ACCOUNTING FOR SPATIAL CORRELATION USING RADIAL SMOOTHERS IN STATISTICAL MODELS USED FOR DEVELOPING VARIABLE-RATE TREATMENT PRESCRIPTIONS

K.S. McCarter

Department of Experimental Statistics
Louisiana State University
Baton Rouge, Louisiana

E. Burris

Louisiana State University AgCenter
Northeast Research Station
St. Joseph, Louisiana

ABSTRACT

Variable-rate treatment prescriptions for use on commercial farms can be developed from embedded field trials on those farms. Such embedded trials typically involve non-random, high-density sampling schemes that result in large datasets and response variables exhibiting spatial correlation. In order to accurately evaluate the significance of the effects of the applied treatments and the measured field characteristics on the response of interest, this spatial correlation must be accounted for in the statistical analysis of the data. One approach is to use a fully parametric model that accounts for the treatment and design structures of the experiment as well as any residual spatial correlation. For example, the MIXED procedure in SAS® includes a variety of parametric spatial covariance structures that can presumably be used for this purpose. However, we have found that because of the large size of the datasets that result from precision agriculture experiments, MIXED is often unable to fit models that include one of these parametric spatial structures. Another approach is to use a model that incorporates a non-parametric smoother to account for any residual spatial correlation, in addition to a parametric component that accounts for the treatment and design structures of the experiment Such semi-parametric models utilize fewer computing resources and can be used with large datasets. The GLIMMIX procedure in version 9.2 of SAS® includes a radial smoother that can be used for this purpose. We demonstrate the use of radial smoothers in GLIMMIX to fit semi-parametric models that account for spatial correlation. We compare inferences from models that account for spatial correlation using radial smoothers to those from models that do not account for spatial correlation. In addition, we discuss several important issues that arise when fitting models utilizing radial smoothers, such as selecting the number of knots to use in the radial smoother.

Keywords: semi-parametric model, radial smoother, linear mixed model, spatial correlation, precision agriculture.

INTRODUCTION

The purpose of this article is to demonstrate the use of nonparametric smoothers in mitigating spatial correlation in precision agriculture data in order to improve the inferential validity of the statistical analysis of that type of data.

Spatial correlation can have any of several causes. It can result from the effects of unmeasured or otherwise unaccounted-for field characteristics that vary spatially across a field. Failure to adequately account for trend in a particular dimension can induce spatial correlation among the residuals in that dimension. Such correlation would tend to be positive, with residuals having a greater likelihood of being similar when they are close together than when they are far apart. Spatial correlation can also result from problems involving the measurement process itself. For example, in measuring cotton yield the cotton passes by a yield monitor as it travels through the chute on its way to the hopper. If the cotton becomes temporarily jammed near the yield monitor, measured yield values can either increase or decrease substantially, depending on which side of the monitor the jam occurs. Once the jam clears, yield measurements would move substantially in the opposite direction until normal flow was restored. If they occur frequently enough, such jams could induce a negative spatial correlation in the yield measurements over short distances.

In some cases spatial correlation can be modeled directly using a parametric error covariance structure. Several types of spatial covariance structures have been investigated in the spatial statistics literature, and modern statistical software typically provides such modeling capability. For example, both the MIXED and GLIMMIX procedures in SAS® provide the ability to fit parametric error covariance structures. This approach can be problematic, however, if the software is not implemented using algorithms designed for use with large datasets. It has been our experience that with datasets as large as those typically generated by precision agriculture applications, general purpose statistical software is often unable to fit parametric covariance structures. The software runs out of memory while trying to fit the covariance structure.

If residual spatial correlation is the result of unaccounted-for trend in a particular dimension, then it stands to reason that accounting for that trend should reduce or eliminate the spatial correlation in that dimension. In order to account for trend parametrically using common statistical software, the trend must be well-approximated by a function possessing a relatively simple mathematical representation (e.q., polynomial). Even when this is the case, incorporating a parametric trend function in a statistical model requires knowing what that trend function is. How can this trend function be determined? When the space over which trend exists is one-dimensional, residual plots can be helpful in detecting unaccounted-for trend and understanding its nature. In two dimensions, residual plots may still be somewhat useful for detecting the existence of trend. However, in two dimensions residual plots are harder to assess and are therefore less useful in determining a suitable parametric form for that trend. As a result, for the kinds of two-dimensional, geo-referenced data generated by precision agriculture applications, it can be very difficult to develop fully parametric models that eliminate unaccounted-for trend as a source of residual spatial correlation.

Instead of accounting for residual trend parametrically, an alternative is to model the trend nonparametrically. There is a vast literature in nonparametric smoothing we can draw upon for this purpose, and it is not our purpose to go into this literature in detail. Suffice it to say that in one dimension there are many kinds of smoothers available. While they are fewer in number, smoothers are also available for two and higher dimensions. One of the advantages of nonparametric smoothing splines is there ability to follow quite complicated trend curves (in one dimension) or surfaces (in two dimensions) without having to specify a particular parametric functional form.

Nonparametric smoothers can be combined with parametric model components that account for the treatment and design structures of an experiment. The resulting model is called a semi-parametric model. Semi-parametric models utilize fewer computing resources than those that model the spatial covariance structure parametrically. They can therefore be used with the large datasets common to precision agriculture research. Certain types of semi-parametric models containing penalized splines have representations as mixed models, and can therefore be fit with mixed-model software. The GLIMMIX procedure in version 9.2 of SAS® includes several nonparametric smoothers. Some of these smoothers are included in a model as fixed effects, while others are included as random effects. The penalized radial smoothing spline, the nonparametric smoother of primary interest in this paper, is incorporated in a model as a random effect. Our goal in using semi-parametric models is to improve the validity of inferences involving the treatments in a precision agriculture experiment. This is accomplished by using nonparametric smoothers to account for residual spatial trend, which in turn can reduce or eliminate residual spatial correlation.

In semi-parametric models where a smoothing spline is implemented as a random effect, it may seem somewhat mysterious as to what the smoother is actually doing. On the other hand, semi-parametric models implemented using fixed effect splines are generally easier to understand. The thing to keep in mind is that the mixed-model implementation of a semi-parametric model is just a particular representation of an underlying model that has a fixed-effect representation as well, and these two representations are equivalent. If we can understand the fixed-effects representation of the model, then we understand what the random effects representation is doing also.

To get a better understanding of semi-parametric models, we first demonstrate their use in a very simplified, one-dimensional analysis of covariance (ANCOVA) situation (Milliken and Johnson, 2002). In this context we fit a semi-parametric model in which the nonparametric component is implemented as a fixed effect. We show that this model is essentially an ANCOVA model where the nonparametric spline plays the role of the covariate. We then fit a semi-parametric model that incorporates a radial smoother. When implemented as a random effect, a radial smoother is essentially a penalized nonparametric smoothing spline. The advantage of radial smoothers is that they easily extend to two or more dimensions, and hence can be used in precision agriculture settings. We then show that these two models give essentially the same results. This provides a basis for understanding what mixed-model semi-parametric models, and in particular models incorporating radial smoothers, are doing.

We then consider a case study using data from an actual precision agriculture experiment. The data from the experiment are analyzed two ways. The data are first analyzed using a model that makes no attempt to account for residual spatial correlation. We then analyze the data using a semi-parametric model that incorporates a two-dimensional penalized radial smoothing spline. The radial smoothing spline accounts for enough of the residual trend that the residual spatial correlation is essentially eliminated. We compare the inferences regarding the treatments for the two models, and demonstrate that the inferences that are made depend on whether or not the presence of spatial correlation is addressed.

ONE-DIMENSIONAL EXAMPLE USING SIMULATED DATA

In this first example we analyze a simulated dataset constructed to illustrate several of the assertions made in the introduction. The data are generated from an ANCOVA model utilizing a one-way treatment structure and a completely randomized design structure (Milliken and Johnson, 2009). The single treatment has two levels. The affect of the covariate on the response includes linear and quadratic terms, as well as an additional component to create oscillation around their sum. The generated error terms are independent and identically distributed (iid) zero mean Normal variates with constant variance.

The covariate is restricted to one dimension in order to more easily compute correlations between residuals. So that these correlations can be interpreted properly, the covariate values are equally spaced. For our purposes it suffices to focus on the correlation between adjacent (i.e., lag-one) residuals along the dimension of the covariate.

Five models are fit to the data to help illustrate the points being made: (1) an ANOVA model that ignores the covariate, (2) a linear ANCOVA model, (3) a quadratic ANCOVA model, (4) a semi-parametric model using a second-order truncated power function (TPF) spline with five knots, and (5) a semi-parametric model with a penalized radial smoothing spline using 10 knots. The first three are no doubt familiar to the reader, so they will not be described in detail. The last two involving splines may be less familiar, and so we will describe them briefly.

The semi-parametric model using a second-order TPF spline with five knots is given by $y_{ij} = \mu_{ij} + e_{ij}$, where

$$\mu_{ij} = \beta_{0i} + \beta_1 T_j + \beta_2 T_j^2 + \sum_{m=1}^{5} \gamma_m \left(\left| T - \kappa_m \right|_+ \right)^2 .$$

In this expression κ_m is the m-th knot, and $\left| T_j - \kappa_m \right|_+$ is the positive part of $T_j - \kappa_m$; that is,

$$\left| T_j - \kappa_m \right|_+ = \begin{cases} T_j - \kappa_m & \text{if } T_j > \kappa_m \\ 0 & \text{if } T_j \le \kappa_m . \end{cases}$$

Note that

$$\mu_{1j} - \mu_{2j} = \beta_{01} - \beta_{02} ,$$

so that the difference between the treatment means is the same across the entire range of the covariate. The parameters β_{0i}, β_1, β_2, γ_m, $m = 1,\ldots,5$ are all fixed effects. This model is fit with the following SAS® statements.

```
proc glimmix data=ex01;
class TRT ;
effect SPL = spline(T / basis=TPF() degree=2
                    knotmethod=equal(5));
model Y = TRT SPL;
run;
```

The “effect” statement creates the linear, quadratic, and truncated power spline variables and stores them all together under the name SPL. All are included in the model as fixed effects by including the SPL identifier in the model statement as shown.

The semi-parametric model incorporating a penalized radial smoothing spline is more complicated. Because of space limitations in this paper we refer the interested reader to Ruppert et al. (2003) or the SAS® documentation for details. The following SAS® statements fit this model.

```
proc glimmix data= ex01;
class TRT;
model Y = TRT T;
random T / type=rsmooth knotmethod=equal(10);
run;
```

Boxplots of the observed data and fitted values from the ANOVA model are given in Figure 1. Plots of the observed and fitted values for the ANCOVA and semi-parametric models are given in Figure 2. Selected results for all five models are given in Table 1 and Table 2.

The boxplots in the left panel of Figure 1 reveal only a slight difference between the two samples. The ANOVA comparing treatment means is not significant, with a p-value of 0.4019. The point estimate of the difference between the means of treatments one and two is -15.38, and a 95% confidence interval estimate of this difference is [-51.63, 20.89]. Clearly the problem with this analysis is that we have completely ignored the affect that the covariate has on the response, which can be seen in the plot in the right panel.

Figure 2 shows attempts to account for this trend using the ANCOVA models and the semi-parametric models. The linear ANCOVA model picks up the general increasing linear trend in the data, but ignores the curvilinear trend that exists. The quadratic ANCOVA model accounts for both the linear and the quadratic trends, but it is not equipped to handle the oscillation that exists about the quadratic trend. On the other hand, both of the semi-parametric spline models are able to detect and model this oscillation, resulting in better fits to the data. Note that because the semi-parametric model with the radial smoother is penalized whereas the model using the TPF spline is not, the model using the five-knot TPF spline is able to follow the data more closely than the model using the ten-knot radial smoother. It is clear that both of the semi-parametric models give a much better fit to the data than either of the parametric ANCOVA models.

Table 1 gives estimates of the error variance as well as inferences for the fixed effects for all five models. As stated previously, because it completely ignores the information provided by the covariate the ANOVA model is unable to detect the

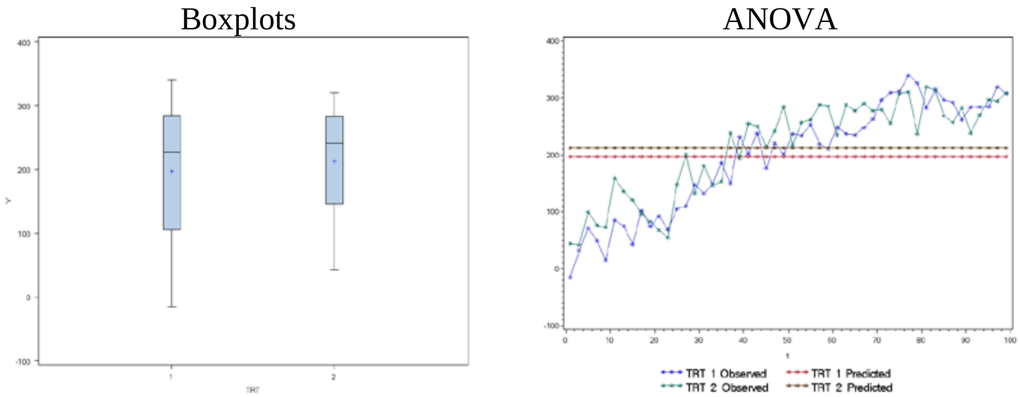

Figure 1. Box-plots of observed values, and observed and fitted values from the ANOVA.

difference between the treatment means. On the other hand, both of the ANCOVA models and both of the semi-parametric models are able to detect the difference between the treatment means, although the linear ANCOVA model is just barely significant at the 5% level. From the table we see that the more we account for the effect of the covariate on the response, several things happen. The estimate of the error variance decreases, the p-value for the comparison of treatment means decreases, and the confidence interval for the difference between treatment means become more precise. In short, we are able to make sharper, more accurate inferences the more accurately we model the relationship between the response and the covariate. This is what we expect from using an ANCOVA, and it should be clear that the splines are serving the role of covariates. There is more that can be said, though, and this involves the residual correlation.

Table 2 gives the lag-one residual correlation, by treatment, for each of the five models. From the plots of the observed and fitted values for the ANOVA model and the two ANCOVA models in Figures 1 and 2, it is clear that residuals that are close with respect to their covariate values will tend to be more similar than residuals that are far apart. This induces a correlation among the residuals in that dimension. However, this relationship among the residuals is not an accurate reflection of the probabilistic relationship between the true error terms that they are estimating. Recall that the data were generated using stochastically independent errors. In the ANOVA model, where the trend along the covariate dimension has been completely ignored, the lag-one residuals are the most-highly correlated. In the linear ANCOVA model we account for some of the trend in that direction, and as a result the residual correlation has decreased for that model. However, we haven't accounted for the trend completely, and because of this the lag-one residual correlation is still statistically significant. The quadratic ANCOVA model accounts for even more of the trend, and hence the lag-one residual correlation decreases further, although it is still significant for the residuals under treatment 1. On the other hand, the residual correlations for each of the semi-parametric models are not significantly different from zero. The inclusion of the smoothing splines allows the fitted model to follow the complicated trend associated with the covariate. As a result the residuals are not

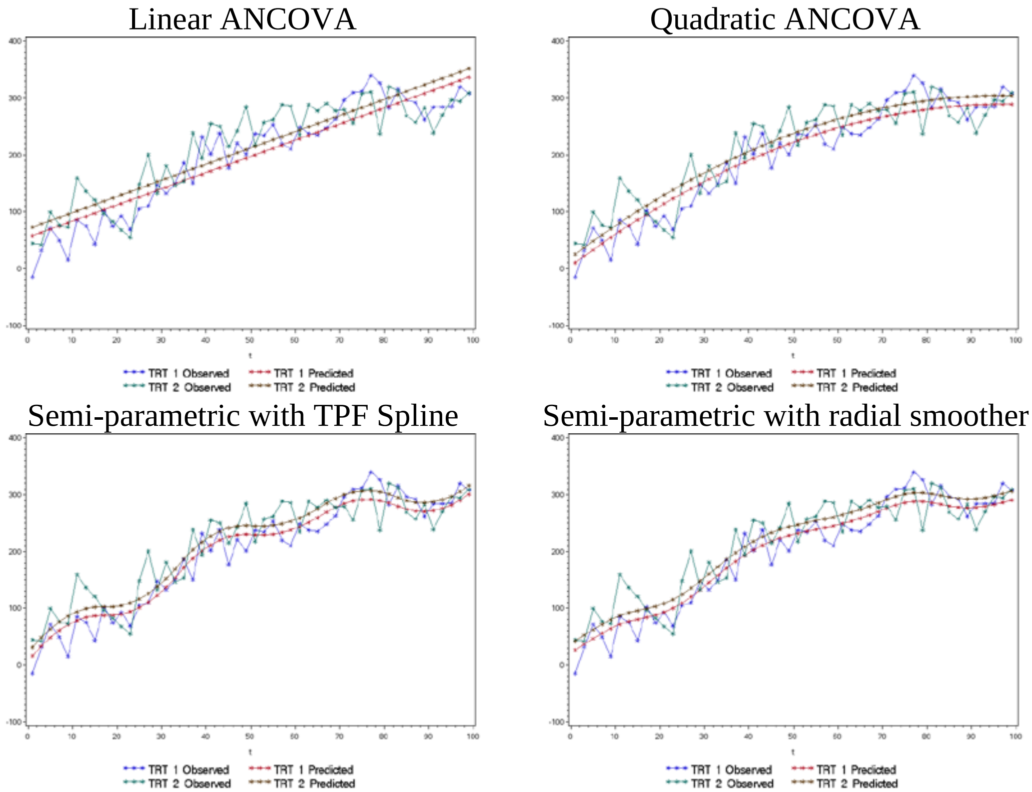

Figure 2. Observed and fitted values for the ANCOVA models and the semi-parametric spline models.

polluted by unaccounted-for trend and therefore exhibit the uncorrelated nature of the underlying error terms they are estimating.

These results demonstrate two statements made in the introduction. First, that failure to adequately account for trend in a particular dimension can induce spatial correlation among the residuals in that dimension. Second, to the extent that residual spatial correlation is due to unaccounted-for trend, by accounting for that trend we can reduce, and possibly eliminate, that residual spatial correlation.

Table 1. Estimates of error variance and inferences regarding treatment effects for each model.

Model	Estimate of Error Variance	Treatment p-value	TRT 1 – TRT 2 95% CI
ANOVA	8,342.85	0.4019	[-51.63, 20.89]
Linear ANCOVA	1,470.21	0.0477	[-30.60, -0.16]
Quadratic ANCOVA	957.82	0.0147	[-27.67, -3.09]
Semi-parametric w/TPF	777.37	0.0070	[-26.46, -4.30]
Semi-parametric w/radial smoother	792.55	0.0076	[-26.57, -4.19]

Table 2. Lag 1 residual correlation, by treatment, for each model.

Model	Correlation	(p-value)
	Treatment 1	Treatment 2
ANOVA	0.944 (<0.0001)	0.874 (<0.0001))
Linear ANCOVA	0.533 (<0.0001)	0.506 (0.0002)
Quadratic ANCOVA	0.298 (0.0374)	0.257 (0.0752)
Semi-parametric w/TPF	0.080 (0.583)	0.071 (0.626)
Semi-parametric w/radial smoother	0.088 (0.550)	0.089 (0.544)

These results also provide a way of interpreting what semi-parametric models like these are doing. In the way we are using them here they are essentially semi-parametric ANCOVA models that can be used to account for complicated trend that is left unaccounted for by the parametric part of the model.

Precision Agriculture Case Study

We now apply the techniques discussed above to the analysis of a precision agriculture dataset. The data for this case study were obtained from an on-farm field trial conducted on a commercial cotton farm in northeast Louisiana. The purpose of the study was to evaluate the effects of a nematicide and three nitrogen rates on cotton lint yield in order to develop a treatment prescription for future use on that field. Burris et al. (2009) used this data in developing a treatment prescription for that field. Their prescription incorporated producer preferences and was created using automated procedures, both of which were defined and described in McCarter et al. (2007). The statistical methodology they used was also described in McCarter et al. (2007) and is in the same vein as that detailed by Willers et al. (2008). In this case study we extend that methodology by the addition of a radial smoothing spline to the model, and illustrate how such semi-parametric models can be used to reduce or eliminate spatial correlation, thereby increasing the validity of the statistical results, and hence improving the resulting treatment prescription.

Description of the field trial

From prior research the field was known to vary spatially with respect to soil type. Apparent soil electroconductivity (EC_a) measurements were taken across the entire field and used as a proxy for soil type. From the raw EC_a values researchers created an ordinal variable defining three soil-type categories representing low, medium, and high quantities of soil clay content.

There were two applied treatments used in the experiment, a nematicide and nitrogen fertilizer. The nematicide treatment consisted of two levels: applied at a fixed rate or not applied at all. Three nitrogen rates were used, which we will call levels 1, 2, and 3 in this paper.

The experiment was laid out in three replicates, with the six nitrogen-nematicide treatment combinations assigned at random to plots within each replicate. Plots extended the length of the field and were each 24 rows wide. Nitrogen application equipment spanned 12 rows. Nitrogen application passes were nested within the 24-row wide treatment plots, requiring 2 application passes within each treatment plot. Nematicide application passes were nested within Nitrogen application pass. The nematicide application equipment spanned 4 rows, requiring 3 nematicide application passes per nitrogen application pass.

At harvest, a yield monitor on the cotton picker measured cotton lint yield every two seconds as it traversed the field. Yield data were spatially referenced using a gps receiver mounted on the picker. The cotton picker spanned 6 rows. Harvest passes were nested within application pass, and hence there were 2 harvest passes per nitrogen application pass. Note that the two harvest passes within a Nitrogen application pass each covered half of the middle nematicide application pass within that Nitrogen application pass. This increases the potential for correlation among yield measurements within the two harvest passes within a Nitrogen application pass.

The dataset contained 6008 yield measurements. Precision agriculture experiments can produce large datasets, and this is actually a relatively small dataset for this type of application.

Statistical analysis

The measured response variable, cotton lint yield, contains several sources of variability that can be divided into the following categories: the applied treatments, the observed field characteristic, and the variability induced by the conduct of the experiment. In addition, because yield measurements are taken repeatedly as the cotton picker traverses the field, the yield measurements may exhibit spatial correlation.

The applied treatments consist of nitrogen rate (NRATE) and nematicide (NMTCD). The observed field characteristic that we consider in this analysis is the EC_a zone (EC_ZONE), which as described previously serves as a proxy for soil type. A mixed model analysis of variance is used to model cotton lint yield (YLD) as a function of the applied treatments and the measured field characteristic (McCarter et. al, 2007; Willers et al., 2008). The variables NRATE, NMTCD, and EC_ZONE are fixed effects. These variables are included in the model as main effects. All two- and three-way interactions between these variables are included in the model as well.

There are several possible sources of variation resulting from the conduct of the experiment that should be considered for inclusion in the model as random effects. There is potential random variation among REPs, and the 24-row PLOTs to which the treatments were randomized. Other potential sources of variation include nitrogen application pass (APASS) and harvest pass (HPASS). All were included in the initial model in order to assess their significance and to determine the random effects part of the model. Once the random effects part of the model has been determined, the model can then be used to evaluate the effects of NRATE and NMTCD within each EC_ZONE.

We first fit a model that assumes an independent error structure for the R-side random effects (i.e., model residuals). The first step in developing the model involves determining the random-effects part of the model. Random effects that were initially included include REP, PLOT, APASS, and HPASS. Even though REP was initially conceived of as a blocking factor when the researchers designed the study, the field locations corresponding to the REPs were very large and did not represent natural blocks. Estimates of the variance components for REP and PLOT were zero, and hence REP and PLOT were removed from the model. The random effects due to application pass (APASS) and harvest pass within application pass (APASS*HPASS) were significant, and were retained in the model. The resulting model was fit using the following GLIMMIX code.

```
proc glimmix data=yield;
class EC_ZONE NRATE NMTCD APASS HPASS ;
model YLD = EC_ZONE | NRATE | NMTCD
        LOC_X LOC_Y
        / ddfm=satterth;
random APASS;
random APASS*HPASS;
run;
```

Table 3 shows the dimensions of the X and Z design matrices and the number of covariance parameters in the model. There are 36 application passes in the dataset. Of these, one contains a single harvest pass, while the rest contain two. Hence there are $36 + 35 \times 2 + 1 = 107$ random effects in the model, as reflected in the "Columns in Z" row of the table. There are two G-side covariance parameters, corresponding to the variances associated with the APP_PASS and APP_PASS*HPASS random effects, respectively. The single R-side covariance parameter is the variance of the residual term.

The covariance parameter estimates are given in Table 4. There is significant variability among nitrogen application passes, as well as between harvest passes within each nitrogen application pass.

Table 5 gives tests of the fixed effects. The three-way interaction between EC_ZONE, NRATE, and NMTCD is not quite significant. On the other hand, the two-way interaction between EC_ZONE and NRATE and the two-way interaction between EC_ZONE and NMTCD, are both significant. This implies that the effects of NRATE and NMTCD depend on EC_ZONE, and hence a variable rate treatment prescription involving NRATE and NMCTD would be appropriate for the field.

Figure 3 contains graphics produced by GLIMMIX that are useful for checking distribution assumptions about the random effects. The empirical distribution of the Studentized conditional residuals is fairly symmetric. The tails of the distribution do appear to be heavier than that of a Normal distribution, but other

Table 3. Model dimensions table

Description	Model without radial smoother	Model with radial smoother
G-side covariance parameters	2	3
R-side covariance parameters	1	1
Columns in **X**	50	50
Columns in **Z**	107	702
Subjects (Blocks in **V**)	1	1
Max Observations per Subject	6008	6008

than that the normality assumption does not appear to be violated to a great extent.

Figure 4 shows an empirical semivariogram of the conditional residuals from this model. It clearly shows that observations close together are more similar than observations farther apart. In particular, residuals for observations closer than about 30 distance units are correlated, with residuals for observations closer together being more highly correlated. Residuals for observations that are more than 30 distance units apart are essentially uncorrelated. Note that adjacent observations within a harvest pass are approximately 4 distance units apart, and observations in adjacent harvest passes

Table 4. Covariance parameter estimates.

Model	Covariance Parameter	Point Estimate	Standard Error
Without radial smoother	APASS	693.37	400.55
	APASS*HPASS	1173.55	324.12
	Residual	15649	287.53
With radial smoother	APASS	113.81	307.04
	APASS*HPASS	1242.82	330.23
	Variance of radial smoother	5.95	0.74
	Residual	11129	215.52

Table 5. Type-III tests of fixed effects.

Effect[*]	Without radial smoother p-value	With radial smoother p-value
E	<0.0001	0.8535
N	0.4932	0.3657
E*N	0.0010	0.2425
M	0.0197	0.0118
E*M	0.0001	0.0273
N*M	0.9174	0.3803
E*C*M	0.0643	0.2271
LOC_X	0.0003	0.8995
LOC_Y	<0.0001	0.8232

* E=EC_ZONE, N=NRATE, M=NMTCD

can be as close as 6 distance units. Hence residuals for several observations within a given harvest pass and between several harvest passes may be correlated. From this it is clear that the data violate the assumption of residual independence this model imposes, and therefore any inferences drawn from this model regarding fixed effects are suspect.

We next fit a semi-parametric model that includes a two-dimensional radial smoothing spline. The purpose of the radial smoother is to account for trend across the field so that the residual spatial correlation is significantly reduced or eliminated. The GLIMMIX statements below fit this model.

```
proc glimmix data=yield;
class EC_ZONE NRATE NMTCD APASS HPASS ;
model YLD = EC_ZONE | NRATE | NMTCD
        LOC_X LOC_Y
        / ddfm=satterth;
random APASS;
random APASS*HPASS;
random LOC_X LOC_Y
  / type=rsmooth knotmethod=data(knotinfo) knotinfo;
run;
```

For this analysis 625 equally spaced knots were laid out across a lattice spanning the field. Thirty knots were removed because they occurred in an area where no yield measurements were obtained (i.e., the location of the farm house). This left 595 knots for the analysis, which were stored in a dataset named "knotinfo." In the code above the knots were then read into GLIMMIX using the "data(knotinfo)" option of the "random" statement. The dimensions of the X and Z design matrices for this model are found in Table 3. Recall that in the original model without the radial smoother, there were 107 columns in the Z matrix, corresponding to the 107 random effects in the model. By using a radial

Model without radial smoother Model with radial smoother

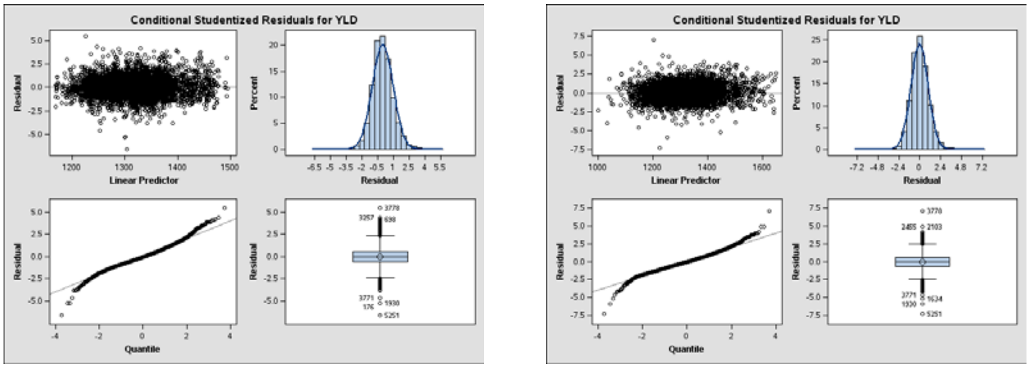

Figure 3. Summary of Studentized conditional residuals.

smoother with 595 knots, we have added 595 random effects to the original model. There are therefore 702 random effects in this model, and hence 702 columns in the Z matrix, as shown in the table. The model without the radial smoother has 2 G-side covariance parameters. The model with the radial smoother has one additional G-side covariance parameter, the variance of the radial smoother, for a total of 3 G-side covariance parameters.

The covariance parameter estimates for this model are given in Table 4. The variation from one nitrogen application pass to another is not significant in this model. On the other hand the variation between harvest passes within nitrogen application passes is still significant. Note also that the estimate of the residual variance in this model is almost 30% less than what it was in the model without the radial smoother.

From Figure 3 we see that, as with the previous model, the empirical

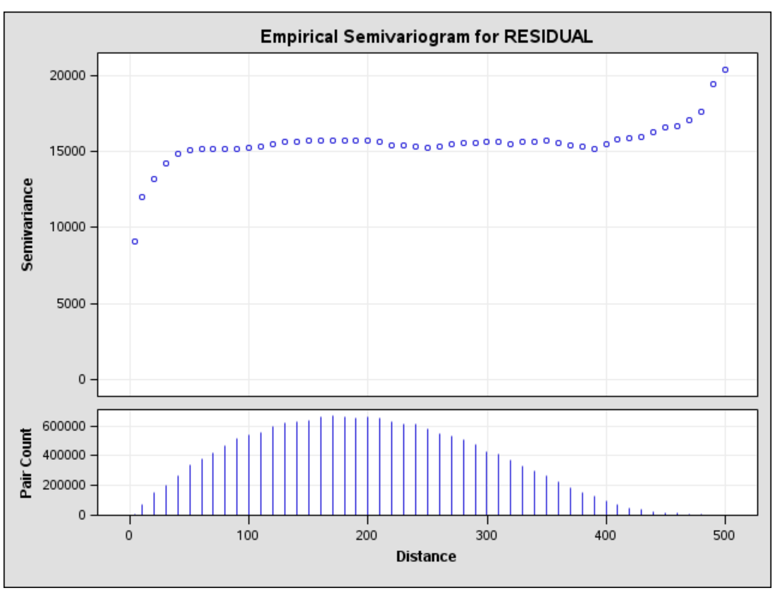

Figure 4. Empirical semivariogram of the conditional residuals for the model without the radial smoother.

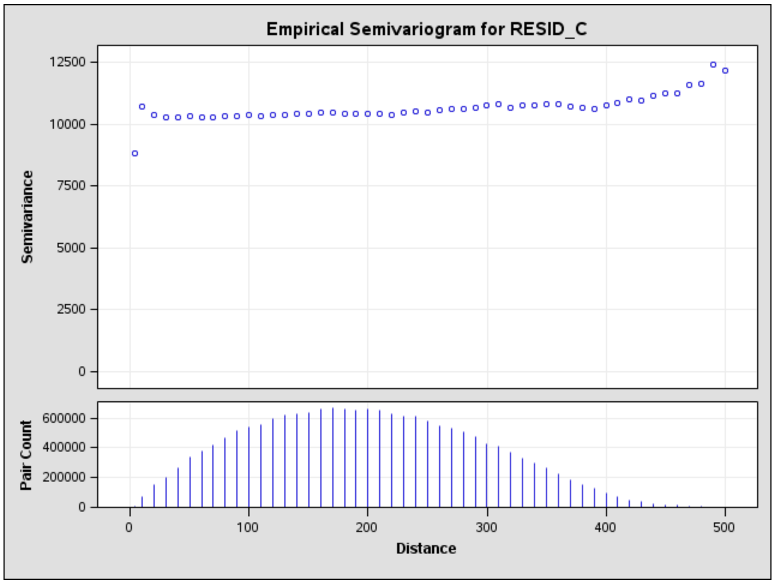

Figure 5. Empirical semivariogram of the conditional residuals for the model with the radial smoother.

distribution of the Studentized conditional residuals has somewhat heavier tails than that of the Normal distribution, but other than that the normality assumption does not appear to be violated to a great extent.

Figure 5 shows an empirical semivariogram for the conditional residuals from this model. At distances greater than about 4 distance units, the value of the semivariogram is very close to the estimated residual variance of 11129 (see Table 4). This implies that residuals separated by more than 4 distance units are uncorrelated, or nearly so. Since adjacent yield points within a harvest pass are separated by about 4 units, on average, and yield points in adjacent harvest passes are separated by at least 6 distance units, we conclude that the conditional residuals are uncorrelated and that the tests involving the fixed effects are reliable.

Table 5 summarizes the tests of the fixed effects in the model. The three-way interaction between EC_ZONE, NRATE, and NMTCD is not significant. The two-way interaction between EC_ZONE and NRATE, which was significant in the model without the radial smoother, is not significant in this model. In fact, none of the effects involving NRATE are significant. Hence NRATE does not appear to be having much of an impact on cotton lint yield on this particular field. On the other hand, the two-way interaction between EC_ZONE and NMTCD and the main effect of NMTCD, both of which were significant in the model without the radial smoother, remain significant in this model. The results of these tests indicate that the prescription involving the nematicide treatment will vary depending on the EC_ZONE, while the prescription for the nitrogen treatment can be a blanket treatment.

SUMMARY AND CONCLUSIONS

Unaccounted-for spatial trend is a source of spatial correlation. To the extent that spatial correlation is due to unaccounted-for trend, it may be possible to reduce or eliminate it by adequately accounting for that trend. These assertions were demonstrated using a simulated one-dimensional dataset. In that example we showed that semi-parametric models incorporating smoothing splines can be quite effective at removing complicated spatial trend, resulting in greatly reduced spatial correlation. The radial smoother in the GLIMMIX procedure of version 9.2 of SAS® is a penalized smoothing spline implemented as a random effect and can be used for this purpose. We applied this methodology to an actual precision agriculture dataset. The data were modeled two ways: first, without accounting for residual spatial correlation, and second, using a radial smoother. The residuals from the model without the radial smoother exhibited significant spatial correlation. In contrast, the residuals from the model with the radial smoother were essentially free of spatial correlation. Inferences involving fixed effects were different for the two models, illustrating the fact that the results of statistical analyses, and hence treatment prescriptions that are based on those results, are affected by whether spatial correlation has been accounted for.

REFERENCES

Burris, E., D. Burns, K.S. McCarter, C. Overstreet, M.C. Wolcott. 2009. Evaluation of the effects of Telone II (fumigation) on nitrogen management and yield in Louisiana delta cotton. Precision Agriculture. DOI 10.1007/s11119-009-9129-x.

McCarter, K.S., E. Burris, G.A. Milliken, E.L. Clawson, H.Y. Wong, J.L. Willers. 2007. Specifications of a prototype software system for developing variable-rate treatment prescriptions for use in precision agriculture. Proceedings of the Nineteenth Annual Kansas State University Conference on Applied Statistics in Agriculture. Manhattan, Kansas. Ed. John E. Boyer.

Milliken, G., D. Johnson. 2002. Analysis of Messy Data, Volume III: Analysis of covariance. Chapman Hall/CRC, Boca Raton, FL.

Milliken, G., D. Johnson. 2009. Analysis of Messy Data, Volume I: Designed experiments. 2nd edition. Chapman Hall/CRC, Boca Raton, FL.

Ruppert, D., M.P. Wand, R.J. Carroll. 2003. Semiparametric regression. Cambridge University Press, New York, NY.

SAS Institute Inc. 2008. *SAS/STAT® 9.2 User's Guide*. Cary, NC: SAS Institute Inc.

Willers, J.L, G.A. Milliken, C.G. O'Hara, and J.N. Jenkins. 2008. Defining the experimental unit for the design and analysis of site-specific experiments in commercial cotton fields. Agricultural Systems. 96: 237-249.