

SPATIAL ANALYSIS OF NDVI READINGS WITH DIFFERENT SAMPLING DENSITIES

H. Zhang, Y. Lan, R. Lacey, W. C. Hoffmann, J. K. Westbrook

ABSTRACT. *Advanced remote sensing technologies provide researchers an innovative way to collect spatial data in precision agriculture. Sensor information and spatial analysis together allow for a detailed understanding of the spatial complexity of a field and its crop. The objective of the study was to describe field variability in the normalized difference vegetation index (NDVI) and characterize the spatial structure of NDVI data by geostatistical variogram analysis. Data sets at three different sampling densities were investigated and compared to characterize NDVI variation within the specified study area. Variograms were computed by Matheron's method of moments (MoM) estimator and fitted by theoretical models. The fitted spherical model was determined to be the best model for the data analysis in the study. The range of spatial dependence of the NDVI data was 40 m for a sampling area of 4 m × 3 m. Knowing the amount of remotely sensed data needed to characterize the spatial variation of the field with NDVI allows us to save sampling costs and prescribe site-specific nitrogen and other agrichemical applications.*

Keywords. *NDVI, Remote sensing, Spatial statistics, Variogram.*

Advanced remote sensing technologies provide researchers an innovative way to collect spatial data in precision agriculture. Many commercially available sensors or optical instruments provide the capability of acquiring real-time spectral information from vegetation. Studies have suggested that crop spectral reflectance can be used to assess plant nutrient and pigment status (Goel et al., 2003a; Osborne et al., 2002), monitor plant conditions at various scales (Blackmer et al., 1994; Plant, 2001), and crop biophysical variables (Thenkabail et al., 2000; Goel et al., 2003b).

Canopy spectral reflectance properties based on the absorption of light at a specific wavelength are associated with specific plant characteristics. The spectral reflectance in the visible wavelengths (400-700 nm) is low because of the high absorption of light energy by chlorophyll. The reflectance in the near-infrared (NIR) wavelengths (700-1300 nm) is high because of the multiple scattering of light by different leaf tissues (Taiz and Zeiger, 2006). Vegetation indices have been developed with the reflectance data from red and NIR wave-

lengths and are often used to monitor crop growth conditions. The normalized difference vegetation index (NDVI) is a good indicator of vegetation, crop biomass, and health in agricultural applications (Rouse et al., 1973; Tucker, 1979). NDVI is calculated as: $NDVI = (NIR - Red) / (NIR + Red)$, where Red and NIR stand for the spectral reflectance measurements acquired in the red and near-infrared regions, respectively. Healthier crop canopies will absorb more red and reflect more near-infrared light than stressed or unhealthy canopies, and consequently have a higher NDVI value. Sembring et al. (1998) found that NDVI was a good indicator of nitrogen (N) uptake of winter wheat. Freeman et al. (2007) collected NDVI with Greenseeker (NTech Industries, Inc., Ukiah, Cal.) handheld sensors and plant height measurements on individual corn plants at various growth stages and related them to individual plant biomass, forage yield, and N uptake. Bronson et al. (2005) used NDVI collected from different sensors to give a better estimation of in-season plant N status.

Remotely sensed data and spatial analysis together allow for a detailed understanding of the spatial complexity of a field and its crop. Determination of the spatial variability of field parameters is usually based on the concept that sampled values at nearby locations are more similar than those farther apart. Measurements from the field are usually gathered as point data, such as samples from an individual plant. Spatial analysis methods can be used to interpolate measurements to create a continuous surface map or to describe its spatial pattern (Cressie, 1993). As a powerful tool in geostatistics, variograms (also referred to as semivariograms) characterize the spatial dependence of data and give the range of spatial correlation, within which the values are correlated with each other and beyond which they become independent. The parameters of the best fitted model for a variogram can be used for kriging (Matheron, 1963; Stein and Corsten, 1991). Kriging has been recommended as the best method to interpolate point data since it minimizes the error variance using a

Submitted for review in December 2009 as manuscript number IET 8353; approved for publication by the Information & Electrical Technologies Division of ASABE in January 2011. Presented at the 2009 ASABE Annual Meeting as Paper No. 095982.

The authors are **Huihui Zhang, ASABE Member Engineer**, Graduate Student, Department of Biological and Agricultural Engineering, Texas A&M University, College Station, Texas; **Yubin Lan, ASABE Member Engineer**, Agricultural Engineer, USDA-ARS Areawide Pest Management Research Unit, College Station, Texas; **Ronald Lacey, ASABE Member Engineer**, Professor, Department of Biological and Agricultural Engineering, Texas A&M University, College Station, Texas; **W. Clint Hoffmann, ASABE Member Engineer**, Agricultural Engineer, and **John K. Westbrook**, Research Meteorologist, USDA-ARS Areawide Pest Management Research Unit, College Station, Texas. **Corresponding author:** Yubin Lan, USDA-ARS Areawide Pest Management Research Unit, 2771 F&B Road, College Station, Texas 77845; phone: 979-260-3759; fax: 979-260-9386; e-mail: yubin.lan@ars.usda.gov.

weighted linear combination of the data (Panagopoulos et al., 2006). There are also numerous studies demonstrating the benefits of geostatistical analysis techniques to agricultural management. Heisel et al. (1996) used kriging to map the density of weeds in winter wheat. Stewart et al. (2002) used geostatistical methods to interpolate data and produce maps of a field representing the spatial variability of all the soil and wheat properties. With the aid of these maps and empirical modeling techniques, relationships between the wheat and soil factors were determined. Yamagishi et al. (2003) investigated the spatial variability of crop biomass and determined if site-specific management could be applied to a small field by using a variogram.

The large amount of remotely sensed data also could increase the sampling costs, provide redundant information, and require complicated data analysis techniques. The issue has drawn considerable attention to specify the sampling requirements needed to accurately analyze the spatial property of an object. The objectives of this study were to describe the variability of a soybean field in NDVI, characterize the spatial structure of NDVI with different sampling data sets using variogram analysis, and determine an optimum sampling size that could adequately describe the field variation in canopy NDVI for future studies.

MATERIALS AND METHODS

STUDY SITE

The study site consisted of a 15 m × 65 m area within an approximately 1 ha soybean field near College Station, Texas (30.37055° N, 96.21610° W). The soybeans (variety HBK C5025, Hornbeck Seed Co., Dewitt, Ark.) were planted on 31 March 2008 with a row spacing of 1 m and with the rows oriented in the east-west direction. Nitrogen was applied as ammonium sulfate (336 kg ha⁻¹) broadcast prior to planting and incorporated into the beds. The field was irrigated weekly as needed during the pod fill period.

SAMPLING DESIGN AND DATA COLLECTION

At the end of June 2008, the plants on the north side of plot reached senescence, while the other plants were growing vigorously. To assess the spatial variation of soybean plants within this area, spectral reflectance measurements were conducted along four selected transects with a FieldSpec handheld hyperspectroradiometer (Analytical Spectral Devices, Inc., Boulder, Colo.). The distance between selected transects was 3 m. The sampling position along each transect

was marked by a 2 m interval, and the spatial coordinates were recorded using a GPS receiver (2 m × 3 m sampling area). Each transect comprised 32 sampling points for a total of 128 observations (fig. 1).

The FieldSpec handheld hyperspectroradiometer was positioned with a nadir view from a height of about 2 m above the ground. With an angular field of view of 25°, it scanned approximately 0.62 m² of field area. The spectroradiometer collected data from the canopy in a wavelength range from 325 nm to 1075 nm with a sampling interval of 1.6 nm. The spectroradiometer output 512 continuous data points with each reading. A sunny day was chosen for the field test, and all data were collected around solar noon. Instrument optimization and white reference measurements were performed prior to sample measurements using RS³ software (Analytical Spectral Devices, Inc., Boulder, Colo.). The spectroradiometer was adjusted to ten scans per dark current, and the integration time was set at 217 ms. All reflectance measurements taken from each transect were completed within half an hour. The reflectance values at 680 nm in the red region and at 800 nm in the NIR region were chosen to calculate NDVI (Castro-Esau et al., 2006).

STATISTICAL ANALYSIS

Descriptive statistics for NDVI values were calculated and the outliers and anomalies were examined using R (ver. 2.8.0, The R Foundation for Statistical Computing, Vienna, Austria). Autocorrelation analysis was applied to each transect. The spatial structure of the NDVI readings was determined using geostatistical techniques and variogram analysis.

Variograms

Variograms were computed for three data sets with different sampling densities. The first data set was all NDVI data at a 2 m sampling interval (2 m × 3 m sampling area), the second data set was reduced to sampling points at a 4 m interval (4 m × 3 m sampling area), and the third data set was reduced to sampling points at a 6 m interval (6 m × 3 m sampling area). This resulted in 128, 64, and 40 measurements being used for the 2, 4, and 6 m spatial analyses, respectively. The procedures for detecting trend and anisotropy were performed.

The experimental variograms were computed using Matheron's method of moments (MoM) estimator (Matheron, 1965). The estimator is given by the following equation:

$$\gamma(h) = \frac{1}{2m(h)} \sum_{i=1}^{m(h)} [z(x_i) - z(x_i + h)]^2 \quad (1)$$

where $\gamma(h)$ is an unbiased estimate of the variance of the $m(h)$ pairs of NDVI readings; $m(h)$ is the number of sampling pairs separated by a lag h for $i = 1, 2, \dots, m(h)$; and $z(x_i)$ and $z(x_i + h)$ are the NDVI values at locations x_i and $(x_i + h)$, respectively.

Theoretical Models

The experimental variograms were fitted (based on a weighted least squares approximation) with theoretical models that provided three key parameters: the nugget variance, the sill variance, and the range of spatial dependence. These model parameters described the spatial structure of the NDVI readings. The sample sizes of 128 and 64 in this study satis-

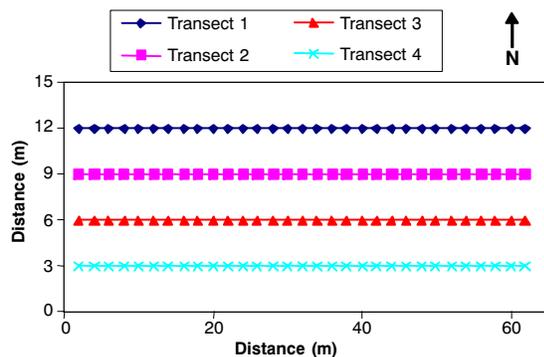


Figure 1. Locations of sampling points in the study area.

fied the requirement of acquiring reliable estimation of a variogram by MoM (Webster and Oliver 1992; Kerry and Oliver 2007). Therefore, the spherical and exponential models were fitted to the variograms computed from the 2 m interval and 4 m interval data sets. For the 6 m interval data set, the variogram was estimated by the maximum likelihood (ML) approach and compared to those estimated by MoM (Lark, 2000). On the basis of the least sum of squares or Akaike information criterion (AIC; Akaike, 1973), a good fit model was chosen. Given a data set, several models may be ordered according to their AIC. The one with the lowest AIC would be the best. The parameters of the model were used for kriging, which is a method of interpolation to predict unknown values from data observed at known locations.

The spherical model is one of the most commonly used models for experimental data (Webster and Oliver, 2007) and is expressed as:

$$\gamma(h) = \begin{cases} c_0 & \text{when } h = \varepsilon \text{ (very small lag)} \\ c_0 + c \left(\frac{3h}{2a} - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right) & \text{when } 0 < h \leq a \\ c_0 + c & \text{when } h > a \end{cases} \quad (2)$$

where c_0 is the nugget variance, c is the sill, h is the lag, and a is the range.

The exponential model has been used commonly because of its generality. In the isotropic case, it is given by the following equation:

$$\gamma(h) = \begin{cases} c_0 + c[1 - \exp(h/r)] & \text{for } h > 0 \\ 0 & \text{for } h = 0 \end{cases} \quad (3)$$

The non-linear parameter r defines the spatial scale of the variation. The sill is approached asymptotically. For practical purposes, $a = 3r$ is regarded as the effective range of the exponential model, which is the lag at which the sill reaches approximately $c_0 + 0.95c$ (Webster and Oliver, 2007).

In all, the procedure for modeling a variogram involves both visual inspection and statistical fitting: first plot the variogram, then choose one or more models with the right shape to represent the major trends in the data, next fit each model in turn by weighted least squares, and finally inspect the result by plotting the fitted models on the variogram. Among the plausible models, the one with smallest mean square or smallest residual sum of square will be chosen. A smaller nugget (error) and lower nugget-to-sill ratio would also be considered.

The variogram analyses, experimental variogram computing, model fitting, and kriging were performed with the geOR package in R software.

RESULTS AND DISCUSSION

DESCRIPTIVE STATISTICS

The NDVI data for the four transects are plotted in figure 2. The NDVI tendency of each transect was different. The lowest NDVI value was found in transect 2, where the plants were senescent and very dry. The descriptive statistics of the

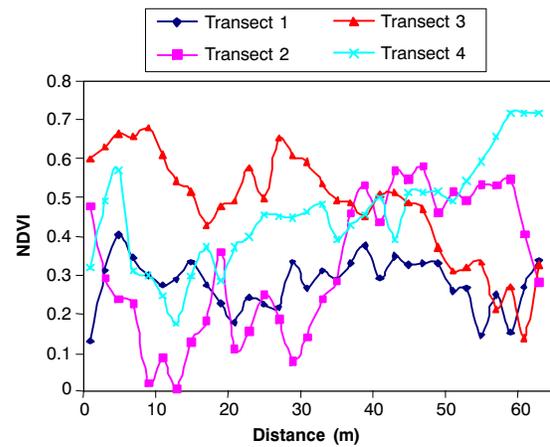


Figure 2. NDVI data for the four transects.

NDVI data at the three sampling intervals were calculated (table 1). The means, medians, and standard deviations of the three data sets were similar. Student t-tests were performed for the three data sets, and there was no significant difference among means. In other words, the decrease in the sampling density did not affect the properties of the NDVI data. No transformation of the data was necessary for geostatistical analysis.

GEOSTATISTICAL ANALYSIS

The existence of anisotropy was assessed first. Anisotropy was tested in four directions (0° , 45° , 90° , and 135°). The direction of the maximum continuity was found along the transects, and the direction of the minimum continuity was perpendicular to the transects, as there were more data points along a transect than perpendicular to a transect. The calculation of semivariance was restricted to the direction of the transects only. The NDVI variograms computed for the 2 m interval, 4 m interval, and 6 m interval data sets are shown in figure 3. The shapes of the variograms were a little wavelike, which indicated the periodicity since the distance of the transects remained constant. There were no evident differences in shape and semivariance magnitude among sample densities. The variability increased while the lag distance increased until about 50 m.

Exponential and spherical models were fitted to the variogram computed for the 2 m interval and 4 m interval data sets (figs. 4 and 5). Figure 6 indicates that the variogram computed for the 6 m interval data set was fitted with exponential and spherical models estimated by weighted least squares and maximum likelihood.

The variogram parameters are summarized in table 2. For the 2 m interval data set, the sum of squares of the exponential model and spherical model were the same, but the exponen-

Table 1. Descriptive statistics of NDVI for the three data sets in the study area.

Parameter	Data Set		
	2 m	4 m	6 m
Mean	0.3850	0.3810	0.3780
Median	0.3721	0.3876	0.3653
SD	0.1601	0.1640	0.1589
Skewness	-0.0198	-0.19	-0.097
CV (%)	41.59	43.08	42.04
Count	128	64	44

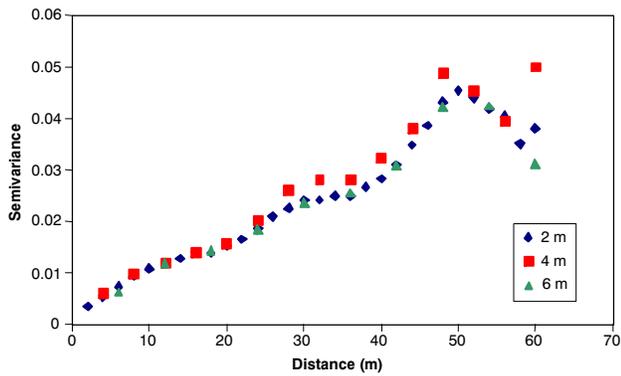


Figure 3. Variograms of three NDVI data sets in the study field for 2 m, 4 m, and 6 m intervals.

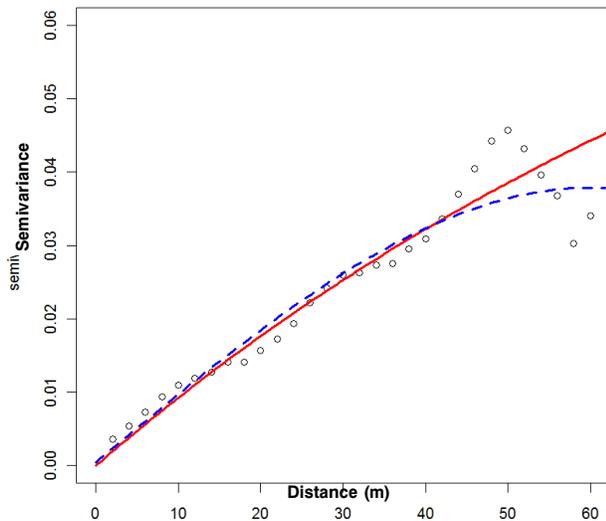


Figure 4. Variogram of NDVI data set (2 m interval) in the study field: experimental variogram calculated by method of moments estimator (circles), and exponential (solid line) and spherical (dashed line) models fitted by weighted least squares.

tial model had a smaller nugget and nugget-to-sill ratio. For the 4 m interval data set, both the exponential and spherical models had zero nuggets. The sum of squares of the spherical model was smaller than that of the exponential model. Moreover, the range of the exponential model was larger than the length of the transect. The range of the spherical model was 40 m, beyond which the NDVI became independent. Although maximum likelihood is suggested for calculation of a variogram when the data size is less than 50, both models fitted for the 6 m interval data set had high nugget-to-sill ratios in this case. Also from figure 6, the maximum likelihood models were far away from the variogram and could not give a better fit or more information than the other two models estimated using MoM. Overall, the spherical model for the 4 m interval dataset had a range of 40 m, zero nugget and nugget-to-sill ratio, the smallest sum of squares, and good fit with visual inspection (fig. 5). The parameters of the spherical model for the 4 m interval data set were used for kriging and predicting NDVI values at unsampled locations.

With 64 known sampling points and the parameters of the chosen model, kriging was performed to produce the NDVI data map of the study area (fig. 7). This map describes the spatial variation of NDVI within the study field in a better

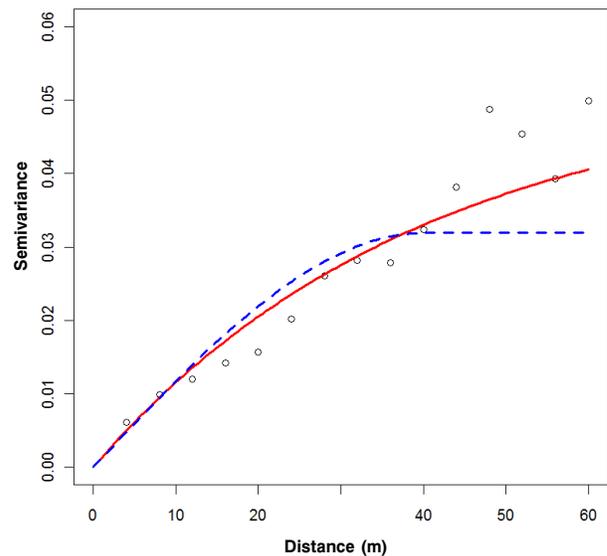


Figure 5. Variogram of NDVI data set (4 m interval) in the study field: experimental variogram calculated by method of moments estimator (circles), and exponential (solid line) and spherical (dashed line) models fitted by weighted least squares.

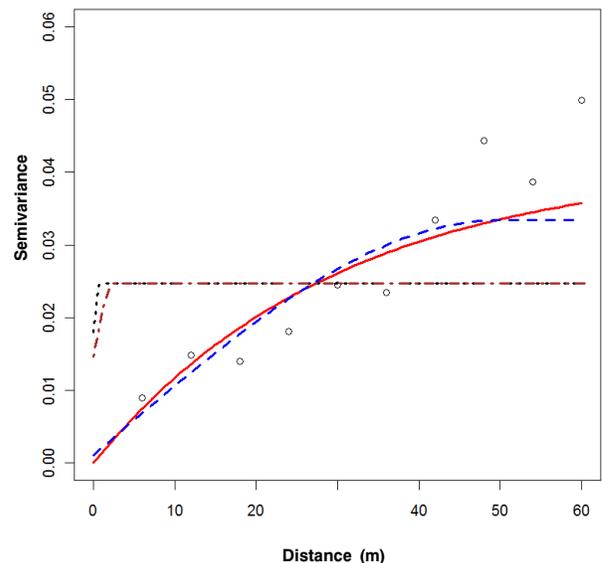


Figure 6. Variogram of NDVI data set (6 m interval) in the study field: experimental variogram calculated by method of moments estimator (circles), exponential (solid line) and spherical (dashed line) models fitted by weighted least squares, and exponential (dotted line) and spherical (dot-dash line) models fitted by maximum likelihood.

way, adds more information, and provides better understanding than classical descriptive statistics analysis. Within the 15 m × 65 m area, the NDVI values were very low on the northwest side, where the soybean plants were yellow, dry, and had stopped growing. The variability perpendicular to the transects was large.

CONCLUSIONS

This study revealed that the remotely sensed normalized difference vegetation index (NDVI) was suitable to describe crop ground cover and crop growing status. The NDVI data, analyzed by a geostatistical method, variogram, and kriging,

Table 2. Parameters of the exponential and spherical models fitted to the experimental variogram estimated by method of moments (MoM) ($n = 128$) and maximum likelihood (ML; $n = 64$) that describe the spatial structure of NDVI in the study field.

Interval	Model	Range (m)	Nugget	Sill	Nugget % ^[a]	Sum of Squares	AIC
2 m	Exponential	209	0.0003	0.0728	0.4	0.0108	
	Spherical	60	0.0005	0.0379	1.3	0.0108	
4 m	Exponential	119	0	0.0522	0	0.0061	
	Spherical	40	0	0.0319	0	0.0036	
6 m	Exponential	89	0	0.0413	0	0.0047	
	Spherical	50	0.0011	0.0434	25.34	0.0051	
	ML/exponential	0.15	0.018	0.0247	72.87		-30.01
	ML/spherical	2.5	0.0147	0.0247	59.51		-30

^[a] Percentage nugget is calculated as $\text{Nugget/Sill} \times 100$.

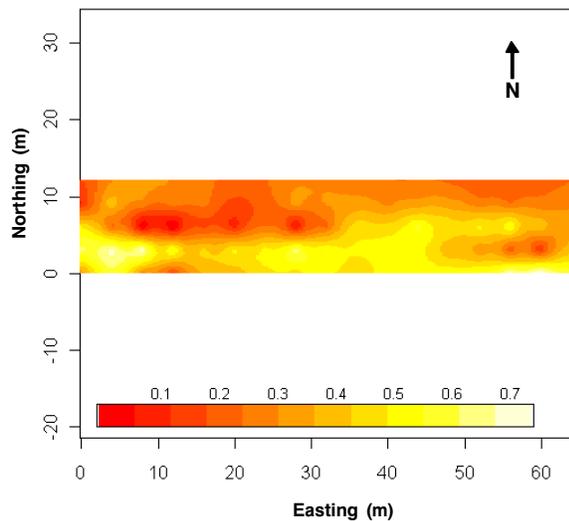


Figure 7. NDVI map of the study field.

gave a good description of the spatial variation within the field. In the study, the spatial dependence of the NDVI data was 40 m with a sampling area of $4 \text{ m} \times 3 \text{ m}$. Although it is possible to increase the sampling interval to 6 m without lost spatial information, the parameters of the fitted model are not accurate enough for kriging. Compared to a sampling interval of 2 m, the use of the 4 m interval data set reduces the processing of redundant data without affecting the quality of the variation described. Knowing the amount of remotely sensed data needed to characterize the spatial variation of the field with NDVI allows us to save sampling costs and prescribe nitrogen and other agrichemical applications.

The study only considered three sampling densities for this field. The distance between transects was consistent. Different sampling area and spatial structure of other remotely sensed data may be examined in a future study.

ACKNOWLEDGEMENTS

The authors express their sincere gratitude to Mr. Mike O'Neil for helping with field test.

REFERENCES

Akaike, H. 1973. Information theory and an extension of maximum likelihood principle. In *Proc. 2nd Intl. Symp. on Information Theory*, 267-281. B. N. Petrov and F. Coaki, eds. Budapest, Hungary: Akademia Kiado.

- Blackmer, T. M., J. S. Schepers, and G. E. Varvel. 1994. Light reflectance compared with other nitrogen stress measurements in corn leaves. *Agron. J.* 86(6): 934-938.
- Bronson, K. F., J. D. Booker, W. J. Keeling, R. K Boman, T. A. Wheeler, R. J. Lascano, and R. L. Nichols. 2005. Cotton canopy reflectance at landscape scale as affected by nitrogen fertilization. *Agron. J.* 97(3): 654-660.
- Castro-Esau, K. L., G. A. Sanchez-Azofeifa, and B. Rivard. 2006. Comparison of spectral indices obtained using multiple spectroradiometers. *Remote Sensing Environ.* 103(3): 276-288.
- Cressie, N. A. C. 1993. *Statistics for Spatial Data*. New York, N.Y.: John Wiley and Sons.
- Freeman, K. W., K. Girma, D. B. Arnall, R. W. Mullen, K. L. Martin, R. K. Teal, and W.R. Raun. 2007. By-plant prediction of corn forage biomass and nitrogen uptake at various growth stages using remote sensing and plant height. *Agron. J.* 99(2): 530-536.
- Goel, P. K., S. O. Prasher, J. A. Landry, R. M. Patel, R. B. Bonnell, A. A. Viau, and J. R. Miller. 2003a. Potential of airborne hyperspectral remote sensing to detect nitrogen deficiency and weed infestation in corn. *Computers and Electronics in Agric.* 38(2): 99-114.
- Goel, P. K., S. O. Prasher, J. A. Landry, R. M. Patel, A. A. Viau, and J. R. Miller. 2003b. Estimation of crop biophysical parameters through airborne and field hyperspectral remote sensing. *Trans. ASAE* 46(4): 1235-1246.
- Heisel, T., C. Andersen, and A. K. Ersboll. 1996. Annual weed distributions can be mapped with kriging. *Weed Res.* 36(4): 325-337.
- Kerry, R., and M. A. Oliver. 2007. Comparing sampling needs for variograms of soil properties computed by the methods of moments and residual maximum likelihood. *Geoderma* 140(4): 383-396.
- Lark, R. M. 2000. Estimating variograms of soil properties by the method-of-moments and maximum likelihood. *European J. Soil Sci.* 51(4): 717-728.
- Matheron, G. 1963. Principles of geostatistics. *Economic Geology* 58(8): 1246-1266.
- Matheron, G. 1965. Les variables régionalisées et leur estimation: Une application de la théorie defonctions aléatoires aux sciences de la nature. Paris, France: Masson et Cie.
- Osborne, S. L., J. S. Schepers, D. D. Francis, and M. R. Schlemmer. 2002. Detection of phosphorus and nitrogen deficiencies in corn using spectral radiance measurements. *Agron. J.* 94(6): 1215-1221.
- Panagopoulos, T., J. Jesus, M. Antunes, and J. Beltrao. 2006. Analysis of spatial interpolation for optimising management of a salinized field cultivated with lettuce. *European J. Agron.* 24(1): 1-10.
- Plant, R. E. 2001. Site-specific management: The application of information technology to crop production. *Computers and Electronics in Agric.* 30(1-3): 9-29.
- Rouse, J. W., R. H. Haas, J. A. Schell, and D. W. Deering. 1973. Monitoring vegetation systems in the Great Plains with ERTS.

- In *Proc. 3rd ERTS Symp.*, 309-317. NASA Special Publication SP-351, volume I. Washington, D.C.: NASA.
- Sembiring, H., W. R. Raun, G. V. Johnson, M. L. Stone, J. B. Solie, and S. B. Phillips. 1998. Detection of nitrogen and phosphorus nutrient status in winter wheat using spectral radiance. *J. Plant Nutrition* 21(6): 1207-1232.
- Stein, A., and L. C. A. Corsten. 1991. Universal kriging and cokriging as a regression procedure. *Biometrics* 47(2): 575-588.
- Stewart, C. M., A. B. McBratney, and J. H. Skerritt. 2002. Site-specific durum wheat quality and its relationship to soil properties in a single field in northern New South Wales. *Precision Agric.* 3(2): 155-168.
- Taiz, L., and E. Zeiger. 2006. *Plant Physiology*. 4th ed. Sunderland, Mass.: Sinauer Associates, Inc.
- Thenkabail, P. S., R. B. Smith, and E. De Pauw. 2000. Hyperspectral vegetation indices for determining agricultural crop characteristics. *Remote Sensing Environ.* 71(2): 158-182.
- Tucker, C. J. 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing Environ.* 8(2): 127-150.
- Webster, R., and M. A. Oliver. 1992. Sample adequately to estimate variograms of soil properties. *J. Soil Sci.* 43(1): 177-192.
- Webster, R., and M. A. Oliver. 2007. *Geostatistics for Environmental Scientists*. 2nd ed. Chichester, U.K.: John Wiley and Sons.
- Yamagishi, J., T. Nakamoto, and W. Richner. 2003. Stability of spatial variability of wheat and maize biomass in a small field managed under two contrasting tillage systems over 3 years. *Field Crops Res.* 81(2-3): 95-108.