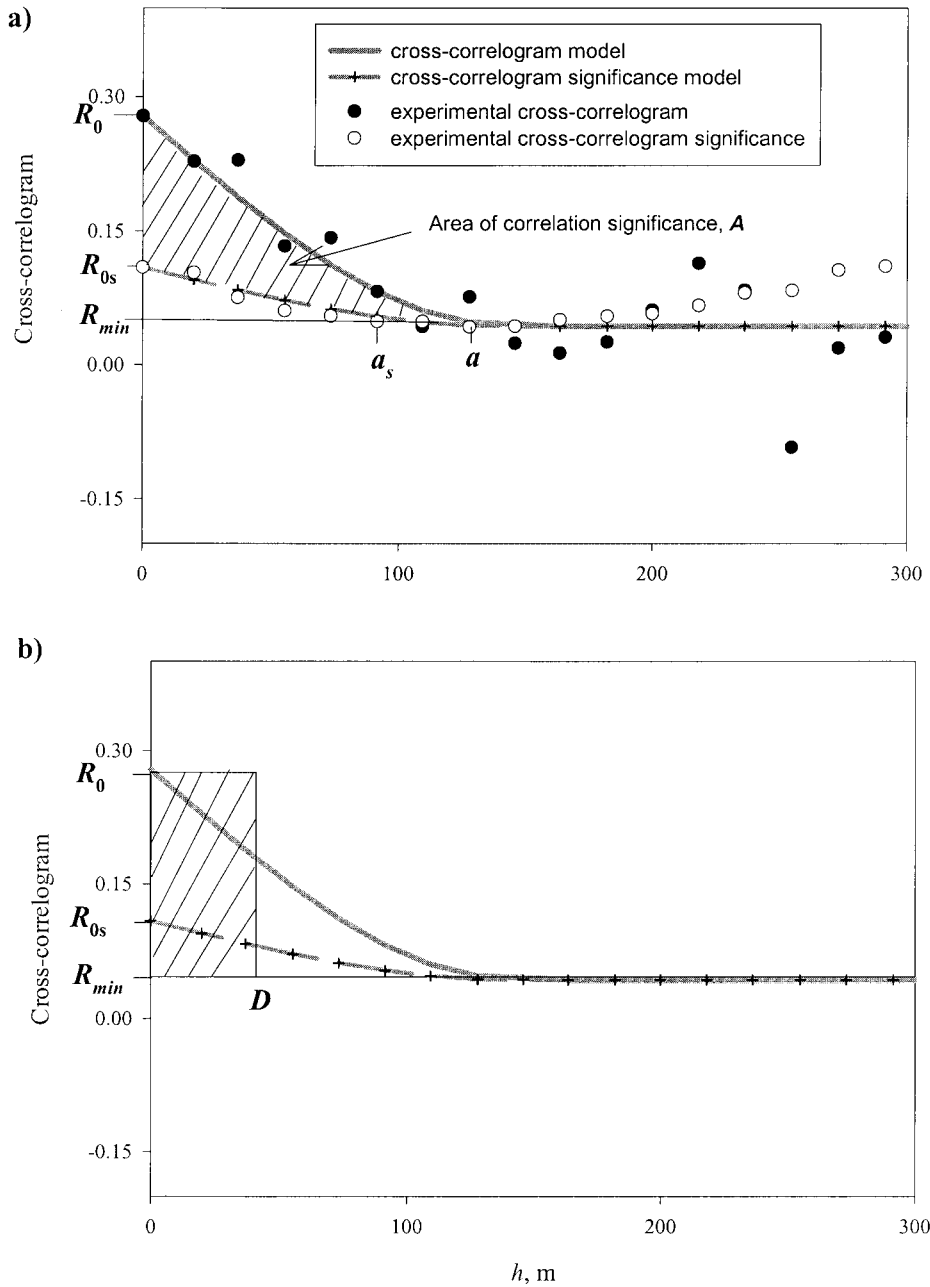
Kravchenko and Bullock (2002) found that within-field variability in topography affected spatial variability of soybean protein and oil concentrations. For example, soybean seed of high or low protein values were found in areas with certain topographical properties, particularly with areas of high or low elevation with respect to mean field elevation. The size of the area where significant correlation existed between protein concentration and elevation could be estimated based on the cross-correlogram range. Moreover, the ranges of significant cross-correlation between protein concentrations and elevation were shown to be related to the spatial variability of field elevation, and, hence, could be predicted based on the elevation data.

Although the strength and year-to-year consistence of the relationship between soybean protein and oil concentration and topography require further study before actual differential soybean harvesting can be attempted; however, in this study the available information serves as a good basis for deriving a size for the areas suitable for site-specific management based on the spatial variability of the variable of interest and field topography.

## DERIVATION OF THE AREA SIZE

We propose that the size of the area for potential differential harvesting based on soybean protein or oil

Dep. of Crop Sciences, 1102 S. Goodwin Ave., Univ. of Illinois, Urbana, IL 61801. Received 6 Mar. 2001. \*Corresponding author (dbullock@uiuc.edu).



**Fig. 1.** Examples of (a) an experimental cross-correlogram and cross-correlogram significance limit ( $P = 0.05$ ) along with cross-correlogram and cross-correlogram significance models, and (b) cross-correlogram models with effective correlation distance,  $D$ . Shaded area represents the amount of significant correlation between either protein or oil concentration and elevation existing over the distance  $a$ .

concentration relationships with topography can be defined using cross-correlogram data, similarly to the definition of the cell size for site-specific management proposed by Han et al. (1994). For deriving the area size, we assume that the total amount of the significant spatial correlation between either protein or oil concentration and elevation is represented by the difference between the areas under the curve of the cross-correlogram and under the curve of cross-correlogram significance (Fig. 1). For negative correlations, the areas above the respective curves are used. Since cross-correlogram is by definition a correlation coefficient between the data separated by various distances, its statistical significance is

determined as that of a Pearson correlation coefficient taking into account the number of data pairs used to calculate the cross-correlogram value at each distance. The definition of the size is meaningful only if the cross-correlogram values are statistically significant at distances ranging from zero (or close to zero) to several lag sizes. The distance at which cross-correlogram becomes insignificant is further called cross-correlogram range of significance,  $a$ . The other parameters used in further discussion include cross-correlogram value at zero distance,  $R_0$ , which is equal to a Pearson correlation coefficient, and the critical significant cross-correlogram value at the distance  $a$ ,  $R_{min}$  (Fig. 1a).

If the experimental cross-correlogram is statistically significant within a certain range,  $a$ , and can be fitted with a mathematical model  $f(h)$ , where  $h$  is the separation distance, then, the area under the cross-correlogram curve on interval from 0 to  $a$ , can be defined as

$$F(h) = \int_0^a f(h)dh. \quad [1]$$

The respective area under the curve of the cross-correlogram significance limit is obtained as

$$S(h) = \int_0^a s(h)dh, \quad [2]$$

where  $s(h)$  is the mathematical model describing the curve. The area of correlation significance,  $A$ , which defines the total amount of significant spatial correlation between the variables, is equal to

$$A = F(h) - S(h) = \int_0^a f(h)dh - \int_0^a s(h)dh \quad [3]$$

The area between the curves  $f(h)$  and  $s(h)$ ,  $A$ , corresponds to an area of a rectangle, one side of which is equal to  $(R_0 - R_{\min})$  and the other side represents an effective correlation distance,  $D$  (Fig. 1b). At this distance, the amount of all significant spatial correlation between the studied variables is accounted for (Han et al., 1994) and we suggest using it as an indicator of the optimal size of the area for site-specific management based on elevation. The rationale behind using the effective correlation distance,  $D$ , instead of the correlogram range of significance,  $a$ , is that at the distances close to  $a$  the correlation with topography is very weak. Hence, using  $a$  would overestimate the size of the area suitable for site-specific management with meaningfully strong relationship between the variable of interest and topography. In terms of the effective correlation distance,  $D$ , the area of significant correlation will be equal to

$$A = (R_0 - R_{\min})D. \quad [4]$$

To find the area of significant correlation,  $A$ , mathematical models  $f(h)$  and  $s(h)$  for fitting experimental cross-correlograms and cross-correlogram significance limits need to be defined. Depending on the shape of the experimental cross-correlogram we propose two types of mathematical models to fit the experimental data. A spherical model of the type

$$f(h) = R_0 - (R_0 - R_{\min})[1.5(h/a) - 0.5(h/a)^3], \quad \text{for } 0 \leq h \leq a \quad [5]$$

was used for the cross-correlograms that had the highest value at zero  $h$ , and decreased rapidly with increasing  $h$ . For the cross-correlograms with a relatively moderate decrease at increasing  $h$ , we used a power model:

$$f(h) = R_0 - (R_0 - R_{\min})(h/a)^3, \quad \text{for } 0 \leq h \leq a. \quad [6]$$

The spherical model (Eq. [5]) was also used to fit the curves of cross-correlogram significance limit:

$$s(h) = R_{0s} - (R_{0s} - R_{\min})[1.5(h/a_s) - 0.5(h/a_s)^3], \quad \text{for } 0 \leq h \leq a_s$$

$$s(h) = R_{\min}, \quad \text{for } a_s \leq h \leq a, \quad [7]$$

where  $R_{0s}$  was a critical significant cross-correlogram value at zero distance, and  $a_s$  was a distance at which critical significant cross-correlogram values approached  $R_{\min}$  value (Fig. 1a).

For the fields where cross-correlogram was fitted with a spherical model (Eq. [5]), one can calculate the area of correlation significance,  $A$ , by substituting models describing experimental cross-correlograms and cross-correlogram significance limits (Eq. [5] and [7]) in corresponding Eq. [1] and [2] and performing integration, as

$$A = a(3/8R_0 + 5/8R_{\min}) - [a_s(3/8R_{0s} + 5/8R_{\min}) + (a - a_s)R_{\min}]. \quad [8]$$

The effective correlation distance,  $D$ , is obtained from combining Eq. [4] and [8] as

$$D = \frac{a(3/8R_0 + 5/8R_{\min}) - [a_s(3/8R_{0s} + 5/8R_{\min}) + (a - a_s)R_{\min}]}{(R_0 - R_{\min})}. \quad [9]$$

For the fields where the experimental cross-correlogram was fitted with the power model (Eq. [6]), the area of significant correlation is defined by substituting Eq. [6] and [7] in Eq. [1] and [2] with following integration as

$$A = a(3/4R_0 + 1/4R_{\min}) - [a_s(3/8R_{0s} + 5/8R_{\min}) + (a - a_s)R_{\min}], \quad [10]$$

and the corresponding effective correlation distance for these fields will be equal to

$$D = \frac{a(3/4R_0 + 1/4R_{\min}) - [a_s(3/8R_{0s} + 5/8R_{\min}) + (a - a_s)R_{\min}]}{(R_0 - R_{\min})}. \quad [11]$$

If we assume that the change in cross-correlogram significance limit values is negligible over the distance and the curve for the cross-correlogram significance limit can be approximated by a straight line of the type  $s(h) = R_{\min}$  (Fig. 1), we can derive simplified equations for the effective correlation distance. The simplified effective correlation distance,  $D_{\text{simple}}$ , for the variables with cross-correlograms fitted with spherical model can be calculated as

$$D_{\text{simple}} = \frac{a(3/8R_0 + 5/8R_{\min}) - aR_{\min}}{(R_0 - R_{\min})} = 3/8a. \quad [12]$$

For variables with cross-correlograms fitted with the power model, the simplified effective correlation distance is equal to

$$D_{\text{simple}} = \frac{a(3/4R_0 + 1/4R_{\min}) - aR_{\min}}{(R_0 - R_{\min})} = 3/4a. \quad [13]$$

The assumption that the curve for the cross-correlogram significance limit can be approximated by a straight

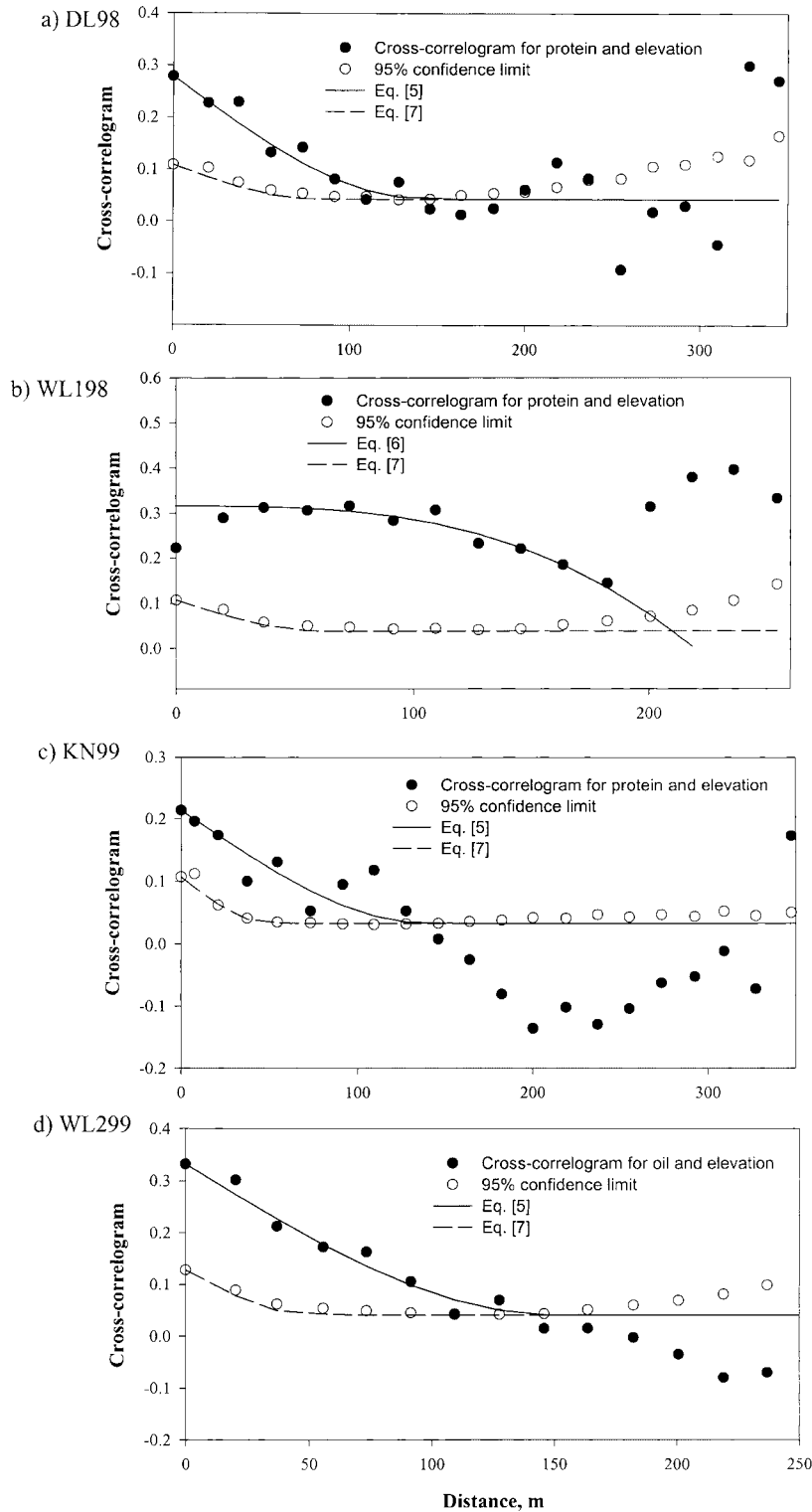


Fig. 2. Experimental cross-correlograms and cross-correlogram significance limits ( $P = 0.05$ ) along with cross-correlogram and cross-correlogram significance models for (a) protein and elevation data from the DL98 field, (b) protein and elevation data from the WL198 field, (c) protein and elevation data from the KN99 field, and (d) oil and elevation data from the WL299 field.

line results in slightly overestimated cross-correlogram significance areas. Hence, the effective distances calculated using Eq. [12] and [13] will be higher compared

with the true values (Eq. [9] and [11]). Equations [12] and [13], however, can be used as rule of thumb approximations for effective correlation distances.

## MATERIALS AND METHODS

Data on soybean protein and oil concentrations from four agricultural fields located in eastern Illinois were used in the study. Soybean seed samples were collected in the fall of 1998 from the DL98 and WL198 fields and in the fall of 1999 from the KN99 and WL299 fields. The number of soybean samples collected were equal to 334, 339, 336, and 235 for the DL98, WL198, KN99, and WL299 fields, respectively. Detailed elevation measurements were taken for each field with distance between the measurements of  $\approx 10$  m for the WL198 and WL299 fields and from 2 to 60 m for the DL98 and KN99 fields. A detailed description of the procedures for soybean protein and oil data sampling and elevation measurements is given in Kravchenko and Bullock (2002).

For each of the fields, we calculated cross-correlograms for protein vs. elevation and oil vs. elevation data (Goovaerts, 1997). Significance levels of the cross-correlograms were calculated for each lag distance based on the number of data pairs available at this lag for the cross-correlogram calculation. For determination of the differential harvesting area, we used the cross-correlograms which were significant at zero distances or at distances close to zero, and for which the range of significant cross-correlogram values extended to several lag distances.

## RESULTS AND DISCUSSION

For determination of the effective correlation distances we used cross-correlograms for protein vs. elevation data from the DL98, WL198, and KN99 fields, and the cross-correlogram for oil vs. elevation data from the WL299 field (Fig. 2). Experimental cross-correlograms and cross-correlogram significance curves were fitted with mathematical models described above in Eq. [5] and [6]. The spherical model (Eq. [5]) was used to fit protein vs. elevation cross-correlograms from the DL98 and KN99 fields and oil vs. elevation cross-correlograms from the WL299 field. The power model (Eq. [6]) was used to fit protein vs. elevation cross-correlogram from the WL198 field. Cross-correlogram significance curves for all fields were fitted with Eq. [7]. Experimental cross-correlograms and their respective fitted models are shown in Figs. 2a, 2b, 2c, and 2d for the DL98, WL198, KN99, and WL299 fields, respectively.

Parameters of the models were estimated during model fitting, and effective correlation distances,  $D$ , were calculated based on the model parameters using Eq. [9] for data from the DL98, KN99, and WL299 fields and Eq. [11] for data from the WL198 field. Simplified effective correlation distances,  $D_{\text{simple}}$ , were calculated

using cross-correlogram ranges of significance,  $a$ , (Eq. [12] and [13]). Model parameters as well as calculated  $D$  and  $D_{\text{simple}}$  values are presented in Table 1. For the data with cross-correlograms rapidly decreasing with  $h$ , the effective correlation distances were relatively small ( $\approx 50$  m), while for the WL198 data set with cross-correlogram decreasing less rapidly, the effective correlation distance was equal to 152 m. The effective correlation distances calculated using simplified equations were 6 to 9 m larger than the  $D$  values from Eq. [9] and [11].

Kravchenko and Bullock (2002) showed that elevation data can be used for approximation of the significant cross-correlogram range. Ranges of elevation variograms or distances at which noticeable changes in variogram slopes occurred were similar in size to the significant cross-correlogram range,  $a$ . At the DL98 field, the range of significant cross-correlogram was similar to the range of the first of the nested structures in the elevation variograms (145 m). At the WL198 and WL299 fields, the range of significant cross-correlograms were close to the distances at which there was a change in elevation variogram slope (147 m) and (100 m). At the KN99 field, the range of significant cross-correlogram corresponded to the elevation variogram range (164 m). The effective correlation distances calculated based on the elevation data using simplified equations can, therefore, provide an approximation for the size of the differential harvesting area for the fields without soybean sampling (Table 1). Significant cross-correlograms were observed in fields where topography was represented by diverse landscape forms with relatively large areas occupied by summits as well as foot slopes or depressions (Kravchenko and Bullock, 2002). In such fields, the potential for differential harvesting might be better than in the fields with a single dominating landscape form. The sharper contrast between protein or oil concentration values of summits and depressions might be expected in years with particularly dry or wet weather. Hence, in such years the differential harvesting might be more beneficial than in years with moderate weather conditions.

## CONCLUSIONS

In this study we developed a procedure for determining the size of the areas within a given agricultural field applicable for implementing site-specific management

**Table 1. Parameters of the mathematical models used to fit the experimental cross-correlograms and cross-correlogram significance curves along with effective correlation distances proposed for differential soybean harvesting based on topography.**

Field	Model parameters						$D$	$D_{\text{simple}}$	$D_{\text{simple}}$ from elevation data
	Model	$R_0^\dagger$	$R_{\text{min}}^\ddagger$	$a^\S$	$R_{0s}^\parallel$	$a_\#$			
				m				m	
DL98(protein)	Eq. [4]	0.279	0.043	140	0.109	80	44 Eq. [9]	53 Eq. [12]	54
WL198(protein)	Eq. [5]	0.316	0.038	210	0.107	60	152 Eq. [11]	158 Eq. [13]	110
KN99(protein)	Eq. [4]	0.214	0.033	140	0.107	50	45 Eq. [9]	53 Eq. [12]	62
WL299(oil)	Eq. [4]	0.332	0.041	130	0.128	50	43 Eq. [9]	49 Eq. [12]	38

$^\dagger R_0$  is the cross-correlogram value at zero distance.

$^\ddagger R_{\text{min}}$  is the critical significant cross-correlogram value at the distance  $a$ .

$^\S a$  is the cross-correlogram range of significance.

$^\parallel R_{0s}$  is a critical significant cross-correlogram value at zero distance.

$^\# a_\#$  is a distance at which critical significant cross-correlogram values approached  $R_{\text{min}}$  value.

tasks, including differential harvesting, tillage, fertilizer application, irrigation, etc. The requirements for the successful application of the procedure are that a relationship exists between the variable of interest (which may be either crop yields, or soil N, P, and K levels, or soil physical properties) and topography, and that the relationship extends over a substantial area (relatively large ranges of spatial correlation). An example of application of the procedure was presented in this study. The variables of interest were soybean protein and oil concentrations, and the procedure was used for calculation of a size of an area suitable for differential soybean harvesting. The size of the area appropriate for differential harvesting was assumed to be equal to an effective correlation distance calculated based on the significant experimental cross-correlograms of protein or oil concentration and elevation. The proposed procedure was applied to calculate the effective correlation distances for protein and oil distributions of the four studied fields. Depending on the shape of the experimental cross-

correlograms, the effective correlation distances ranged from 50 to 150 m in different fields.

### REFERENCES

- Gessler, P.E., O.A. Chadwick, F. Chamran, L. Althouse, and K. Holms. 2000. Modeling soil-landscape and ecosystem properties using terrain attributes. *Soil Sci. Soc. Am. J.* 64:2046–2056.
- Goovaerts, P. 1997. *Geostatistics for natural resources evaluation*. Oxford University Press, New York.
- Han, S., J.W. Hummel, C.E. Goering, and M.D. Cahn. 1994. Cell size selection for site-specific crop management. *Trans. ASAE* 37:19–26.
- Kravchenko, A.N., and D.G. Bullock. 2000. Correlation of grain yield with topography and soil properties. *Agron. J.* 92:75–83.
- Kravchenko, A.N., and D.G. Bullock. 2002. Spatial variability of soybean quality data as a function of field topography: I. Spatial data analysis. *Crop Sci.* 42:804–815 (this issue).
- Moore, I.D., P.E. Gessler, and G.A. Nielson. 1993. Soil attribute prediction using terrain analysis. *Soil Sci. Soc. Am. J.* 57:443–452.
- Timlin, D.J., Ya. Pachepsky, V.A. Snyder, and R.B. Bryant. 1998. Spatial and temporal variability of corn grain yield on a hillslope. *Soil Sci. Soc. Am. J.* 62:764–773.