Spatial Modeling and Variability Analysis for Modeling and Prediction of Soil and Crop Canopy Coverage Using Multispectral Imagery from an Airborne Remote Sensing System

Y. Huang, Y. Lan, Y. Ge, W. C. Hoffmann, S. J. Thomson

Abstract. *Spatial modeling and variability analysis of soil and crop canopy coverage has been accomplished using aerial multispectral images. Multispectral imagery was acquired using an MS-4100 multispectral camera at different flight altitudes over a 40 ha cotton field. After the acquired images were geo-registered and processed, spatial relationships between the aerial images and ground-based soil conductivity and NDVI (normalized difference vegetation index) measurements were estimated and compared using two spatial analysis approaches (model-driven spatial regression and data-driven geostatistics) and one non-spatial approach (multiple linear regression). Comparison of the three approaches indicated that OLS (ordinary least squares) solutions from multiple linear regression models performed worst in modeling ground-based soil conductivity and NDVI with high AIC (Akaike information criterion) (-668.3 to 2980) and BIC (Bayesian information criterion) (-642.4 to 3006) values. Spatial regression and geostatistics performed much better in modeling soil conductivity, with low AIC (2698 to 2820) and BIC (2732 to 2850) values. For modeling ground-based NDVI, the AIC and BIC values were -681.7 and -652.1, respectively, for spatial error regression and -679.8 and -646.5, respectively, for geostatistics, which were only moderate improvements over OLS (-668.3 and -642.4). Validation of the geostatistical models indicated that they could predict soil conductivity much better than the corresponding multiple linear regression models, with lower RMSE (root mean squared error) values (0.096 to 0.186, compared to 0.146 to 0.306). Results indicated that the aerial images could be used for spatial modeling and prediction, and they were informative for spatial prediction of ground soil and canopy coverage variability. The methods used for this study could help deliver baseline data for crop monitoring with remote sensing and establish a procedure for general crop management.*

Keywords. *Aerial multispectral imagery, Crop management, Geostatistics, Image processing, Spatial regression.*

Remote sensing technology has the potential to maximize the economic benefit and minimize environmental pollution through improved crop monitoring and management. Airborne remote sensing offers flexibility for rapid image acquisition with high spatial resolution at different flight altitudes. In agricultural research and applications, multispectral imaging systems mounted on aircraft are cost-effective in providing an important data source of crop, soil, or ground cover information (Moran et al., 1997; Senay et al., 1998; GopalaPillai and Tian, 1999; Yang and Anderson, 1999; Medlin et al., 2000; Yang and Everitt, 2002; Pinter et al., 2003; Dobermann and Ping, 2004; Ye et al., 2007; Huang et al., 2008; Inman et al., 2008; Yang et al., 2009).

Conventional statistical techniques often cannot be effectively implemented for spatial data analysis because the techniques do not consider spatial connections or trends between data points (typically neighborhood points). Development of spatial statistics began with mapping, surveying, and geography (Getis et al., 2004). The techniques of spatial statistics have been formalized and developed since the 1950s (Cressie, 1993). With the development of geographic information system (GIS) technology, spatial statistical techniques have drawn considerable attention and have been widely applied in spatial data modeling and analysis for social sciences such as geography, sociology, demography, and economics and for natural sciences such as geophysics, biology, epidemiology, and agriculture. Spatial statistical techniques have promoted the development and applications of statistics, GIS, remote sensing, and scientific computing as well.

Submitted for review in September 2009 as manuscript number IET 8228; approved for publication by the Information & Electrical Technologies Division of ASABE in August 2010.

Mention of trademark, vendor, or proprietary product does not constitute a guarantee or warranty of the product by the USDA and does not imply its approval to the exclusion of other products that may also be suitable.

The authors are **Yanbo Huang, ASABE Member Engineer,** Agricultural Engineer, USDA-ARS Crop Production Systems Research Unit, Stoneville, Mississippi; **Yubin Lan, ASABE Member Engineer,** Agricultural Engineer, USDA-ARS Areawide Pest Management Research Unit, College Station, Texas; **Yufeng Ge, ASABE Member Engineer,** Assistant Research Engineer, Department of Crop and Soil Sciences, Texas A&M University, College Station, Texas; **Wesley C. Hoffmann, ASABE Member Engineer,** Agricultural Engineer, USDA-ARS Areawide Pest Management Research Unit, College Station, Texas; and **Steven J. Thomson, ASABE Member Engineer,** Agricultural Engineer, USDA-ARS Crop Production Systems Research Unit, Stoneville, Mississippi. **Corresponding author:** Yanbo Huang, USDA-ARS Crop Production Systems Research Unit, 141 Experiment Station Rd., Stoneville, MS 38776; phone: 662-686-5354; fax: 662-686-5372; e-mail: yanbo.huang@ars.usda.gov.

Spatial data have been analyzed by using both spatial econometrics and geostatistics (Calderon, 2009). Although both methods solve problems in spatial data modeling and analysis using different statistical approaches to quantify two-dimensional and two-directional data dependence and heterogeneity in space, the comparison is not straightforward. Based on the works of Haining (1990) and Anselin (1988), most of the work in spatial data modeling and analysis can be referred to as either the model-driven approach or the data-driven approach (Anselin, 1990). The model-driven approach is employed by spatial econometrics with spatial regression analysis, starting with a model structure specification that is then fitted with the data. Methods in this category deal with model parameter estimation and model structure specification diagnostics in spatial models (Paelinck and Klaassen, 1979; Anselin, 1988; Griffith, 1992). The data-driven approach in geostatistics uses variogram and kriging methods, and assumes randomness in data distribution (i.e., the null hypothesis) based on a normal or randomization approach (Schabenberger and Pierce, 2002). The spatial pattern, spatial structure, and spatial interaction are derived from the data only, without constraints of a pre-conceived theoretical specification. For applications, the model-driven approach mainly deals with spatial modeling related to regional and urban economics, and the data-driven approach focuses on studies of issues in geophysics, biology, and agriculture.

In agricultural crop management, few studies to date have used the model-driven approach, instead relying primarily on the data-driven approach. Anselin et al. (2004) presented a case study using data from Argentina to model the spatial structure of yield data for corn nitrogen response. The overall objective of their study was to determine the feasibility of using spatial econometric analysis on yield monitor data from a combine to estimate site-specific crop response. Results indicated nitrogen response differences by landscape position, and that site-specific application could be profitable if indicated by spatial modeling. Non-spatial models did not necessarily indicate profitability. Their study indicated that some type of spatial analysis should be implemented if spatial data are to be used for estimation of site-specific response. Lambert et al. (2004) compared four approaches that incorporated spatial correlation into regression models: spatial econometrics, polynomial trend regression, classical nearest neighbor analysis, and geostatistics. With the data studied, spatial econometrics, geostatistics, and spatial trend analysis offered stronger statistical evidence of spatial heterogeneity of corn yield response to nitrogen than the ordinary least squares (OLS) or nearest neighbor analysis. Like the study by Anselin et al. (2004), all spatial models indicated potential profitability of using variably applied nitrogen.

Geostatistics (variogram modeling of regression residuals) is a widely employed approach to model datasets that exhibit spatial correlation in agriculture. This approach is especially well established in the soil science literature, where sample data are point-based and sparse (Odeh et al., 1994; Knotters et al., 1995; Odeh et al., 1995; Hengl et al., 2004; Ge et al., 2007a, 2007b). Parameters for the variogram of regression residuals can be estimated by either an empirical general least squares (EGLS) fitting procedure (Hengl et al., 2004; Ge et al., 2007a, 2007b) or residual maximum likelihood (REML) (Lark, 2000; Kerry and Oliver, 2007). Compared with EGLS, REML has potential advantages for spatial modeling. First, it can estimate both regression coefficients and residual covariance structure simultaneously. Second, it performs better than the method of moment estimator (which is employed for EGLS) for small sample sizes (Kerry and Oliver, 2007).

Geostatistics also provides an interpolation technique (kriging) that can predict response variables at unsampled locations. Yao et al. (2003) used a co-located cokriging estimator to investigate soil nutrient mapping with soil sampling data and an aerial hyperspectral image. A single hyperspectral image band was selected for each soil nutrient factor based on the correlation between the nutrient factor and the image band. When compared with regression analysis, results from co-located cokriging showed better correlations and accuracy with smaller RMSE (root mean squared error). Misaghi et al. (2004) developed a model to forecast strawberry yield using several different artificial neural network techniques. Data collected included aerial imagery, soil parameters, and plant parameters from a 0.8 ha strawberry field. Geostatistical techniques were used to produce more input by interpolation for training and testing the models. Three network models (multilayer perceptrons, generalized feed-forward, and modular neural networks) showed the best results for yield prediction vs. actual yield. Bajwa and Mozaffari (2007) tested various spatial models for analyzing variations in green normalized difference vegetative index (GNDVI), a vegetation index derived from aerial remote sensing data utilizing green and near-infrared wavelengths in response to nitrogen treatments and petiole nitrate content. Five spatial models were developed by incorporating the spatial correlation structure in the regression model through various covariance models. Ortiz et al. (2007) studied spatiotemporal variability of southern root-knot nematode infection in cotton for site-specific management. Semivariograms were used to study the spatiotemporal variability of root-knot nematode, and canonical correlation and cross-correlograms were used to study spatial correlation between root-knot nematode and soil properties. Soil properties highly correlated with root-knot nematode population density were entered into a logistic regression model to create a map of probability risk for root-knot nematode population density. Ge et al. (2007a) used the regression-kriging method to account for spatial dependence among soil samples and aid in prediction model development with soil reflectance spectra measured with a spectroradiometer in visible and NIR wavelengths. For each soil property, semivariance analysis of prediction residuals from principal component regression revealed strong spatial dependence in Na; medium spatial dependence in Ca, Mg, and sand; weak spatial dependence in K and P; and a pure nugget effect in Zn and clay. Fitted theoretical semivariograms were then used to develop the regression-kriging models. Both the principal component regression and regression-kriging models were tested with the validation set, and their prediction capability was evaluated by R^2 and RMSE. The results showed that the regression-kriging models were able to predict most soil properties with reasonably high R^2 and low RMSE.

A new airborne multispectral remote sensing system with built-in camera stabilization was developed for crop management (Lan et al., 2009). Experiments were conducted to evaluate the performance of the system integrated with automatic camera stabilization using a three-CCD (charge-coupled device) MS 4100 multispectral camera (Geospatial Systems, Inc., West Henrietta, N.Y.). Multispectral images were acquired using the integrated imaging system over a 115 ha cotton field. With the processed imagery, spatial relationships between the aerial images and ground-measured values of normalized difference vegetation index (NDVI) and soil apparent electrical conductivity were modeled and estimated.

The goal of the study presented herein is to investigate and compare the model-driven and data-driven approaches in spatial modeling and analysis for spatial prediction of ground soil and crop canopy coverage variability using aerial multispectral images obtained by the airborne imaging system described by Lan et al. (2009).

Objectives

- To apply the spatial econometric and geostatistical methods for modeling ground soil apparent electrical conductivity and crop canopy vigor using airborne remote sensing imagery.
- To compare different spatial models to evaluate their performance in modeling ground soil apparent electrical conductivity and crop canopy vigor.

Materials and Methods

Field of Interest

A 40 ha crop field for this study was located in the eastern part of Burleson County, Texas (30° 33′ 35.61″ N, 96° 26′ 25.89″ W). Dominant soil types in the field include a Ships clay (very-fine, mixed, active, thermic Chromic Hapluderts) and a Belky clay (fine, mixed, active, thermic Entic Hapluderts). The field has been on a corn-cotton rotation in recent years. Cotton was planted in March 2007 and harvested in September. Conventional methods were practiced for insect and weed control.

Aerial Multispectral Imagery

Lan et al. (2009) described a TerraHawk system (Frontier Electronic Systems Corp., Stillwater, Okla.) for automated airborne imaging using an MS 4100 multispectral camera, position-based camera triggering using GPS, and a system for automatic camera stabilization. This integrated multispectral imaging system was mounted on a single-engine Cessna 210 (Cessna Aircraft Co., Wichita, Kans.) for imaging the field in fly-overs. During the fly-overs, an IRR 180 irradiance radiometer (Frontier Electronic Systems Corp., Stillwater, Okla.) was set up to accumulate incoming solar irradiance. After the fly-overs, the digital numbers of the acquired imagery were converted to percent reflectance using the recorded data of solar irradiance. Figure 1 illustrates a georeferenced color infrared (CIR) image of the cotton field acquired on 20 September 2007 at a flight altitude of 2600 m (image resolution of 1.56 m pixel^{-1}). The red color on the image indicates the cotton crop canopy prior to harvest.

Ground-Based Data Collection

Prior to planting in 2007, soil apparent electrical conductivity (ECa) was measured at shallow (0 to 0.3 m) and deep (0 to 0.9 m) depths using a Veris 3150 EC system (Veris Technologies, Inc., Salina, Kans.). At full canopy, crop vigor

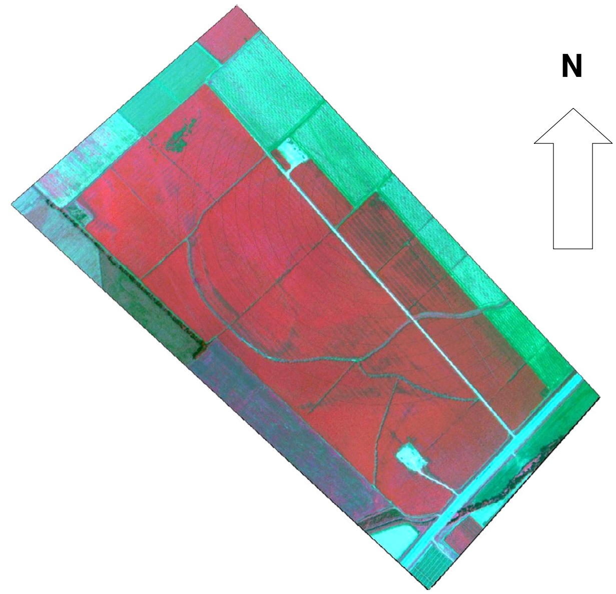

Figure 1. Georeferenced color infrared image of the cotton field acquired on 20 September 2007 using MS 4100 camera in TerraHawk system on Cessna 210.

was measured using a Green Seeker handheld data collection and mapping unit (model 505, NTech Industries, Inc., Ukiah, Cal.), which output NDVI values over the field.

Both Veris and Green Seeker datasets are spatially dense, and it is necessary to reduce the size of the datasets so that they can be effectively handled in statistical software packages. From each of the Veris and Green Seeker datasets, 300 points were randomly extracted to form new Veris and Green Seeker datasets for modeling, and an additional 150 points were extracted to form the datasets for model validation. Figure 2 shows the boundary of the study field, soil types, and the selected modeling and validation data points for both Veris and Green Seeker datasets.

Spatial Data Analysis

In this study, two different methods were used to model ground sensor measurements using aerial remote sensing data. The two methods are model-driven spatial regression and data-driven geostatistics. REML was used to estimate the model parameters in geostatistics; this geostatistical method is referred to as REML-geostatistics for the remainder of the

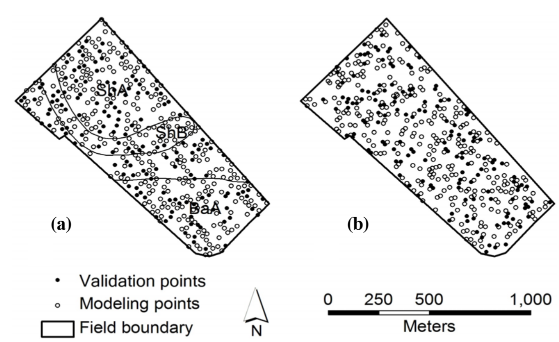

Figure 2. Field boundary of the study field and sample locations for (a) Veris and (b) Green Seeker. Dominant soil types in this field are ShA (Ships clay, 0% to 1% slopes, rarely flooded), ShB (Ships clay, 1% to 3% slopes, rarely flooded), and BaA (Belk clay, 0% to 1% slopes, rarely flooded).

discussion herein. Multiple linear regression as OLS was also implemented as a non-spatial modeling alternative. All spatial models were compared with the non-spatial model for each variable. The predictor variables used to model the ground sensor measurements (Veris measured shallow and deep soil ECa, and Green Seeker measured NDVI) were three original bands (Green, Red, and NIR) and two derived vegetation indices [NDVI = (NIR-Red)/(NIR+Red), and GNDVI = (NIR-Green)/(NIR+Green)] from the aerial CIR image.

The selection of these five variables for modeling was based on our research experience. While factors such as salinity, soil textures, water content, organic matter, cation exchange capacity, macro-nutrients, and drainage class can all affect ECa (Corwin and Lesch, 2005), previous studies have shown that soil clay content and hence soil moisture content is a principal driving factor for the spatial heterogeneity of soil ECa in this area (Ge et al., 2008; Stanislav et al., 2009). Soil moisture content further correlates to crop growth (Stanislav et al., 2009; Stanislav et al., 2010) and canopy optical properties that are captured by different bands of the remote sensing images. Therefore, it is reasonable to use these three original bands and two derived indices to model ground-based measurement such as soil ECa.

The Akaike information criterion (AIC) (Akaike, 1974) and the Bayesian information criterion (BIC) (Schwarz, 1978) were used for model comparison:

$$\mathrm{AIC} = -2\left[N \log\left(\underline{\hat{\theta}}\right) - d\right] \tag{1}$$

$$\mathrm{BIC} = -2\left[N \log\left(\underline{\hat{\theta}}\right) - \frac{d}{2}\log N\right] \tag{2}$$

where N is the sample size of the observations, d is the number of the parameters being estimated of the model, and $\underline{\hat{\theta}}$ is the parameter vector of the model.

AIC and BIC have been used for model comparison in terms of the principle of parsimony (Tukey, 1961). In explaining this principle Tukey (1961) stated that "among model equivalent regressors for y, prefer one with fewest estimated coefficients." Therefore, the best model with minimal AIC and BIC should be the one with least parameters, and the final model will be optimum in terms of minimum AIC and BIC values.

Model-Driven Spatial Regression

The model-driven approach to spatial statistics is concerned with relevance of spatial effects on model specification and estimation (Anselin, 1988). The spatial structure is modeled using simultaneous maximum likelihood estimation. The dependent variable or residuals are a function of a spatially weighted average of neighboring point data. To characterize spatial dependence, spatial weight matrices are constructed and then included in a specially structured regression model.

Spatial weight matrices are designed to represent neighborhood relationships. A spatial matrix of vicinities is defined for introduction of microlocalization factors and spatial dependence within regression models. The matrix is used in spatial models to relate a variable at one point in space to the observations for that variable in other spatial units in the system.

In general, there are two patterns of spatial dependence in regression analysis: spatial lag and spatial error. Correspondingly, there are two methods for introducing the spatial weight matrix (**W**) into the multiple linear regression model:

- Include $\mathbf{W}\underline{\mathbf{y}}$ as an independent variable, which constructs a spatial lag regression model, where $\underline{\mathbf{y}}$ is a vector of observations on the dependent variable. If the spatial lag component is ignored, then the OLS estimates are inconsistent and biased.
- Include $\mathbf{W}\underline{\mathbf{u}}$ as an independent variable of $\underline{\mathbf{u}}$, which constructs a spatial error regression model, where $\underline{\mathbf{u}}$ is the sample vector of the residue of the multiple linear regression model. If the spatial error component is ignored, then the OLS estimates are inefficient but still unbiased.

A spatial lag regression model is a general spatial autoregressive model in which the explanatory variables include a spatial lagged dependent variable (Baller et al., 2001):

$$\begin{aligned} \underline{\mathbf{y}} &= \rho\mathbf{W}\underline{\mathbf{y}} + \mathbf{X}\underline{\boldsymbol{\beta}} + \underline{\boldsymbol{\varepsilon}} \\ \underline{\boldsymbol{\varepsilon}} &\sim \mathrm{ND}\left(\mathbf{0}, \sigma^2\mathbf{I}\right) \end{aligned} \tag{3}$$

where ρ is the coefficient of $\mathbf{W}\underline{\mathbf{y}}$, $\mathbf{X}$ is a matrix of observations on the explanatory variables, $\underline{\boldsymbol{\beta}}$ is the vector of multiple linear regression coefficients on the explanatory variables, $\underline{\boldsymbol{\varepsilon}}$ is the vector of independent and identically distributed (i.i.d.) error terms, ND is the normal distribution, σ^2 is the variance of $\underline{\boldsymbol{\varepsilon}}$, and $\mathbf{I}$ is the identity matrix with ones on the main diagonal and zeros elsewhere.

A spatial error regression model includes an error term that follows a spatial autoregressive process, and the explanatory variables only contain exogenous variables (Haining, 1995):

$$\begin{aligned} \underline{\mathbf{y}} &= \mathbf{X}\underline{\boldsymbol{\beta}} + \underline{\mathbf{u}} \\ \underline{\mathbf{u}} &= \lambda\mathbf{W}\underline{\mathbf{u}} + \underline{\boldsymbol{\varepsilon}} \\ \underline{\boldsymbol{\varepsilon}} &\sim \mathrm{ND}\left(\mathbf{0}, \sigma^2\mathbf{I}\right) \end{aligned} \tag{4}$$

where $\underline{\mathbf{u}}$ is the vector of the spatially autocorrelated error term, $\underline{\boldsymbol{\varepsilon}}$ is the vector of i.i.d. errors, and λ is the coefficient of $\mathbf{W}\underline{\mathbf{u}}$. Other terms are the same as in equation 3.

Data-Driven REML-Geostatistics

The data-driven approach is based on geostatistics. This approach assumes that the spatial structure, such as soil types or management zones in agriculture, is continuous. Spatial modeling with this approach involves fitting spatially distributed data for parametric estimates or an empirical semivariogram of a spatial process.

Geostatistics is a widely employed approach to model datasets that exhibit spatial correlation in agriculture based on variogram modeling of regression residuals. Geostatistics also provides an interpolation technique (kriging) that can predict response variable at unsampled locations (Journel and Huijbregts, 1981). Geostatistics and regression can be integrated to combine the features of both techniques.

Table 1. Multiple linear regression models of Veris measured shallow and deep soil apparent electrical conductivity and Green Seeker measured NDVI.

	Model	AIC	BIC
Veris measured shallow soil ECa	y = 471.8 - 44.42(Green) + 42.27(Red) -3.262**(NIR) + 1836*(NDVI) - 1820(GNDVI)	2980	3006
Veris measured deep soil ECa	y = -110.3 + 21.87(Green) - 6.197(Red) - 7.579**(NIR) + 27.59(NDVI) + 718.7(GNDVI)	2921	2947
Green Seeker measured NDVI	y = 0.9717 + 0.0696(Green) - 0.1057*(Red) + 0.0102(NIR) - 3.43(NDVI) + 2.726(GNDVI)	-668.3	-642.4

Table 2. Calibrated spatial regression models and AIC and BIC values of Veris measured shallow and deep soil apparent electrical conductivity and Green Seeker measured NDVI.

	Model[a]	AIC	BIC
Veris measured shallow soil ECa			
Spatial lag regression	y = 0.9392***(**W***y*) + 1.649(Green) + 7.493*(Red) - 3.608(NIR) - 1307*(NDVI) - 17.36(GNDVI) - 40.87	2820	2850
Spatial error regression	y = 63.81 + 6.977(Green) - 34.88*(Red) - 3.608(NIR) + 179.9(NDVI) + 361.8(GNDVI) + u u = 0.9519***(**W***u*)	2827	2827
Veris measured deep soil ECa			
Spatial lag regression	y = 0.9416***(**W***y*) + 58.16*(Green) - 36.7*(Red) - 7.247**(NIR) + 272.4(NDVI) + 2128(GNDVI) - 374.4*	2767	2796
Spatial error regression	y = -341.6* + 60.44*(Green) - 34.88*(Red) - 8.26**(NIR) - 1359*(NDVI) + 2370(GNDVI) + u u = 0.9496***(**W***u*)	2769	2798
Green Seeker measured NDVI			
Spatial lag regression	y = 0.5941**(**W***y*) + 0.0749(Green) - 0.0989(Red) + 0.0066(NIR) - 3.238(NDVI) + 3.059(GNDVI) + 0.2538	-674.7	-645.1
Spatial error regression	y = 1.95* - 0.0042(Green) - 0.0746(Red) + 0.0166*(NIR) - 2.012(NDVI) - 0.0485(GNDVI) + u u = 0.8069***(**W***u*)	-681.7	-652.1

[a] Statistically significant coefficients are indicated by asterisks, where * indicates p < 0.1, ** indicates p < 0.01, and *** indicates p < 0.001. Parameters with no asterisks are therefore not significant at the 0.1 level.

Table 3. REML-geostatistical models of Veris measured shallow and deep soil apparent electrical conductivity and Green Seeker measured NDVI.

	Model[a]	AIC	BIC
Veris measured shallow soil ECa	y = 88.8 + 2.13(Green) + 10.09(Red) - 4.087(NIR) + 262.1(NDVI) + 194.2(GNDVI) c_0 = 431.3; c_1 = 1207; a = 684; ns = 26.3%	2757	2791
Veris measured deep soil ECa	y = -158.2 + 39.95(Green) - 23.69(Red) - 5.596(NIR) - 925.6(NDVI) + 1603(GNDVI) c_0 = 220.7; c_1 = 816.5; a = 235; ns = 21.3%	2698	2732
Green Seeker measured NDVI	y = 1.726 - 0.0078(Green) - 0.0623(Red) + 0.0151(NIR) - 1.582(NDVI) - 0.1692(GNDVI) c_0 = 0.0045; c_1 = 0.0021; a = 400; ns = 68.2%	-679.8	-646.5

[a] c_0 = nugget, c_1 = partial sill, a = range of the fitted variogram models, and $ns = c_0/(c_0 + c_1)$.

Geostatistics takes into consideration that the residuals from multiple linear regression are spatially dependent in which OLS regression estimates are no longer optimum. In a matrix representation, geostatistical regression can be expressed as:

$$\underline{\mathbf{y}} = \mathbf{X}\underline{\boldsymbol{\beta}} + \mathbf{W}\underline{\mathbf{e}} \tag{5}$$

where $\underline{\mathbf{e}}$ is the vector of regression residuals, and $\mathbf{W}$ is the residual covariance matrix, which is determined by a covariance function having three parameters (c_0 as the nugget variance, c_0+c_1 as the sill variance, and a as the range). Other terms are the same as in equation 3. In this study, the model parameters ($\underline{\boldsymbol{\beta}}$ and $\mathbf{W}$) were estimated by REML, i.e., maximizing the residual maximum likelihood function, and the spherical covariance function was chosen to model the residual covariance structure among the selected sampling points.

Software Tools

All the models were computed using R statistical software (www.r-project.org). Spatial lag and error regressions were performed in the spdep contributed package (http://cran.mtu.edu). The REML-geostatistics was performed in the geoR contributed package (http://cran.mtu.edu).

Results and Discussion

The results of the multiple linear regression model calibration are presented in table 1. The statistically significant coefficients are indicated by asterisks, where * indicates p < 0.1, ** indicates p < 0.01, and *** indicates p < 0.001. Parameters with no asterisks are therefore not significant at the 0.1 level.

Spatial Regression

The spatial lag and error regression models were determined for Veris measured shallow and deep soil ECa and Green Seeker measured NDVI, respectively. Table 2 shows the results of the calibrated spatial regression models. The results indicated that the spatial regression models should have a much better ability in prediction of the spatially varied data over the field with significantly lower AIC and BIC values compared to the OLS multiple linear regression models.

Based on the AIC and BIC values, the spatial lag regression provided better models for Veris measured shallow and deep soil ECa, and the spatial error regression provided the better model for Green Seeker measured NDVI. The comparison of the magnitude of the ρ and λ values (larger values indicate higher spatial autocorrelation) indicated that the Veris measurements (deep and shallow)

Table 4. RMSE values of OLS and REML-geostatistical model validation of Veris measured shallow and deep soil apparent electrical conductivity and Green Seeker measured NDVI.[a]

Model	Veris Shallow Soil ECa	Veris Deep Soil ECa	Green Seeker NDVI
OLS	0.146	0.306	0.165
REML-geostatistics	0.096	0.186	0.149

[a] $$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(\frac{y_{i,measured} - y_{i,predicted}}{y_{i,measured}}\right)^2}{n}}$$

have a higher spatial continuity (in terms of the spatial lag models for spatial correlation) than Green Seeker NDVI (in terms of the spatial error model).

REML-Geostatistics and Model Validation

Table 3 shows the results of the calibrated models of REML-geostatistics. The results indicated that the REML-geostatistical models are even better than the spatial regression models for soil conductivity, with significantly lower AIC and BIC values. Therefore, compared to the spatial regression models and the OLS multiple linear regression models, the REML-geostatistical models should

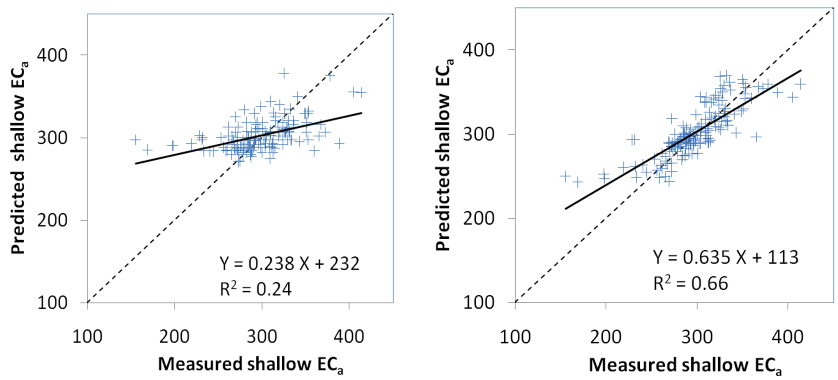

Figure 3. Predicted versus Veris measured shallow soil apparent electrical conductivity (mS m^{-1}) of the validation set using (left) ordinary least squares and (right) REML-geostatistics (dashed line is 1:1 line).

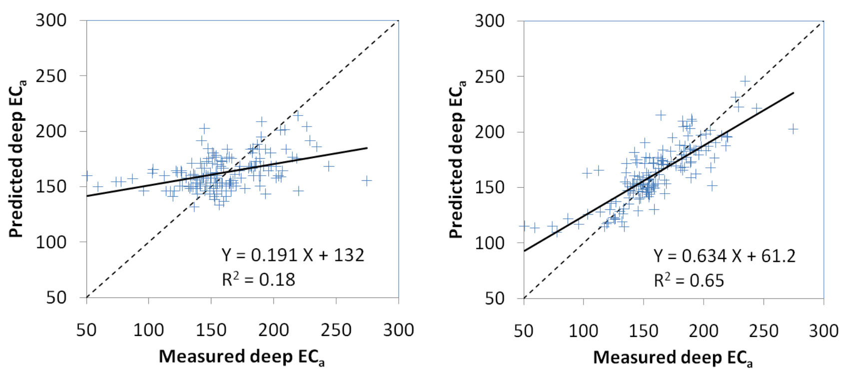

Figure 4. Predicted versus Veris measured deep soil apparent electrical conductivity (mS m^{-1}) of the validation set using (left) ordinary least squares and (right) REML-geostatistics (dashed line is 1:1 line).

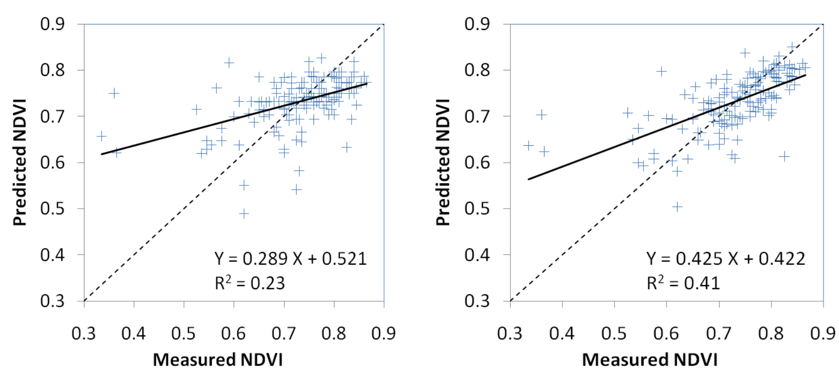

Figure 5. Predicted versus Green Seeker measured NDVI of the validation set using (left) ordinary least squares and (right) REML-geostatistics (dashed line is 1:1 line).

have the best ability in prediction of the spatially varied soil conductivity data over the field. Comparison of the *ns* ratios (lower values indicate higher spatial autocorrelation) also indicated that the Veris measurements (deep and shallow) have a higher spatial continuity than the Green Seeker measured NDVI in terms of the REML-geostatistical models.

Table 4 shows the RMSE values of OLS and REML-geostatistical model validation for Veris measured shallow and deep soil ECa and Green Seeker measured NDVI. The scatter plots of the model validation are shown in figures 3, 4, and 5. The results indicated that the REML-geostatistical models provided consistent predictions. The measured values were significantly better than the corresponding OLS models, with smaller RMSE values and higher correlation between the predicted values and the measured values along the 1:1 line.

The results in table 1 indicate that the OLS models can be used to predict ground-based measurements by airborne remote sensing imagery. However, the statistical significances of the models are relatively low. It seems that remote sensing imagery is capable of capturing the large-scale variation, and hence the general trends in the ground-based dataset can be modeled and predicted by remote sensing imagery. However, there is significant small-scale (or local) variation in the datasets that cannot be effectively modeled and predicted by remote sensing imagery. Therefore, we have to resort to spatial prediction methods to model this small-scale variation by the data itself. This is the reason why spatial regression and REML-geostatistics outperformed the OLS method. With the results in tables 2 and 3, based on comparison of the AIC and BIC values, the REML-geostatistical models were best for all variables, and the OLS models fared poorly for all variables. The spatial lag and error regressions provided comparable models to REML-geostatistics. For the spatial regression models, both ρ and λ are significant at $p < 0.001$ level, indicating very strong spatial correlation in lags and errors. By accounting for spatial dependence, regression coefficients change. Apparently, coefficients estimated by the spatial regression and REML-geostatistics are more reasonable, and the OLS coefficients are less efficient.

Statistically significant correlation between the aerial CIR image and Veris measured ECa data indicated that, even though soils under the crop canopy cannot be directly seen from the remote sensing image, some soil properties can still be inferred from remote sensing images. For example, in these non-saline fields, soil ECa is primarily determined by soil clay content and moisture condition (Corwin and Lesch, 2005). Soils with moderate ECa (such as loamy soils) have a set of desirable attributes (such as favorable hydraulic conductivity, good structure for root development, good aeration conditions, etc.) that stimulate crop growth and produce a greater amount of biomass, which further influences the optical properties of the crop canopy. It is through this soil-crop interaction that soil ECa can be modeled by remote sensing imagery. This would also hold for any soil property that is a crop growth limiting factor during a season (such as macro- and micro-nutrient levels, pH, or lime requirement).

It is interesting to note that the Green Seeker NDVI resulted in models do not fit as well as Veris measured shallow and deep soil ECa for all three prediction methods: OLS, spatial regression, and REML-geostatistics. Veris measured shallow and deep soil ECa were modeled comparably, with much lower AIC and BIC values. A number of reasons may explain why the Veris method gave better results. First, the Green Seeker NDVI measurements might have contained a lot of random noise, which made the modeling result inferior. This can be verified by larger nugget to sill ratios (*ns*) in the NDVI residuals (table 3). Second, the Green Seeker data were highly skewed to the right with relatively few high values, whereas the Veris datasets were quite symmetrical. The normality assumption made by OLS, spatial regression, and REML-geostatistics might be violated by the skewed distribution of Green Seeker NDVI.

In this study, three ground-based (and spatially dense) variables (Veris measured deep and shallow ECa and Green Seeker measured NDVI) were focused on. For these types of data (also including spatially dense yield monitor data), kriging might be sufficient to produce high-resolution GIS layers for site-specific crop management. However, there are many other relevant data, such as soil property and crop physiological measurements, that can be collected through soil and crop sampling. One feature of this type of data is that they are expensive and spatially sparse. Auxiliary datasets obtained from remote sensing imagery would be very useful as co-variables to interpolate these datasets into GIS layers. Methods discussed in this study (spatial regression and REML-geostatistics) would provide versatile techniques for data analysis; this indicates advantages to using remote sensing imagery.

Conclusions

This study investigated and compared three methods (OLS multiple linear regression, spatial regression, and REML-geostatistics) in spatial modeling and analysis for spatial prediction of ground soil and crop canopy coverage variability using aerial multispectral images. A number of conclusions are drawn from the results of the study:

- Although ground-based measurement can be modeled and predicted by airborne remote sensing imagery with the OLS method, the remote sensing imagery is only capable of capturing the large-scale variation and modeling and predicting general trends in the ground-based dataset. To model and predict the significant portion of small-scale, local variation in the datasets, spatial prediction methods such as spatial regression and geostatistics are needed.
- The significant correlation between the aerial CIR image and Veris measured ECa data indicated that some soil properties can be inferred via remote sensing, even though soils under the crop canopy cannot be seen directly from the remote sensing image. The moderate correlation between the aerial CIR image and Green Seeker measured NDVI might be caused by highly right-skewed, non-normally distributed data due to random noise in the measurements.
- Airborne multispectral imagery provided informative co-variables to interpolate the spatially dense, ground-based variables (Veris measured deep and shallow ECa and Green Seeker measured NDVI) into GIS layers. The methods of spatial regression and geostatistics provided versatile techniques for analysis of the datasets.

ACKNOWLEDGMENTS

Special thanks to Dr. Dan Martin and Mr. Lee Denham for their help in making the successful flyover in image acquisition.

REFERENCES

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. Automatic Control* 19(6): 716-723.

Anselin, L. 1988. *Spatial Econometrics: Methods and Models.* Boston, Mass.: Kluwer.

Anselin, L. 1990. What is special about spatial data? Alternative perspectives on spatial data analysis. In *Spatial Statistics: Past, Present, and Future*, 63-77. D. Griffith, ed. Ann Arbor, Mich.: Institute of Mathematical Geography.

Anselin, L., R. Bongiovanni, and J. Lowenbert-Deboer. 2004. A spatial econometric approach to the economics of site-specific nitrogen management in corn production. *American J. Agric. Econ.* 86(3): 675-687.

Bajwa, S. G., and M. Mozaffari. 2007. Effect of N availability on vegetative index of cotton canopy: A spatial regression approach. *Trans. ASABE* 50(5): 1883-1892.

Baller, R., L. Anselin, S. Messner, and D. Hawkins. 2001. Structural covariates of U.S. county homicide rates: Incorporating spatial effect. *Criminology* 39(3): 561-590.

Calderon, G. F.-A. 2009. Spatial regression analysis vs. kriging methods for spatial estimation. *Intl. Adv. Econ. Res.* 15(1): 44-58.

Corwin, D. L., and S. M. Lesch. 2005. Apparent soil electrical conductivity measurements in agriculture. *Computers and Electronics in Agric*. 46: 11-43.

Cressie, N. A. 1993. *Statistics for Spatial Data*. Revised ed. New York, N.Y.: John Wiley and Sons.

Dobermann, A., and J. L. Ping. 2004. Geostatistical integration of yield monitor data and remote sensing improves yield maps. *Agron. J.* 96(1): 285-297.

Ge, Y., J. A. Thomasson, C. L. Morgan, and S. W. Searcy. 2007a. VNIR diffuse reflectance spectroscopy for agricultural soil property determination based on regression-kriging. *Trans. ASABE* 50(3): 1081-1092.

Ge, Y., J. A. Thomasson, R. Sui, and J. Wooton. 2007b. A comparison of regression and regression-kriging for soil characterization using remote sensing imagery. ASABE Paper No. 071001. St. Joseph, Mich.: ASABE.

Ge, Y., J. A. Thomasson, R. Sui, C. L. Morgan, S. W. Searcy, and C. B. Parnell. 2008. Spatial variation of fiber quality and associated loan rate in a dryland cotton field. *Precision Agric.* 9(4): 181-194.

Getis, A., J. Mur, and H. G. Zoller, eds. 2004. *Spatial Economics and Spatial Statistics.* New York, N.Y.: Palgrave Macmillan.

GopalaPillai, S., and L. Tian. 1999. In-field variability detection and spatial yield modeling for corn using digital aerial imaging. *Trans. ASAE* 42(6): 1911-1920.

Griffith, D. A. 1992. Simplifying the normalizing factor in spatial autoregressions for irregular lattices. *Papers in Regional Sci.* (71)1: 71-86.

Haining, R. P. 1990. *Spatial Data Analysis in the Social and Environmental Science.* Cambridge, U.K.: Cambridge University Press.

Haining, R. P. 1995. Data problems in spatial econometric modeling. In *New Direction in Spatial Econometrics*, 156-171. L. Aneslin and R. Florax, eds. Berlin, Germany: Springer.

Hengl, T., G. B. M. Heuvelink, and A. Stein. 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma* 120(1-2): 75-93.

Huang, Y., Y. Lan, and W. C. Hoffmann. 2008. Use of airborne multi-spectral imagery in pest management systems. *Agric. Eng. Intl.* vol. X: manuscript IT 07 010.

Inman, D., R. Khosla, R. Reich, and D. G. Westfall. 2008. Normalized difference vegetation index and soil color-based management zones in irrigated maize. *Agron. J.* 100(1): 60-66.

Journel, A. G., and C. H. J. Huijbregts. 1981. *Mining Geostatistics.* San Diego, Cal.: Academic Press.

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.

Knotters, M., D. J. Brus, and J. H. O. Voshaar. 1995. A comparison of kriging, co-kriging, and kriging combined with regression for spatial interpolation of horizon depth with censored observations. *Geoderma* 67(3-4): 227-246.

Lambert, D. M., J. Lowenberg-DeBoer, and R. Bongiovanni. 2004. A comparison of four spatial regression models for yield monitor data: A case study from Argentina. *Precision Agric.* 5(6): 579-600.

Lan, L., Y. Huang, D. E. Martin, and W. C. Hoffmann. 2009. Development of an airborne remote sensing system for crop pest management: System investigation and verification. *Applied Eng. in Agric.* 25(4): 607-615.

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.

Medlin, C. R., D. R. Shaw, P. D. Gerard, and F. E. LaMastus. 2000. Using remote sensing to detect weed infestation in *Glycine max*. *Weed Sci.* 48(3): 393-398.

Misaghi, F., S. Dayyanidardashti, K. Mohammadi, and M. R. Ehsani. 2004. Application of artificial neural network and geostatistical methods in analyzing strawberry yield data. ASABE Paper No. 041147. St. Joseph, Mich.: ASAE.

Moran, M. S., Y. Inoue, and E. M. Barnes. 1997. Opportunities and limitations for image-based remote sensing in precision crop management. *Remote Sensing Environ.* 61(3): 319-346.

Odeh, I. O. A., A. B. McBratney, and D. J. Chittleborough. 1994. Spatial prediction of soil properties from landform attributes from landform attributes derived from a digital elevation model. *Geoderma* 63(3-4): 197-214.

Odeh, I. O. A., A. B. McBratney, and D. J. Chittleborough. 1995. Further results on prediction of soil properties from terrain attributes: Heterotropic co-kriging and regression-kriging. *Geoderma* 67(3-4): 215-226.

Ortiz, B. V., D. G. Sullivan, C. Perry, and G. Vellidis. 2007. Spatial variability of root knot nematodes in relation to within field variability of soil properties. ASABE Paper No. 071196. St. Joseph, Mich.: ASABE.

Paelinck, J. H. P., and L. H. Klaassen. 1979. *Spatial Econometrics*. Farnborough, U.K.: Saxon House.

Pinter, P. J., Jr. J. L. Hatfield, J. S. Schepers, E. M. Barnes, M. S. Moran, C. S. T. Daughtry, and D. R. Upchurch. 2003. Remote sensing for crop management. *Photogrammetric Eng. and Remote Sensing* 69(6): 647-664.

Schabenberger, O., and F. J. Pierce. 2002. *Contemporary Statistical Models for the Plant and Soil Sciences.* Boca Raton, Fla.: CRC Press.

Schwarz, G. E. 1978. Estimating the dimension of a model. *Annals of Statistics* 6(2): 461-464.

Senay, G. B., A. D. Ward, J. G. Lyon, N. R. Fausey, and S. E. Nokes. 1998. Manipulation of high spatial resolution aircraft remote sensing data for use in site-specific farming. *Trans. ASAE* 41(2): 489-495.

Stanislav, S., C. Morgan, A. Thomasson, R. Sui, and T. Cothren. 2009. Implementing soil-specific seeding rates to increase field-scale cotton lint. In *Proc. Beltwide Cotton Conf.*, 1249-1254. Cordova, Tenn.: National Cotton Council.

Stanislav, S., C. L. S. Morgan, A. J. Thomasson, R. Sui, J. T. Cothren, and E. Hequet. 2010. The effect of soil specific seeding rates on water stress and cotton yield. In *Proc. Beltwide Cotton Conf.*, 124. Cordova, Tenn.: National Cotton Council.

Tukey, J. W. 1961. Discussion, Emphasizing the connection between analysis of variance and spectrum analysis. *Technometrics* 3(2): 191-219.

Yang, C., and G. L. Anderson. 1999. Airborne videography to identify spatial plant growth variability for grain sorghum. *Precision Agric.* 1(1): 67-79.

Yang, C., and J. H. Everitt. 2002. Relationships between yield monitor data and airborne multispectral multidate digital imagery for grain sorghum. *Precision Agric.* 3(4): 373-388.

Yang, C., J. H. Everitt, J. M. Bradford, and D. Murden. 2009. Comparison of airborne multispectral and hyperspectral imagery for estimating grain sorghum yield. *Trans. ASABE* 52(2): 641-649.

Yao, H., L. Tian, G. Wang, and I. A. Colonna. 2003. Soil nutrient mapping using aerial hyperspectral imaging and soil sampling data: A geostatistical approach. ASABE Paper No. 031046. St. Joseph, Mich.: ASABE.

Ye, X., K. Sakai, S. Asada, and A. Sasao. 2007. Use of airborne multispectral imagery to discriminate and map weed infestations in a citrus orchard. *Weed Biol. and Mgmt.* 7(1): 23-30.