

University of Nebraska - Lincoln  
**DigitalCommons@University of Nebraska - Lincoln**

---

Faculty Publications, Department of Statistics

Statistics, Department of

---

2007

# Changing the Support of a Spatial Covariate: A Simulation Study

Tisha L. Hooks

*University of Nebraska - Lincoln*

Jeffrey F. Pedersen

*University of Nebraska - Lincoln, [jpedersen1@unl.edu](mailto:jpedersen1@unl.edu)*

David B. Marx

*University of Nebraska-Lincoln, [david.marx@unl.edu](mailto:david.marx@unl.edu)*

Roch E. Gaussoin

*University of Nebraska - Lincoln, [rgaussoin1@unl.edu](mailto:rgaussoin1@unl.edu)*

Follow this and additional works at: <http://digitalcommons.unl.edu/statisticsfacpub>



Part of the [Other Statistics and Probability Commons](#)

---

Hooks, Tisha L.; Pedersen, Jeffrey F.; Marx, David B.; and Gaussoin, Roch E., "Changing the Support of a Spatial Covariate: A Simulation Study" (2007). *Faculty Publications, Department of Statistics*. 40.

<http://digitalcommons.unl.edu/statisticsfacpub/40>

This Article is brought to you for free and open access by the Statistics, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications, Department of Statistics by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

# Changing the Support of a Spatial Covariate: A Simulation Study

Tisha Hooks,\* Jeffrey F. Pedersen, David B. Marx, and Roch E. Gaussoin

## ABSTRACT

Researchers are increasingly able to capture spatially referenced data on both a response and a covariate more frequently and in more detail. A combination of geostatistical models and analysis of covariance methods may be used to analyze such data. However, very basic questions regarding the effects of using a covariate whose support differs from that of the response variable must be addressed to utilize these methods most efficiently. In this experiment, a simulation study was conducted to assess the following: (i) the gain in efficiency when geostatistical models are used, (ii) the gain in efficiency when analysis of covariance methods are used, and (iii) the effects of including a covariate whose support differs from that of the response variable in the analysis. This study suggests that analyses which both account for spatial structure and exploit information from a covariate are most powerful. Also, the results indicate that the support of the covariate should be as close as possible to the support of the response variable to obtain the most accurate experimental results.

T. Hooks and D.B. Marx, Dep. of Statistics, Univ. of Nebraska–Lincoln, Lincoln, NE 68583-0963; J.F. Pedersen, USDA-ARS, NPA Grains, Bioenergy, and Forage Research, Univ. of Nebraska–Lincoln, Lincoln, NE 68583-0897; R.E. Gaussoin, Dep. of Agronomy and Horticulture, Univ. of Nebraska–Lincoln, Lincoln, NE 68583-0915. Received 25 July 2006.  
\*Corresponding author (THooks@winona.edu).

**Abbreviations:** AICC, corrected Akaike information criteria.

RECENT ADVANCES in precision agriculture have provided researchers with the ability to collect various measurements such as infrared and visible light reflectance data (Servilla, 1998), which are indicative of such factors as moisture status during various stages of crop development (Bryant et al., 2003), and “on-the-go” data during harvest such as electrical conductivity readings (McGuire, 2003), yield, and test-weight readings (Wehrspann, 2000). Similar data are also available from satellite imagery (Frazier et al., 2004). These data points are typically associated with extremely dense spatial coordinates, thus creating the opportunity to use these measurements as covariates for the primary response variable to possibly increase experimental precision. As technologies continue to improve concerning on-the-go data collection and the precision of imagery, the importance and potential impact of utilizing such data in planned experiments will increase.

In addition to the growing availability of massive amounts of spatially coordinated data, researchers have witnessed a rapid increase in the speed and power of computers, which allows researchers to effectively manage such data. Also, a collection of geostatistical models allows the researcher to both characterize and account for the underlying spatial patterns in their data, leading to potentially more precise estimation (Journel and Huijbregts, 1978; Isaaks and Srivastava, 1989; Cressie, 1993). A good introduction to geostatistical methods is given by Littel et al. (1996) in a chapter dealing with spatial variability. Ultimately, the parallel

Published in *Crop Sci* 47:622–628 (2007).

doi: 10.2135/cropsci2006.07.0490

© Crop Science Society of America

677 S. Segoe Rd., Madison, WI 53711 USA

This document is a U.S. government work and is not subject to copyright in the United States.

increases in computer technology and the ability to collect (or access) spatial data has created a need for research on how to manage data and use resources efficiently.

As mentioned, these large data sets provide researchers with the opportunity to use some measurements as covariates, thereby improving estimation by utilizing information about one variable that is contained in another. Analysis of covariance methods are used to analyze such data (Searle, 1971). However, basic questions need to be answered regarding the use of intensively collected data points as covariates in analyzing data collected over an entire plot (e.g., yield) or data collected at a single point in a plot to represent the entire plot (e.g., soil chemistry from a soil probe). For example, would greater experimental precision be obtained by utilizing all intensively collected data points from a plot as covariates for a trait such as yield, or would some subset of the intensively collected data points provide greater experimental precision? This question is related to what is widely known in geostatistics as the change of support problem (Olea, 1991; Cressie, 1993; Schabenberger and Gotway, 2005). The support of the data refers to the length, area, or volume that a measured datum represents. Note that in many cases the data is collected at a single point; thus, it is said to have “point support.” If all intensively collected data points are utilized in the analysis (e.g., by obtaining a block average of all data points included in a plot and using the block average as the new variable), then the support of the data has been changed. Effectively, this block average is a new variable, and the statistical and spatial properties of this new variable differ from those of the original. In particular, the spatial structure and parameters such as the range and sill of the corresponding semivariogram for this new variable are altered.

For example, a variable of point support may be associated with a semivariogram such as the one shown in Fig. 1. This particular example illustrates a spherical model. This is one of the geostatistical models referenced earlier, which allows us to characterize and account for underlying spatial variability. In general, the semivariogram is a measure of the average dissimilarity between data separated by a distance  $h$ . Note that since this function is a measure of dissimilarity, we see that the value of the semivariogram increases with lag distance  $h$ . The parameters of the semivariogram are the range, sill, and nugget. For the spherical semivariogram, the range is defined as the critical distance above which observations become independent and beyond which the model function returns a constant value, the sill. The sill is equal to the variance of independent observations. Finally, the nugget describes microscale variation that may cause a discontinuity at the origin. Changing the support of the data by averaging over observations will change these parameter values and may possibly change the results of the analysis considerably

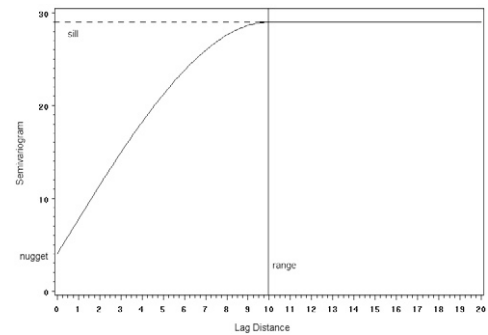


Figure 1. Example of a spherical semivariogram.

(Clark, 1979). In addition to the conventional change of support problem, questions also arise regarding the effects of conducting an analysis of covariance when the response variable and the covariate are of different supports.

To answer such questions, one would ideally know the true values of the treatment and response variables along with their spatial structure and conduct numerous replicates of each experiment to place confidence in the results. Simulation studies provide such capacity. Therefore, the objectives of this research were to conduct a simulation study to explore (i) the gain in efficiency when methods that exploit spatial structure are utilized, (ii) the gain in efficiency when methods that exploit information from a covariate are utilized, and (iii) the effects of including a covariate whose support differs from that of the response variable in the analysis.

## MATERIALS AND METHODS

The simulated experiment consisted of five replications of five treatments. The treatments were laid out in a completely randomized design on a  $5 \times 5$  arrangement of plots and were randomly assigned to the plots in each iteration. Within each plot, another  $5 \times 5$  grid of points was constructed (Fig. 2). For each of the 625 points, both a spatial floor ( $Y$ ) and a spatial covariate ( $X$ ) were generated using the method of Gaussian cosimulation (Oliver, 2003). This method is described as follows: let

$$Y = \mu_1 + L_1 \mathbf{Z}_1$$

and

$$X = \mu_2 + L_2(\rho \mathbf{Z}_1 + \sqrt{1 - \rho^2} \mathbf{Z}_2)$$

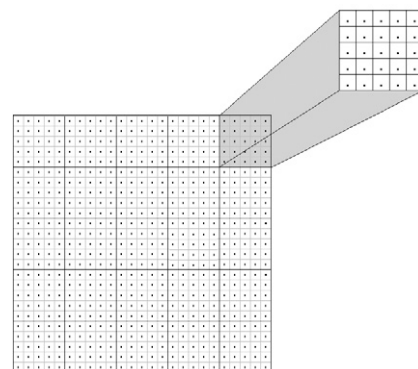


Figure 2. Layout for data generation.

where  $L_1$  and  $L_2$  are the square roots (which can be obtained via methods such as the Cholesky decomposition and spectral decomposition) of given covariance matrices,  $\mu_1$  and  $\mu_2$  are the means of  $Y$  and  $X$ , and  $\mathbf{Z}_1$  and  $\mathbf{Z}_2$  are vectors of independent normally distributed random variables with mean 0 and variance 1. Then, it is easily shown that the covariance of  $Y$  is  $\text{cov}(Y)=L_1L_1'$ , the covariance of  $X$  is  $\text{cov}(X)=L_2L_2'$ , and the cross-covariance of  $Y$  and  $X$  can be written as  $\text{cov}(Y,X)=\rho L_1L_2'$  (Oliver, 2003).

Note that the parameter  $\rho$  ( $-1 \leq \rho \leq 1$ ) determines the strength of the relationship that exists between the spatial floor and the covariate.

In this simulation, the spherical covariance function with a nugget of zero was used for the construction of both variables. The function is as follows:

$$C(h) = \begin{cases} \sigma^2 \left\{ 1 - \frac{3}{2} \left( \frac{h}{a} \right) + \frac{1}{2} \left( \frac{h}{a} \right)^3 \right\} & \text{if } 0 \leq h \leq a \\ 0 & \text{if } h > a \end{cases}$$

where  $h$  is the distance between observations,  $a$  is the range of the corresponding spherical semivariogram, and  $\sigma^2$  is the sill of the semivariogram.  $Y$  was simulated with a range of 25 and a sill of 5.  $X$ , the covariate, was simulated with a range of 15 and a sill of 5. Finally, two correlation values were considered when modeling the cross-covariance between the spatial floor and the covariate so that both a weak and a strong relationship between the two variables could be considered:  $\rho = .3$  and  $\rho = .7$ . Treatment effects were generated with the following treatment vectors  $\boldsymbol{\tau}$ :

$$\boldsymbol{\tau} = \left(-1, -\frac{1}{2}, 0, \frac{1}{2}, 1\right),$$

$$\boldsymbol{\tau} = \left(\sqrt{\frac{1}{3}}, \sqrt{\frac{1}{3}}, \sqrt{\frac{1}{3}}, \frac{-3}{2}\sqrt{\frac{1}{3}}, \frac{-3}{2}\sqrt{\frac{1}{3}}\right), \text{ and}$$

$$\boldsymbol{\tau} = (0, 0, 0, 0, 0)$$

The first vector represents the case where treatment effects are equally spaced, while the second vector represents the case where treatments are set up in a maximum–minimum configuration. Note that both cases represent the same noncentrality parameter, hence the null hypothesis for no treatment effect should be rejected similarly. Finally, the last vector corresponds to the case where there is no treatment effect.

A response variable and a covariate were created for each plot as follows. A response consisting of the sum of the spatial floor and the treatment effect was generated for each of the 625 points. The response variable for each of the 25 plots was then generated by averaging the responses of all points within a plot. Note that this is analogous to a researcher collecting data on a response such as yield over an entire plot. Then, to investigate whether or not greater experimental precision can be obtained by utilizing all intensively collected data points, three different covariates were considered. First, the covariate for each plot was taken to be the center point of the  $5 \times 5$  grid contained

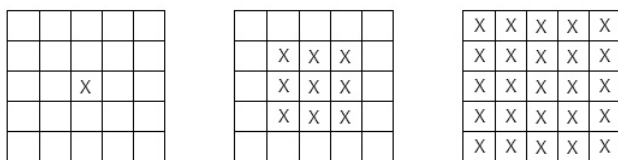


Figure 3. Illustration of the covariate selection. (a) The covariate is obtained from the central point of the plot. (b) The covariate is obtained by averaging the observations in the central  $3 \times 3$  square of the plot. (c) The covariate is obtained by averaging all 25 observations in the plot.

in that plot (Fig. 3a). Second, the covariate for each plot was obtained by averaging the simulated covariate values for all data points in the central  $3 \times 3$  square of each plot (Fig. 3b). These two situations investigate the effects of using only a subset of the collected data points on experimental precision. Finally, the covariate was obtained by averaging all 25 of the simulated covariate values within each plot (Fig. 3c) to represent the case where all intensively collected data points are incorporated into the analysis. Note that in this case, the support of the response variable and covariate are identical.

Recall that this study considered two values for  $\rho$ . In both cases, 1000 data sets were simulated using the same seed value for each  $\rho$ ; thus, the  $i$ th iteration for  $\rho = .3$  had exactly the same treatment randomization and  $\mathbf{Z}_1$  and  $\mathbf{Z}_2$  vectors as the  $i$ th iteration for  $\rho = .7$ . These data sets were subsequently analyzed in four ways using SAS PROC MIXED (SAS Institute, 2003). First, the data were analyzed using a traditional analysis of variance (non-spatial analysis with no covariate). This analysis ignores the spatial structure of the response variable and does not utilize any information that the covariate may contribute. Second, the data were analyzed using an analysis of covariance (nonspatial analysis of covariance), where a separate analysis was conducted for each of the three covariates. This exploits any information the covariate has to offer; however, it still ignores the spatial structure that is present in the data. Next, the data were analyzed using an analysis of variance that included a spatial component in the model but ignored the covariate (spatial analysis with no covariate). Finally, an analysis of covariance (one for each of the three covariates) that included a spatial component in the model was conducted to exploit the spatial structure and the information available from the covariate to improve estimation (spatial analysis of covariance). These analyses were compared on the basis of percent rejection rate of the  $F$ -tests for overall equal means.

A similar simulation study was conducted at the conclusion of the first experiment. In the second study, the response variable was generated by selecting the center observation of each plot to characterize the entire plot. This corresponds to the case where the researcher uses a single data point (e.g., data obtained by means of a core sample) to represent the response of the entire plot. Again, to investigate whether or not greater experimental precision can be obtained by utilizing all intensively collected data points, the three different covariates (Fig. 3) were considered. It was of interest to determine whether or not the covariate whose support was identical or nearest to that of the response resulted in more accurate experimental results. Finally, 1000 data sets were generated and analyzed as described above for each value of  $\rho$  (using the same seed value as in the previous study for each  $\rho$ ), and the analyses were compared on the basis of percent rejection rate of the  $F$ -tests for overall equal means.

A field experiment was conducted to establish credibility of the simulation study. The experiment was conducted on a turf-grass site near Ithaca, NE. The site was divided into a  $5 \times 5$  grid of 25 plots, each plot 1 m square. Each plot was further divided into another  $5 \times 5$  grid (each subdivision 20 cm square). The plots were treated with five different rates of 46-0-0 urea nitrogen:  $\text{kg/ha}^{-1}$ , 6.1  $\text{kg/ha}^{-1}$ , 12.2  $\text{kg/ha}^{-1}$ , 24.4  $\text{kg/ha}^{-1}$ , and 48.8  $\text{kg/ha}^{-1}$ . The experimental design was a  $5 \times 5$  knight's move Latin square. This type of design employs the idea that repetitions of a treatment should be a knight's move (from chess) apart (Martin,

**Table 1a Rejection rates for analyses with response averaged over all 25 observations in the plot and with treatment effects equally spaced.<sup>†</sup>**

No. of observations in covariate	$\rho = .3$				$\rho = .7$	
	NS/NOCOV	S/NOCOV	NS/COV	S/COV	NS/COV	S/COV
1	0.350	0.874	0.386	0.885	0.512	0.963
9	–	–	0.389	0.894	0.556	0.977
25	–	–	0.395	0.899	0.570	0.983

**Table 1b. Rejection rates for analyses with response averaged over all 25 observations in the plot and with treatment effects in maximum–minimum configuration.**

No. of observations in covariate	$\rho = .3$				$\rho = .7$	
	NS/NOCOV	S/NOCOV	NS/COV	S/COV	NS/COV	S/COV
1	0.376	0.880	0.400	0.878	0.520	0.953
9	–	–	0.407	0.888	0.575	0.973
25	–	–	0.411	0.886	0.586	0.975

**Table 1c. Rejection rates for analyses with response averaged over all 25 observations in the plot and with no treatment effect.**

No. of observations in covariate	$\rho = .3$				$\rho = .7$	
	NS/NOCOV	S/NOCOV	NS/COV	S/COV	NS/COV	S/COV
1	0.051	0.041	0.053	0.042	0.049	0.044
9	–	–	0.054	0.041	0.048	0.032
25	–	–	0.047	0.042	0.050	0.030

<sup>†</sup>NS/NOCOV, nonspatial analysis with no covariate; S/NOCOV, spatial analysis with no covariate; NS/COV, nonspatial analysis of covariance; S/COV, spatial analysis of covariance.

1986). Visual quality measurements were taken on a scale of 1 to 9 with a 9 indicating best turfgrass quality on each of the 25 points within each plot before nitrogen application, yielding a total of 625 quality measurements. Yield measurements were also taken at each of the 625 points 3 wk after nitrogen application. The response variable for each of the 25 plots was generated in three ways: as the yield measurement of the center point of the 5 × 5 grid within the plot, as the average of the 9 yield measurements in the central 3 × 3 square of each plot, and as the average of the 25 yield measurements within each plot. The covariate for each plot was also constructed in three ways: as the quality measurement of the center point of the 5 × 5 grid contained in the plot, as the average quality measurement of the 9 observations in the central 3 × 3 square of each plot, and as the average of all 25 quality measurements within each plot. The data were analyzed for each of the three response variables using the maximum likelihood method in SAS PROC MIXED as follows: using a traditional analysis of variance, using an analysis of covariance for each of the three covariates (with a quadratic polynomial of the quality measurements as the covariate), using a spatial analysis of variance, and using a spatial analysis of covariance for each of the three covariates. The analyses were compared based on their values of the corrected Akaike information criteria (AICC).

## RESULTS

Tables 1a–1c show the rejection rates of the *F*-tests for overall equality of means when the response variable is calculated as an average over all 25 observations in a plot. As expected, the rejection rates for the spatial models are higher than the rejection rates for the nonspatial models, and the rejection rates are higher when the covariate is included in the analysis

than when it is ignored. Moreover, note that the increase in power obtained by using an analysis of covariance is more substantial when the correlation between the response and the covariate is strong. As shown in Table 1c, all analyses come close to the nominal 5% level when the null hypothesis of no treatment effect is simulated. Finally, when the correlation between the response and covariate is strong, the rejection rates are higher for the model whose covariate consists of 9 observations than for the model whose covariate consists of a single observation, and they are the highest for the model whose covariate is comprised of all 25 observations in the plot. However, when the correlation between the response and covariate is weak, there is very little difference between the rejection rates for all three covariate analyses (Tables 1a and 1b). In this case, it appears that the support of the covariate does not have a dramatic result on the analysis.

When the response variable is generated from the central observation of each plot, the simulation study yields rejection rates slightly higher than the nominal 5% level when the null hypothesis of no treatment effect is true (Table 2c). Therefore, the rejection rates in Tables 2a and 2b have been adjusted to the 5% rejection rate. For example, the spatial analysis of variance rejects the null hypothesis 8.8% of the time when the null hypothesis is true. The 50 smallest *p* values in this case are all less than .0214; thus, to obtain a true rejection rate of 5% when the null hypothesis is true, a significance level of .0214 is used. Tables 2a and 2b shows the rejection rates of the *F*-tests for overall equality of means (using the adjusted significance levels) when the response variable is generated from the single central observation of each plot. Again, the rejection rates are higher for the spatial models than for the nonspatial models, and the rejections rates are higher when a covariate is used than when it is ignored. Also, as the correlation between the response and the covariate increases, there is a more sizeable increase in the power of the analysis when the covariate is added to the model. Finally, when the correlation between the covariate and response is strong, the rejection rates are highest for the model whose covariate also consists of a single observation. However, when the correlation between the covariate and response is weak, there is hardly any difference between the rejection rates (Tables 2a and 2b). Again, in this case, it appears that the support of the covariate has little effect on the results.

Tables 3a–3c show the AICC for the various analyses of the field experiment. Smaller AICC values indicate a better

**Table 2a. Rejection rates (adjusted to 5% significance level) for analyses with response generated from the center observation in the plot and with treatment effects equally spaced.<sup>†</sup>**

No. of observations in covariate	$\rho = .3$				$\rho = .7$	
	NS/NOCO	S/NOCO	NS/COV	S/COV	NS/COV	S/COV
1	0.235	0.486	0.258	0.523	0.419	0.730
9	–	–	0.268	0.496	0.405	0.641
25	–	–	0.290	0.495	0.387	0.557

**Table 2b. Rejection rates (adjusted to 5% significance level) for analyses with response generated from the center observation in the plot and with treatment effects in maximum–minimum configuration.**

No. of observations in covariate	$\rho = .3$				$\rho = .7$	
	NS/NOCO	S/NOCO	NS/COV	S/COV	NS/COV	S/COV
1	0.259	0.484	0.283	0.516	0.450	0.754
9	–	–	0.308	0.501	0.431	0.650
25	–	–	0.321	0.503	0.409	0.560

**Table 2c. Rejection rates for analyses with response generated from the center observation in the plot and with no treatment effect.**

No. of observations in covariate	$\rho = .3$				$\rho = .7$	
	NS/NOCO	S/NOCO	NS/COV	S/COV	NS/COV	S/COV
1	0.062	0.088	0.063	0.095	0.059	0.082
9	–	–	0.061	0.091	0.056	0.100
25	–	–	0.055	0.090	0.061	0.111

<sup>†</sup>NS/NOCO, nonspatial analysis with no covariate; S/NOCO, spatial analysis with no covariate; NS/COV, nonspatial analysis of covariance; S/COV, spatial analysis of covariance.

fit of the model to the data. When analysis of covariance was utilized in a nonspatial analysis, the case in which the covariate was calculated as the average of all 25 observations in the plot yielded the best-fitting model. Also, the spatial analysis of covariance (when the covariate was calculated from the average of all 25 observations) yielded the best-fitting model when the response was generated from the center observation of the plot. When the response was averaged over all 25 observations in the plot and over the 9 observations in the central  $3 \times 3$  square of the plot, the spatial analysis of covariance (when the covariate was calculated from 9 observations) yielded the best-fitting model.

## DISCUSSION AND CONCLUSIONS

First, and not surprisingly, the results of the simulation study indicate that the use of a model which accounts for spatial structure is superior to the use of a conventional model which ignores this component of the response variable. Also, an analysis which makes use of information from a covariate is more powerful than an analysis which ignores the covariate; moreover, the stronger the correlation between the covariate and the response variable, the more valuable is the information the covariate has to offer. It should be mentioned that a negative correlation between the response and the covariate can exist. The results given in this chapter refer to only positive values for the correlation coefficient since it is the magnitude, not the sign, of the

correlation that is important. Though not reported here, the study arrived at similar conclusions when negative values for  $\rho$  were simulated.

In addition, the results indicate that if a strong relationship exists between a response variable and its covariate, using information from the covariate whose support is nearest to that of the response yields the most precise experimental results. When the response is calculated as the average of all 25 observations in the plot and the correlation between the response and covariate is strong, the most powerful analysis is that which utilizes a covariate also obtained from a block average of the 25 points in the plot. Also, when the central point observation of the plot is recorded as the response and the correlation is strong, the analysis is most powerful when the covariate is also of point support. Finally, though the results are not presented, the authors also performed a simulation study in which the response variable was calculated as a block average of the central 9

observations. Not surprisingly, the most powerful analysis in this case is again that which utilizes information from the covariate whose support is identical to that of the response. On the other hand, if a weak relationship exists between the response variable and its covariate, the effects of changing the support of the covariate do not appear to be as dramatic.

Finally, the model information criteria (AICC) for the various analyses of the field experiment can be used to illustrate the results of the simulation study. First, examine the case where the response is averaged over the 9 observations in the central  $3 \times 3$  square of the plot. When accounting for spatial structure, the model which uses the covariate whose support is identical to that of the response (i.e., when the covariate is also calculated from the 9 observations) provides the best fit. Next, examine the other cases where the response was generated from either the central observation of the plot or the average of all 25 observations in the plot. According to the results of the simulation study, the spatial model which uses the covariate whose support is identical to that of the response is expected to provide the best fit. However, when the response is the center observation of the plot, the AICC is smallest when the support of the covariate is the block average of all 25 observations. Also, when spatial structure is considered and the response is the block average of all 25 observations, the AICC is smallest when the support

of the covariate is the block average of the 9 observations in the central  $3 \times 3$  square of the plot. This can be attributed to the spatial structure of the covariable used in this field experiment. In the simulation study itself, no nugget effect was generated. However, a nugget effect does exist in the field experiment. A geostatistical analysis reveals that fitting a spherical model to the covariate yields a nugget of .0055 and a partial sill of .0028. Note that the nugget is almost twice the size of the partial sill. Because of this inherent variability, it is no surprise that averaging over some of the observations to obtain the covariate leads to a better fitting model than if the covariate is obtained from a single observation. The variability is reduced considerably when averaging over 9 observations as opposed to 1; however, the variability does not decrease as drastically when averaging over 25 observations as opposed to 9. Therefore, when the response is averaged over either 9 or 25 observations, the AICC drops considerably as the model changes from using a single point as the covariate to using a block average of 9 observations. As the model changes from using an average of 9 observations to 25 observations, the AICC actually increase, but the criteria are only slightly different.

Note that the simulation study expresses the results of this experiment on average, while the field experiment gives results for only one data set. If the covariable in this experiment would have alternatively possessed a strong spatial structure with a nugget which was small relative to the partial sill, it is expected that the results of the field experiment would have been in agreement with the simulation study. Similarly, since the nugget for the response variable was relatively large in many cases, the nonspatial analysis sometimes yielded a smaller AICC value than the spatial analysis. If the spatial structure of the response had been strong and the nugget effect for the response relatively small in all cases, it is expected that the field experiment would have been in accord with the simulation results.

In conclusion, this study has shown that methods which exploit both underlying spatial structure and information from a covariate are most efficient, especially when a strong correlation exists between the response and the covariate. Also, this study has addressed questions regarding the use of intensively collected data points as covariates. Based on the results of this simulation, the authors recommend the following:

1. If the correlation between the response and the covariate is strong, then the support of the covariate should be as close as possible to the support of the response variable to obtain the most accurate experimental results.
2. If the correlation between the response and the covariate is weak, then the support of the covariate in relation to the support of the response variable has little effect on the analysis. Since it is known what

**Table 3a. Corrected Akaike information criteria (AICC) for analyses with response generated from the center observation in the plot. For the spatial analyses, the nugget and partial sill for the response variable are given.**

No. of observations in covariate	Nonspatial analysis	Spatial analysis
No covariate effect	30.8	35.0 (nugget = .03148, sill = .07487)
1	33.6	39.0 (nugget = .03087, sill = .07039)
9	31.9	36.0 (nugget = .01129, sill = .08414)
25	21.8	28.1 (nugget = 0.0000, sill = .06021)

**Table 3b. AICC for analyses with response averaged over the 9 observations in the central  $3 \times 3$  square of the plot. For the spatial analyses, the nugget and partial sill for the response variable are given.**

No. of observations in covariate	Nonspatial analysis	Spatial analysis
No covariate effect	5.6	10.8 (nugget = .01715, sill = .02260)
1	9.5	15.6 (nugget = .01625, sill = .02378)
9	2.7	1.6 (nugget = .00192, sill = .03800)
25	-3.9	2.0 (nugget = .00770, sill = .01471)

**Table 3c. AICC for analyses with response averaged over all 25 observations in the plot. For the spatial analyses, the nugget and partial sill for the response variable are given.**

No. of observations in covariate	Nonspatial analysis	Spatial analysis
No covariate effect	-0.6	0.3 (nugget = .00708, sill = .02441)
1	3.3	4.7 (nugget = .00651, sill = .02538)
9	0.4	-4.4 (nugget = .00000, sill = .02723)
25	-5.3	-2.9 (nugget = .00192, sill = .01939)

support is best for the covariate when the correlation is strong and that it will make little difference in the case where the correlation is weak, the researcher should once again use a covariate whose support is as close as possible to that of the response variable.

3. If there is little or no spatial structure in the covariate, then using the average of all georeferenced values which are contained in the support of the response variable is optimal.

## References

- Bryant, R., D. Thoma, S. Moran, C. Holifield, D. Goodrich, T. Keefer, G. Page, D. Williams, and S. Skirvin. 2003. Evaluation of hyperspectral, infrared temperature and radar measurements for monitoring surface soil moisture. First Interagency Conference on Research in the Watersheds, Benson, Arizona. 27–30 Oct. 2003. Available at [www.tucson.ars.ag.gov/icrw/Proceedings/Bryant.pdf](http://www.tucson.ars.ag.gov/icrw/Proceedings/Bryant.pdf) (verified 23 Dec. 2006).
- Clark, I. 1979. Practical geostatistics. Applied Science Publishers, London.
- Cressie, N.A.C. 1993. Statistics for spatial data. John Wiley and Sons, New York.
- Frazier, P., D. Lamb, and L. Kumar. 2004. Remote sensing and satellite technology— implementation and capabilities. Australian Fertilizer Industry Conference, Stradbroke Island. 2–5 Aug. 2004. Stradbroke Island, Queensland, Australia.

- Isaaks, E.H., and R.M. Srivastava. 1989. An introduction to applied geostatistics. Oxford Univ. Press, New York.
- Journel, A.G., and Ch.J. Huijbregts. 1978. Mining geostatistics. Academic Press, New York.
- Littell, R.C., G.A. Milliken, W.W. Stroup, and R.D. Wolfinger. 1996. SAS System for Mixed Models. SAS Institute, Cary, NC.
- Martin, R.J. 1986. On the design of experiments under spatial correlation. *Biometrika* 73:247–277.
- McGuire, J. 2003. Precision agriculture—beyond changing fertilizer rates. Ohio Geospatial Technology Conference for Agriculture and Natural Resources, Columbus, Ohio. 24–26 Mar. 2003. Available at [http://geospatial.osu.edu/conference/proceedings/papers/irwin\\_pap.pdf](http://geospatial.osu.edu/conference/proceedings/papers/irwin_pap.pdf) (verified 23 Dec. 2006).
- Olea, R.A. 1991. Geostatistical glossary and multilingual dictionary. Oxford Univ. Press, New York.
- Oliver, D.S. 2003. Gaussian cosimulation: Modelling of the cross-covariance. *Math. Geol.* 35:681–698.
- SAS Institute. 2003. SAS onlinedoc, version 9.1. SAS Institute, Cary, NC.
- Schabenberger, O., and C.A. Gotway. 2005. Statistical methods for spatial data analysis. Chapman and Hall/CRC Press, Boca Raton, FL.
- Searle, S.R. 1971. Linear models. John Wiley & Sons, New York.
- Servilla, M. 1998. The first steps to understanding agriculture remote sensing. *Mod. Agric.* 1:20–22.
- Wehrspann, J. 2000. Harvest by test weight. *Farm Industry News*. Available at [http://farministrynews.com/mag/farming\\_harvest\\_test\\_weight/](http://farministrynews.com/mag/farming_harvest_test_weight/) (verified 23 Dec. 2006).