Defining the experimental unit for the design and analysis of site-specific experiments in commercial cotton fields

Jeffrey L. Willers a,*, George A. Milliken b, Johnie N. Jenkins a, Charles G. O’Hara c, Patrick D. Gerard d, Daniel B. Reynolds e, Debbie L. Boykin f, Paul V. Good g, Kenneth B. Hood h

a Genetics and Precision Agriculture Research Unit, USDA-ARS, Mississippi State, MS 39762, USA
b Department of Statistics, Kansas State University, USA
c GeoResources Institute, Mississippi State, MS, USA
d Experimental Statistics Unit, Mississippi State, MS, USA
e Department of Plant and Soil Sciences, Mississippi State, MS, USA
f Statistics Unit, USDA-ARS, Stoneville, MS, USA
g Good’s Longview Farm, Macon, MS, USA
h Perthshire Farms, Gunnison, MS, USA

Received 20 October 2006; received in revised form 27 September 2007; accepted 28 September 2007
Available online 19 November 2007

Abstract

Designing experiments involves several processes. The first process identifies the experimental units generated by conducting the experiment. The second process is the application of planned treatments to the experimental units. Other processes are the analysis of the experiment and interpretation of results. While traditional experimental small plot designs are useful for investigating many aspects of cotton production, on a large scale, like an entire commercial field, they are difficult to implement because of routine farming operations and topographical variability. As an alternative, by using the capabilities of a variable-rate controller, it is possible to intersect the swath width of the largest farm implement’s boom (or tool bar) along its geo-referenced paths of travel with one or more geo-registered field zones to create experimental units having different shapes and sizes. Defining the experimental units in this manner establishes a site-specific experiment throughout the entire field. Spatial information recovered by Geographic Information System (GIS) processing from these asymmetrical experimental units, when coupled with general linear mixed model methodology, permits the assessment of effects on geo-referenced yield points due to topography, the site-specific and/or traditional farm management practices, and various interactions among them. In this paper, a general methodological approach for analyzing field-sized site-specific experiments is developed and described. An application is demonstrated by analyzing an unreplicated cotton variety trial that included a single site-specific application of a plant growth regulator.
Published by Elsevier Ltd.

Keywords: General linear mixed models; Geographic information systems; Precision agriculture; Site-specific management; On-farm research; Remote sensing; LiDAR

1. Introduction

Site-specific management is experiencing increased popularity on commercial farms in order to manage and take advantage of spatial variability (Reetz and Snyder, 2005; Sonka et al., 1997). Diverse remote sensing systems (Ackermann, 1999; Jensen, 2000; Moran et al., 1997; Pinter et al.,...
2003) acquire the necessary information to make spatial decisions tailored for maximizing crop production (Corwin and Lesch, 2003a; Dupont et al., 2000; Fleming et al., 2004; Willers et al., 1999, 2005). Generally, field topography characteristics are combined with grower recommendations to create a prescription map that provides a variable-rate controller the information needed to apply site-specific practices to a crop. Crop harvesting equipped with differential, global positioning system (Kennedy, 1996) monitors estimate yield at different locations in the field (Birrell et al., 1996; Pierce et al., 1997).

On a commercial farm, the grower is interested in learning if the site-specific or traditional (also called blanket, broadcast, or flat rate) management practices were effective, which means that some type of analysis is necessary. It is also important to learn how these inputs interact with unmanageable, spatially variable effects due to topography, because the goal of precision agriculture is to apply the appropriate amount at the right place at the right time.

The occurrence of high field-to-field variability implies that spatial decisions involving agri-chemical applications are unique to individual fields as well as different locations within them. Under these circumstances, classical small plot designs or trial and error methods of evaluating spatial decisions of a producer are inefficient at best or inappropriate. These methods fall short since they do not utilize the abundant information about geo-referenced characteristics of an entire field or the geo-referenced geometries generated by farm machines as they complete planting, management recommendations, or harvesting operations. Collectively, all of these events generate geo-referenced experimental units of various shapes and sizes that must be considered in order to build an appropriate statistical model and complete an analysis.

Traditionally designed experiments correspond ‘poorly to a real farm context, in which multiple factors vary simultaneously’ (Sonka et al., 1997). Several constraints arise when designing experiments to evaluate site-specific practices applied to a large commercial field. The first is how to define representative units of replication, including the identification of control units (Mead, 1988). Second, the spatial nature of site-specific management makes it impractical to employ a classical designed experiment, which uses similar sized and shaped plots systematically arranged as either whole plots or split plots, because they are not congruent with the scale and pattern of natural variability (Jensen, 2000; Moran et al., 1997; Pinter et al., 2003; Plant et al., 2001) in crop growth. Third, when excessive field variability occurs, including more replications of treatments in any arrangement of systematic plot configurations will not improve the precision of the experiment (Griffin et al., 2005). Fourth, classical symmetrical plot designs often interfere with routine farming operations (Plant, 2007) because the size of commercial farm machines makes it difficult for them to work efficiently in small sized research plots. On the other hand, while larger plots or strips conform better to the size of the machines, they too cannot conform to the natural pattern of the field’s topography, still making it difficult to evaluate treatments. Finally, strategies to evaluate site-specific practices by point sampling within small areas (Dupont et al., 2000) at several locations throughout large fields are too laborious and often have an inadequate sample size in comparison to the degree of field heterogeneity to be informative. Many of these issues also apply to the evaluation of traditional management practices.

Several authors have addressed the problem of how to analyze landscape experiments (Bramley et al., 1999; Conquest, 2000; Hong et al., 2005; Hurlbert, 1984; Lark and Wheeler, 2003; McDonald et al., 2000; Meek and Singer, 2004; Murtaugh, 2000; Nelsen and Palmquist, 2003; Scheiner and Gurevitch, 2001). Of these, only a few include geo-referenced topographic variables as part of the analysis (e.g., Li et al., 2001). Recently, other investigators have developed methods that utilize geo-referenced information. Plant (2007) describes how to make comparisons of means of spatial data in non-replicated field trials using a spatial model for autocorrelated data. Gotway et al. (1997), Grif- fin et al. (2005), and Lowenberg-DeBoer (2002a,b) incorporate the spatial autocorrelation structure of data provided by yield monitors. But for most of these experiments, randomly assigned treatments were uniformly applied to arbitrary plots (or strips) of similar size and shape (e.g., Bermudez and Mallarino, 2007; Doerge and Gardner, 1999). No other studies jointly take into account field topography and the characteristics of the farm equipment that plant and manage the crop as additional sources of variability. The tasks performed by various machines used to produce a crop also influence crop growth (Fig. 1). These effects represent additional sources of variation that must be considered to better evaluate the efficacy of site-specific and traditional management practices. Together, field topography and farming operations cause geographical heterogeneity in the sizes and shapes of different areas of crop growth.

This paper proposes that agricultural experiments involving site-specific practices on commercial farm fields require the integration of the agricultural, Geographic Information Systems (GIS) and statistical disciplines to develop new methods of design and analysis. Therefore, we present an approach to analyze experiments that involves both site-specific and traditional management practices in commercial fields. The order of presentation is to (1) describe two different types of geo-referenced experimental units generated by the various farming operations, (2) introduce a general case model that utilizes geo-referenced information about field topography and yield to evaluate the effects of site-specific and traditional treatments implemented by the producer, and (3) demonstrate the utility of the methodology by presenting a case study analysis of a nonreplicated cotton variety trial which included a single site-specific application of a plant growth regulator.
2. Methodology and concepts for site-specific experiments

Our analysis process evaluates the interaction patterns among the levels of the inputs (e.g., cultivar selection, application rate) and site variables describing the topography of the field. The centerpiece of the methodology is a unique conceptualization of how to define two kinds of experimental units according to the intersection of field topography and the geo-referenced travel paths of different kinds of variable-rate application equipment as they apply various management recommendations. At the end of the season, the travel paths and yield point records logged by yield monitor equipped harvesters are nested within these different kinds of experimental units. All of the geo-referenced information is then used to construct an appropriate statistical model to evaluate how various management practices and attributes of field topography affect crop production. The intersection of site-specific practices with traditional applications (applied without regard to field topography) is also considered by this analysis methodology.

2.1. Defining the experimental units

In a commercial field, site-specific management practices involve several sizes of experimental units. The various management zones of a field are the most important factor determining characteristics of these experimental units. The second factor is the site-specific management practices selected by the producer for each zone and applied by variable-rate equipped machines.

In cotton production, various researchers have established diverse criteria to define different management zones within a field. For instance, Landsat Thematic Mapper imagery for 11 consecutive years from the same cotton field was studied by Boydell and McBratney (2002) as a technique to establish temporally stable regions of similarity. Bronson et al. (2003) examined the effect of landscape position and soil series on cotton phosphorous utilization. Using soil electrical conductivity (ECa) measurements, Corwin et al. (2003b) found significant correlations with several soil properties such as leaching fraction, pH, plant-available water, and salinity with cotton yield and provided valuable information for site-specific management. Fridgen et al. (2004) developed software that used a fuzzy c-means unsupervised classification algorithm to apportion field information into management zones. In the pest management of the tarnished plant bug in cotton, Willers et al. (2005) used unsupervised classification techniques of normalized difference vegetation index (NDVI) values derived from imagery to determine different growth phenology classes. In other crops, such as grain, other techniques have also been investigated (Ping and Dobermann, 2003).

In GIS analysis, a zone is comprised of all the cells in a raster file representation of a geographical attribute that have the same value, regardless of whether or not they are contiguous (GIS Dictionary, 2006; Theobald, 2003). The raster model uses the concept of a zone to represent a feature that is relatively homogeneous (Theobald, 2003); thus, a management zone is a feature where the same recommendation will be applied to an area of a field. Since there are many approaches developed to define management zones, they will be simply called ‘zones’. Irrespective of the processing methodology used to define these zones, the analysis methodology introduced in this paper can still be applied.

More than one kind of zone may occur within a cotton field and any particular zone can differ in arrangement, shape and size. Fig. 2 illustrates an example with two A zones and one each of zones B and C. These three zones, when intersected by geo-referenced travel paths of variable-rate farm equipment, create asymmetrical experimental units where the site-specific treatments are applied. Each experimental unit associated with the site-specific treatments can be labeled using a general notation. The first index \(i\) names the zone, the second index \(j\) identifies a particular travel path (or swath) of the largest variable-rate farm machine where it crosses a zone, and the third
index \((k)\) identifies the replication of the zone in a particular path. In this illustration, there are three site-specific treatment experimental units in the first sprayer application path (beginning at the lower left), five in the second path (beginning at the top center) and five in the third path (which ends at the upper right) for a total of 13 experimental units. The replication index of the various zones in the paths may be incremented according to the direction of travel by the farm equipment. A particular zone may or may not be replicated within all of the travel paths.

At other times, the producer may apply a traditional management practice without regard to knowledge about zones. These traditional practices establish the second kind of experimental unit that has larger size, more symmetrical shape, and requires a change in notation. This change is that now the index \(i\) refers to different choices (or levels) for implementing the traditional management practice. For example, the producer may wish to plant three cotton cultivars in a field. Each cultivar is randomly assigned to different strips across the length of the field (Fig. 3), with the width of each strip still determined by the largest boom (or toolbar) of a farm machine. Typically, this machine will be a ground sprayer that spans 24 rows. In the illustration (Fig. 3), each cultivar (C1, C2, or C3) is planted in adjacent travel paths of a 12-row planter, making the width of a single planter pass one-half the size of the sprayer path. Therefore, two 12-row planter paths of a single cultivar are nested within a single sprayer pass of 24 rows.

Harvesting operations establish additional relationships primarily driven by the characteristics of the site-specific experimental units and not by the traditional practices experimental units. For example, a 6-row wide harvester will have two travel paths nested within each planter pass and will have four paths nested within each sprayer pass (Fig. 3). The four harvest passes within a sprayer pass

---

**Fig. 2.** Experimental unit \((e_{ijk})\) definitions, where the \(k\)th replicate is indexed within three application passes \((j)\) of the farm implement having the largest boom or toolbar traveling over different zones \((i)\). The small field having three zones is positioned within a larger agricultural landscape that extends beyond its exterior boundary.

**Fig. 3.** Crop cultivars (C1, C2, C3, or other types of blanket management practices) can be randomly assigned to the sprayer strips. Within a planter pass (two are shown at the right), 12 rows at a time are planted, establishing two planting passes per each of the nine sprayer passes which span 24 rows. Harvesting operations span six rows at a time, or half of one planter pass. The yield monitor measurements, or swath elements, are shown by the small squares of one harvest pass at the left. This illustrative field contains 36 harvester passes. The illustration identifies five site-specific experimental units and five harvester pass sub-samples within various sizes of additional site-specific experimental units. There are three zones in this imaginary field.
apportion each site-specific experimental unit into four sub-samples. Each of these sub-samples is further divided into several sub-sub-samples called swath elements (Fig. 3). Each swath element describes the yield response for a specified area on the ground logged by the yield monitor as an individual record. For many monitors, yield records (or points) are written to a file every 2 s and estimate the yield for that swath element, taking the length of travel between readings into account. A good practical approximation is that swath elements are identical in size (area) while the sub-sample zones are of different lengths according to the size and shape of the site-specific experimental unit.

Using the coordinates of each swath element’s centroid, GIS processing extracts field topography characteristics (represented in raster file formats (Theobald, 2003)) and other information about the levels, or application rates, of the various management practices (represented in vector file formats (Theobald, 2003)) and attaches them as additional attributes to each yield record (Willers et al., 2004). Once the GIS data processing steps are completed, a geo-referenced attribute table describes the relationships among the various machines, zones, site-specific and traditional management practice recommendations, and yield data. The structure of this attribute table corresponds to the different types of experimental units just discussed. This geo-registered information provides the starting point for evaluating traditional and site-specific management practices.

2.2. Building a generic linear mixed analysis of covariance model

Concepts for GIS processing and the design of experiments are linked together to develop a process for analyzing geo-referenced information collected from the field. In general, the intersecting geometries of the zones and the farm equipment’s characteristics define the design structure of a site-specific experimental design and its different sizes and kinds of experimental units. The site-specific and traditional practices applied to the crop by the producer describe the treatment structure of the field experiment. The design and treatment structures can be put together (Mead, 1988; Milliken and Johnson, 1992) to create a statistical model that describes the response variable and evaluates the effectiveness of both the site-specific and traditional treatments. The model will include terms for each of the treatment structure factors and for each of the different sizes of experimental units. It will have the levels of the traditional management practice and the levels, or rates, of the site-specific practice as fixed effects and will have the various parts of the design structure as random effects. This model also contains, as fixed effects, one or more terms related to the topography variables assessed by one or more kinds of remote sensing systems. This kind of statistical model is called a linear mixed model (Littell et al., 2006).

To better understand the derivation of the final statistical model, two intermediate models are described. The first model describes yield monitor data as functions of farm equipment geometries, several cotton cultivars (e.g., \( i = 1,2,3 = C \)), representing the levels of the traditional practices, and different rates \(( r = 1,2,3,4 = R \)) of an agrichemical (e.g., a plant growth regulator) applied site-specifically to various zones as

\[
y_{ijkmn} = \mu + \gamma_i + \rho_r + (\gamma \rho)_i + s_{ij} + w_{ijk} + p_{ijkl} + h_{ijklm} + e_{ijklmn},
\]

(1)

where \( y_{ijkmn} \) is the yield value of the \( n \)th yield monitor reading in the \( m \)th harvest pass within the \( k \)th planter pass of the \( i \)th experimental unit (or replication of a site-specific recommendation) receiving the \( r \)th rate in the \( j \)th sprayer application pass of the \( l \)th cultivar. The term \( \mu \) is the mean yield of the field, \( \gamma_i \) is the effect of the \( i \)th cotton cultivar, \( \rho_r \) is the effect of the \( r \)th level or rate of the agri-chemical and \((\gamma \rho)_i \) is the interaction between the cotton cultivars and the rate levels. Several random effects (Fig. 3) of the model are \( s_{ij} \) (the \( j \)th sprayer pass for the \( i \)th cultivar), \( w_{ijk} \) (the \( k \)th experimental unit within the \( i \), \( j \)th combination of cultivar and sprayer pass where a particular rate of an agro-chemical was applied), \( p_{ijkl} \) (the \( l \)th planter pass within the \( k \)th experimental unit within the \( j \)th sprayer pass for the \( i \)th cultivar), \( h_{ijklm} \) (the \( m \)th harvester pass within the \( k \)th planter pass within the \( j \)th experimental unit within the \( l \)th planter pass for the \( i \)th cultivar), and the residual, \( e_{ijklmn} \) (the \( n \)th yield monitor observation within the \( m \)th harvester path within the \( l \)th planter pass of the \( k \)th experimental unit in the \( j \)th sprayer application pass of the \( i \)th cultivar).

Model 1 consists of a two-way treatment structure where all combinations of the \( C \) levels of cultivar occur with each of the \( R \) levels of rate. As previously discussed (see Fig. 4, far left), due to variability in field topography, it is possible that not all of the site-specific management rates of the plant growth regulator will be applied to all of the cultivars in every sprayer path, yielding an unbalanced treatment structure. Since plant growth regulator rate is quantitative, the rate \((\rho_v)\) and rate by cultivar interactions \((\gamma \rho)\) terms can be replaced with a linear trend in rate that has a different slope for each cultivar, to produce the second intermediate model as

\[
y_{ijkmn} = \mu_i + \varphi_{rk} + s_{ij} + w_{ijk} + p_{ijkl} + h_{ijklm} + e_{ijklmn},
\]

(2)

where \( \mu_i \) is the mean yield of the \( i \)th cotton cultivar, \( r_{ijk} \) is the plant growth regulator rate applied to the area corresponding to the \( w_{ijk} \) experimental unit and \( \varphi_{rk} \) is the slope representing the \( i \)th cultivar’s response to the plant growth regulator rates. This model assumes that two or more rates are applied to each cultivar somewhere in the field. In other words, it is not required that two or more rates of plant growth regulator occur in every individual sprayer pass for the \( i \)th cultivar.

However, Model 2 does not account for the variability in the yield response described by the topography, or the
site-specific experimental design just described. Other useful concepts are found in Milliken and Johnson (2002), particularly the chapter discussing the analysis of covariance for split-plot and strip-plot design structures.

Milliken (2003) provides an informative discussion on the analysis of multilevel designs relevant to the general site-specific experimental design just described. Other useful concepts are found in Milliken and Johnson (2002), particularly the chapter discussing the analysis of covariance for split-plot and strip-plot design structures.
3. Case study

In this section, the general process just discussed is demonstrated by analyzing an actual site-specific experiment completed on a commercial cotton farm. While the general case was based on a 12-row planter pattern, the example employs an 8-row planter pattern without loss of generality.

3.1. Methods

3.1.1. Field data

During the 2003 production season, the case study site-specific experimental design was established in approximately 8.88 ha of a 64.75 ha field at Good’s Longview Farm, Noxubee County, several kilometers east of Macon, MS. The row spacing was 0.762 m to promote yield production while reducing soil erosion in a dry land production system. Planting occurred within 2 days of 22 April 2003. Twenty-one cotton cultivars were planted in a configuration of twenty-two, 8-row by 804 m long strip plots (Fig. 5, bottom left), with hopper bins thoroughly emptied between cultivars. For this example analysis, only seventeen cultivars were used and represent the levels of the traditional practice. Other information about the field experiment can be found in Willers et al. (2004).

The first application of plant growth regulator (mepiquat pentaborate) to these cultivars was a broadcast application (0.37 L ha\(^{-1}\)) applied as 24-row swaths on 26 June 2003. The second (and last) application was a site-specific plant growth regulator prescription applied in 8-row swaths on 18 July 2003. The four rates of plant growth regulator applied were 0, 0.29, 0.43, and 0.59 L ha\(^{-1}\) according to the zones defined by a prescription map built from

---

Fig. 5. Grayscale multi-spectral image (top) used to establish the spatially variable plant growth regulator application. Expanded from the inset are the planter (and/or sprayer) paths (bottom left) and the two harvest passes per cultivar strip (bottom right). The four colors assigned to the yield points depict some of the asymmetrical experimental units of the site-specific plant growth regulator prescription within each strip.
a multi-band image composite acquired 9 June 2003 (Fig. 5, top). These rates represent the levels of the site-specific management practice.

Harvesting took place 16 September 2003 using a 4-row cotton picker equipped with a differential global positioning system cotton yield monitor. Yield data were collected once every two seconds and (as harvester speed varied) the distance between logged points differed (1.00–4.45 m). The yield monitor supplied several attributes of each yield point (Fig. 5, bottom right), including seed cotton mass (kg ha⁻¹), time and location. Two unique load identification labels (LOAD_ID) for each cultivar strip were numbered as the picker traveled the first harvest path south to north and the second path north to south. Harvesting of cultivar strips occurred in order, proceeding from the west to east.

All other inputs were blanket applications applied to all cultivars and were not included in the analysis.

3.1.2. Imagery and LiDAR acquisitions of field topography

Several field topography attributes were measured across the field at various resolutions by remote sensing techniques. At the coordinate location of the centroid of each swath element, GIS processing attached one or more topography attributes as site characteristics to each record in the yield file (Willers et al., 2004).

Multi-spectral imagery was obtained at pixel resolutions of 0.5 m². The center bands of the imagery were: 450 nm (blue), 550 nm (green), 650 nm (red) and 850 nm (near infrared (NIR)). Three image dates were utilized: bare (blue), 550 nm (green), 650 nm (red) and 850 nm (near infrared (NIR)). NDVI site characteristics at different times of the year were named NDVI 15 JAN (for 15 January 2002), NDVI 9 JUNE (for 9 June 2003), NDVI 17 JULY (for 17 July 2003). Generally, the NDVI varies over the growing season, particularly before peak bloom, due to complex relationships among soil type, weather, and management practices. To capture some of these temporal changes, we calculated a new attribute grid named NDVI DIFFERENCE, which represented the grid subtraction of NDVI 9 JUNE from NDVI 15 JAN. Other details of the GIS processing steps are described in Willers et al. (2004).

3.1.3. Geographic information system processing

After harvest, additional GIS processing (Anonymous, 1997; Theobald, 2003) refined the data files analyzed by general linear mixed statistical models. The basis of the data structure (as dBase IV table derived from a shapefile) was the yield file obtained from the cotton yield monitor. The various site (topography) characteristics (Chang, 2006) associated with each swath element were SLOPE (percent), ASPECT (a categorical number corresponding to compass direction), FLOW ACCUMULATION (m²), CURVATURE (negative to positive real numbers), EUCLIDEAN DISTANCE (m) from a synthetic stream network, and ELEVATION (m). NDVI site characteristics at different times of the year were named NDVI 15 JAN (for 15 January 2002), NDVI 9 JUNE (for 9 June 2003), NDVI 17 JULY (for 17 July 2003). Generally, the NDVI varies over the growing season, particularly before peak bloom, due to complex relationships among soil type, weather, and management practices. To capture some of these temporal changes, we calculated a new attribute grid named NDVI DIFFERENCE, which represented the grid subtraction of NDVI 9 JUNE from NDVI 15 JAN. Other details of the GIS processing steps are described in Willers et al. (2004).

3.1.4. Statistical data processing, model building and analyses

This cultivar trial is an unreplicated strip-plot design (Milliken and Johnson, 1992) and cannot be extensively analyzed by traditional methods of analyses. However, because the data and management practices are geo-referenced, it can be extensively analyzed as a site-specific experimental design. The study is a two-way treatment structure with 17 levels of cotton cultivars and four levels of the site-specific plant growth regulator application. The nine stacked gene cultivars (Plots 6–14, Bollgard® and Roundup Ready®) were planted in one ‘group’ and the eight single gene cultivars (Plots 15–22, Roundup Ready® only) were planted in another ‘group’. (These group labels were not a source of variation used in this analysis.) The design structure was comprised of 17 strips of eight rows intersected by the various spatial management zones (Fig. 5, top and bottom left panel) for the site-specific plant growth regulator application and later harvested in two loads from two passes per cultivar by a 4-row cotton picker (Fig. 5, bottom right panel).

The attribute table (Theobald, 2003) of the summary shapefile was imported into statistical software to start building the statistical model. The statistical model’s treatment variables were (1) the traditional management practices, called CULTIVAR and (2) the site-specific management practices called RATE. Additional variables needed for modeling were constructed and include LOAD_ID, PASS (or direction ‘north’ or ‘south’ traveled by the harvester within a cultivar strip), LOG FLOW ACCUMULATION, LOG EUCLIDEAN DISTANCE, and many other derivatives, like a BOGGINESS INDEX (=log(FLOW ACCUMULATION + 1)/(SLOPE)), or NDVI 17 JULY RECIPROCAL (=NDVI 17 JULY * (1/exp(NDVI DIFFERENCE))). Another variable, called RATECLASS was generated to identify which yield records belonged to the site-specific experimental units where a particular level of plant growth regulator was applied. The rate of plant growth regulator applied (RATE and RATE²) as well as ELEVATION and ELEVATION² were later used in the modeling and were included in interactions with cultivar and the topography variables. LOAD_ID within CULTIVAR, or the two harvester passes within each cultivar strip, was another random effect in
the model. Outlying yield monitor data points occurring during times when the harvester went through a waterway or ditch were deleted. The first and last 9.1 m of each harvest load were also deleted from the data set since the yield monitor often produced extreme readings near the ends of the harvest passes (Ping and Dobermann, 2005).

The response variable was the SEED COTTON YIELD (kg ha\(^{-1}\)) provided by the yield monitor readings. However, the yield monitor data are quite variable from one time point to the next because of the process of measuring yield. Therefore, several other GIS and statistical filtering steps of the yield records were completed (Willers et al., 2004) before analyses.

Field topography variables as covariates as well as their interactions with CULTIVAR are included in the model. The first step was to fit the whole model (i.e., all possible combinations of variables) to the filtered data and compute residuals. Next, a stepwise elimination process removed variables and interactions, one at a time, until all remaining variables were significant at the 0.05 level or less. For the site-specific case study, the general linear mixed model is

\[ y_{i,k,m} = \mu + \phi_i r_{ik} + \psi_i r^2_{ik} + \sum_{g=4}^{G} \beta_{g} X_{gikm} + w_{ik} + h_{ikm} + e_{ikm}. \]  

(9)

For this model, there is only one sprayer and one planter pass per cultivar (Fig. 5, bottom left panel), so two subscripts can be omitted when compared to the more generic Model (3). This mixed regression model (Gotway and Stroup, 1997; Littell et al., 2006), or mixed analysis of covariance model (Milliken and Johnson, 2002), describes seed cotton yield as a function of cultivar, plant growth regulator rate, and the site characteristic variables (Tables 1a–1c). It is also possible to compute least squares means (Littell et al., 2006) for each cultivar and plant growth regulator combination. These least squares means are the predicted values of the cultivar by plant growth regulator combinations as if each was grown and applied on the whole field. Several tables of least square means and interaction plots obtained with Model 9 are found in Willers et al. (2004).

For comparison purposes, another analysis of the final data set was completed without site characteristics and used the plant growth regulator rate as a continuous variable entered as both a linear and a quadratic effect. The general linear mixed model without site characteristic variables is

\[ y_{ikm} = \beta_0 + \beta_1 x_{ik} + \beta_2 x_{ik}^2 + \beta_3 t_{ik} + \beta_4 s_{ik} + w_{ik} + h_{ikm} + e_{ikm}. \]  

The analysis of variance results for this model are presented in Table 2. Tables of least significant difference (LSD) values of the least squares means and interaction plots for Model 10 are in Willers et al. (2004).

### 3.2. Results and discussion

#### 3.2.1. Geographic Information System/Remote Sensing

Images and LiDAR data were acquired for the focus field and various data products were derived (Willers et al., 2004). GIS processing built the prescription map for the site-specific application of the plant growth regulator which was applied with a variable-rate equipped ground sprayer. The crop was harvested with equipment monitoring continuous yield rates as discrete points. The fact that the various types of information were geo-referenced provided the foundation for GIS processing to produce a table of information useful for statistical analysis.

#### 3.2.2. Statistical analyses

The general linear mixed model analysis (Model 9) of the data using the site characteristics and management

### Table 1a

Fixed effects tests of management practices derived from the analysis of covariance model utilizing site (or topography) characteristics

<table>
<thead>
<tr>
<th>Effect</th>
<th>Num DF(^a)</th>
<th>Den DF(^b)</th>
<th>F value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>CULTIVAR</td>
<td>16</td>
<td>17</td>
<td>10.28</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>RATE</td>
<td>1</td>
<td>5979</td>
<td>2.82</td>
<td>0.0933</td>
</tr>
<tr>
<td>RATE(^2)</td>
<td>1</td>
<td>5979</td>
<td>4.45</td>
<td>0.0349</td>
</tr>
<tr>
<td>RATE * CULTIVAR</td>
<td>16</td>
<td>5979</td>
<td>6.18</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>RATE(^2) * CULTIVAR</td>
<td>16</td>
<td>5979</td>
<td>7.41</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

\(^a\) Num DF, numerator degrees of freedom.
\(^b\) Den DF, denominator degrees of freedom.

### Table 1b

Fixed effects tests containing multi-spectral information derived from the analysis of covariance model utilizing site (or topography) characteristics

<table>
<thead>
<tr>
<th>Effect</th>
<th>Num DF(^a)</th>
<th>Den DF(^b)</th>
<th>F value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDVI DIFFERENCE</td>
<td>1</td>
<td>5979</td>
<td>3.09</td>
<td>0.0788</td>
</tr>
<tr>
<td>NDVI 17 JULY</td>
<td>1</td>
<td>5979</td>
<td>2.54</td>
<td>0.1112</td>
</tr>
<tr>
<td>RATE * NDVI 17 JULY * CULTIVAR</td>
<td>17</td>
<td>5979</td>
<td>5.94</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>RATE(^2) * NDVI 17 JULY * CULTIVAR</td>
<td>17</td>
<td>5979</td>
<td>4.85</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>RATE(^2) * NDVI 9 JUNE * CULTIVAR</td>
<td>17</td>
<td>5979</td>
<td>3.18</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>RATE * NDVI DIFFERENCE * CULTIVAR</td>
<td>17</td>
<td>5979</td>
<td>3.85</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>NDVI 17 JULY RECIPROCAL</td>
<td>1</td>
<td>5979</td>
<td>2.79</td>
<td>0.0947</td>
</tr>
</tbody>
</table>

\(^a\) Num DF, numerator degrees of freedom.
\(^b\) Den DF, denominator degrees of freedom.
The practice information was insightful. The geo-referenced centroids of the swath elements and the various management and topography attributes associated with each provided the data density, scale, and spatial context necessary to study relationships among them and their effects on yield. The presence of several interaction terms from the site-specific analysis (Tables 1a–1c) involving CULTIVAR indicate that there are different functions of independent variables that geographically describe the yield response of the cotton cultivars. The occurrence of the plant growth regulator rate variable in several other interaction terms (Tables 1a–1c) indicates that its effects upon yield depends on both the site and cultivar characteristics occurring at the location of each yield point. For example, one three-way interaction (RATE * CURVATURE * CULTIVAR), suggests that applying too much or too little plant growth regulator to some cultivars growing vigorously in places that collect rainwater decreased yield, while the amount applied to other cultivars that grew vigorously where water collects did not decrease yield. The agricultural implication of simultaneous site and cultivar interactions with plant growth regulator rate classes is that spatial prescription maps generated by unilateral processes, while useful, may be less than optimal. This means that the plant growth regulator application map based only upon the NDVI 9 JUNE values obtained earlier in the season was too simplistic. The analysis indicates that there is room for improvement by taking into account the type of cotton cultivar and the hydrological information derived from elevation data obtained by LiDAR sensor system. Additional optimization experiments need to be planned to establish better rules for the spatial application of plant growth regulators (e.g., Kirkpatrick, 2006).

Table 1c
Fixed effects tests that primarily involve LiDAR information derived from the analysis of covariance model utilizing site (or topography) characteristics

<table>
<thead>
<tr>
<th>Effect</th>
<th>Num DF</th>
<th>Den DF</th>
<th>F value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLOPE</td>
<td>1</td>
<td>5979</td>
<td>3.00</td>
<td>0.0834</td>
</tr>
<tr>
<td>LOG FLOW ACCUMULATION</td>
<td>1</td>
<td>5979</td>
<td>21.60</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>CURVATURE</td>
<td>1</td>
<td>5979</td>
<td>544.65</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>LOG EUCLIDEAN DISTANCE</td>
<td>1</td>
<td>5979</td>
<td>3.92</td>
<td>0.0477</td>
</tr>
<tr>
<td>ELEVATION</td>
<td>1</td>
<td>5979</td>
<td>527.19</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>BOGGINESS INDEX</td>
<td>1</td>
<td>5979</td>
<td>0.05</td>
<td>0.8207</td>
</tr>
<tr>
<td>CURVATURE RECIPROCAL</td>
<td>1</td>
<td>5979</td>
<td>678.02</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>LOG FLOW ACCUMULATION * CURVATURE</td>
<td>1</td>
<td>5979</td>
<td>290.96</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>ELEVATION * CULTIVAR</td>
<td>16</td>
<td>5979</td>
<td>10.35</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>RATE * ELEVATION</td>
<td>1</td>
<td>5979</td>
<td>2.80</td>
<td>0.0944</td>
</tr>
<tr>
<td>RATE * CURVATURE * CULTIVAR</td>
<td>17</td>
<td>5979</td>
<td>9.77</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>RATE * LOG EUCLIDEAN DISTANCE * CULTIVAR</td>
<td>17</td>
<td>5979</td>
<td>1.91</td>
<td>0.0135</td>
</tr>
<tr>
<td>RATE2 * LOG FLOW ACCUMULATION * CULTIVAR</td>
<td>17</td>
<td>5979</td>
<td>8.29</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>RATE2 * LOG EUCLIDEAN DISTANCE * CULTIVAR</td>
<td>17</td>
<td>5979</td>
<td>1.89</td>
<td>0.0145</td>
</tr>
<tr>
<td>RATE2 * ELEVATION * CULTIVAR</td>
<td>17</td>
<td>5979</td>
<td>7.40</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>RATE * LOG FLOW ACCUMULATION</td>
<td>1</td>
<td>5979</td>
<td>1.52</td>
<td>0.2176</td>
</tr>
<tr>
<td>RATE2 * CURVATURE * CULTIVAR</td>
<td>17</td>
<td>5979</td>
<td>9.62</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>RATE2 * SLOPE * CULTIVAR</td>
<td>17</td>
<td>5979</td>
<td>64.09</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>RATE * BOGGINESS INDEX * CULTIVAR</td>
<td>17</td>
<td>5979</td>
<td>8.76</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>RATE * SLOPE</td>
<td>1</td>
<td>5979</td>
<td>0.79</td>
<td>0.3748</td>
</tr>
<tr>
<td>LOG FLOW ACCUMULATION * ELEVATION</td>
<td>1</td>
<td>5979</td>
<td>22.91</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>SLOPE * ELEVATION</td>
<td>1</td>
<td>5979</td>
<td>2.89</td>
<td>0.0894</td>
</tr>
<tr>
<td>LOG FLOW ACCUMULATION * NDVI DIFFERENCE</td>
<td>1</td>
<td>5979</td>
<td>51.67</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>ELEVATION2 * CULTIVAR</td>
<td>17</td>
<td>5979</td>
<td>55.82</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

* Num DF, numerator degrees of freedom.
* Den DF, denominator degrees of freedom.

Table 2
Fixed effects tests for the model without site (or topography) characteristics

<table>
<thead>
<tr>
<th>Effect</th>
<th>Num DF</th>
<th>Den DF</th>
<th>F value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>CULTIVAR</td>
<td>16</td>
<td>3139</td>
<td>1.80</td>
<td>0.0257</td>
</tr>
<tr>
<td>RATE</td>
<td>1</td>
<td>6250</td>
<td>23.36</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>RATE * CULTIVAR</td>
<td>16</td>
<td>6250</td>
<td>2.05</td>
<td>0.0082</td>
</tr>
<tr>
<td>RATE2 * CULTIVAR</td>
<td>17</td>
<td>3440</td>
<td>3.43</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

* Num DF, numerator degrees of freedom.
* Den DF, denominator degrees of freedom.
Table 3
Estimates of selected covariance parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>With site characteristics</th>
<th>Without site characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOAD_ID(CULTIVAR)</td>
<td>13.96</td>
<td>321.25</td>
</tr>
<tr>
<td>Sill</td>
<td>32.465</td>
<td>96.110</td>
</tr>
<tr>
<td>Range</td>
<td>8.0767</td>
<td>0.4678</td>
</tr>
<tr>
<td>Nugget</td>
<td>0.0146</td>
<td>0.9114</td>
</tr>
</tbody>
</table>

a LOAD_ID is the attribute identifying one of two harvest passes within each cultivar strip.

RATE * NDVI 17 JULY * CULTIVAR (Tables 1a, 1b, and 1c). So, while both analyses showed that the yield of the cotton cultivars interacts with RATE and RATE$, it is unlikely that both analyses similarly describe how. Comparisons of interaction plots and least squares means presented in Willers et al. (2004) show other differences between the two analyses.

Other comparisons between the models that use or do not use topography covariates can also be made. For instance, with use of the site characteristics, a Gaussian spatial structure modeled the correlation among experimental units of the same plant growth regulator rate within a cultivar and among harvest passes. Estimates for the LOAD_ID(CULTIVAR) variance component, silt, range and nugget (Littell et al., 2006, p. 441) were obtained (Table 3). Next, another analysis without the site characteristics and with a spherical spatial structure (because the Gaussian spatial structure did not converge) provided estimates for the same parameters (Table 3). Except for the range parameter, estimates of covariance parameters are much larger for the model without site characteristics than the model with site characteristics. The larger estimate for the range with the site characteristics model indicates the impact of topography information upon the correlation structure among the yield points. Overall, the trends in these estimates suggest that the availability of site characteristics improved the analysis process.

One final point is that in a commercial field, site-specific experimental units are assigned treatments by the producer’s judgment and not by a process of randomization. The purpose of randomization is to avoid bias. For a site-specific experiment, an exclusively random process to assign a site-specific treatment to an experimental unit will result in an irrelevant assignment of a recommendation to some of these units, because the site characteristics of all replications will not consistently match that particular treatment. Whenever a site-specific recommendation can be associated with topography attributes, the use of randomization to determine which units are assigned site-specific treatments will introduce bias and not avoid it. For example, the random assignment of a variable herbicide rate to an experimental unit where there are no weeds is not practical. With a site-specific experiment, dynamic changes in a variable-rate controller’s instructions along its paths of travel are based upon the characteristics of the zones. This association results in the assignment of different treatments to various asymmetrical experimental units and does so better than a process of randomization. With this perspective, having information about statistical interactions among topography variables should avoid the future assignment of site-specific treatments to inappropriate locations of the field and allow better allocation of resources.

4. Conclusions

A team comprised of the producer, precision agricultural specialist, GIS artist and statistical analyst is necessary to conduct and analyze a site-specific experiment in a commercial field. Defining the experimental units associated with site-specific and traditional practices provides a good starting point for designing and analyzing management practices in commercial agricultural fields. A site-specific experimental design provides an evaluation of effects of both types of management decisions upon yield and describes their interactions with attributes of field topography. The finding that both types of management practices interact with site characteristics indicates that it is crucial to evaluate different combinations of site characteristics in order to recommend the best practices for different locations in the field.

Acknowledgements

The spatial plant growth regulator prescription was applied in cooperation with Dr. Matt Kirkpatrick and Dr. Tom Barber, Plant and Soil Sciences, Mississippi State University. The cultivar trial was planted in cooperation with Good’s Longview Farm, and sponsored by Mr. Dave Roberts, Delta Pine and Land Seed Company, Scott, MS, Mr. Dennis Reginelli, Mississippi Cooperative Extension Service, Noxubee County, MS, and Dr. Will McCarty, Associate Director, Mississippi State University Extension Service. Imagery and LiDAR data were acquired in cooperation with the GeoResources Institute at Mississippi State University. We thank Dr. Gene Burris, LSU, St. Joseph, LA, for helpful comments on the manuscript. Research was supported in part by grant funds from the GeoResources Institute, Mississippi State University and the USDA-ARS, Special Project, Area-Wide Pest Management of the Tarnished Plant Bug. Mississippi Agricultural and Forestry Experiment Station Publication No. J-10877.

References


