DETERMINING THE EFFECTIVENESS OF INCLUDING SPATIAL INFORMATION INTO A NEMATODE/NUTSEDGE PEST COMPLEX MODEL

by

JOEL VETTER

B.S., Sterling College, 2003

A REPORT

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2012

Approved by:

Major Professor
Leigh Murray

Abstract

An experiment was performed in 2005-2006 to determine if the variety of an alfalfa (*Medicago sativa)* crop rotation can effectively reduce the pest complex consisting of the yellow and purple nutsedge (YNS & PNS) weeds and the southern root-knot nematode (SRKN). During the 2005-2006 growing season, six months were selected to take samples from the alfalfa field (three months in 2005 and three months in 2006). The field was divided into 1m x 2m quadrats. Each month eighty quadrats were randomly selected. The counts of PNS, YNS and a soil sample (analyzed for the count of juvenile SRKN) were taken from each quadrat. In this study, two different ways were examined use spatial information provided from the experiment to alter the original model. First spatial information was treated as fixed effects. Second spatial information was treated as random effects by modifying the residual variance matrix using various "spatial" variance-covariance structures. The results were compared to the original Poisson model and the spatial models to each other but did not have an effective way of comparing random effects models with the fixed effects models. For this data, the use of spatial statistics did not improve the original model consistently. This may be partly because of the nature of the experiment. The alfalfa effectively reduced the YNS, PNS, and SRKN counts. The spatial information was generally more useful earlier in the experiment when the YNS, PNS, and SRKN populations were denser.

Table of Contents

List of Figures

List of Tables

Chapter 1 - Introduction

The southern root-knot nematode (SRKN) is a microscopic plant parasite that attacks the roots of its host. Previous studies have shown that the SRKN has developed a mutually beneficial relationship with the weeds, yellow nutsedge (YNS) and purple nutsedge (PNS) (Schroeder et al., 1994; Thomas et al., 1997; Schroeder et al., 2004; Thomas et al., 2004; Schroeder et al., 2005; Thomas et al., 2005). This pest complex is active primarily in the southern and western United States and affects cotton and chile pepper crops, among others. Targeting parts of this pest complex individually hasn't been historically successful (Schroeder et al., 1994, 2004; Thomas et al., 2005). This research was the basis for an experiment to examine if YNS & PNS counts could be used to predict juvenile SRKN counts.

The previous work was the basis of an alfalfa rotation study to examine potential control of the pest complex. An additional objective was to determine if YNS and PNS plant counts could be used to predict SRKN juvenile counts from soil samples as described in Ou et al (2008). This paper describes an extension to the Ou et al. (2008) by evaluating the addition of spatial information to the models.

Alfalfa Rotation Experiment

The alfalfa rotation experiment began in September 2004 at the Leyendecker Plant Science Research Center, New Mexico State University (see Ou et al., 2008 for details of this experiment). An SRKN-resistant alfalfa plant was rotated into a well-prepared field that was heavily infested with the aforementioned pest complex. In addition to being SRKN-resistant, the alfalfa plant competes well for light and other resources against the YNS and PNS weeds. By rotating this alfalfa crop into this infested field, the researchers were targeting the pest complex as a whole with a financially-viable crop instead of using expensive chemicals to attack individual parts of the pest complex.

The experiment continued through two growing seasons and ended in October 2006. The field was flood irrigated once a month from February to September (Ou et al., 2008). The 50x110 meter field was divided into a grid of 50 by 55 plots. The plots themselves measured 1x2 meters. In May, July, and September of both 2005 and 2006, eighty plots were randomly selected to take YNS, PNS and SRKN counts. The (x, y) grid coordinates of each selected plot was recorded along with the sample counts. A 0.25- ×1-m quadrat was laid down in center of the selected plot and the number of YNS and PNS plants was counted. Ten 50-cm^3 soil samples were also taken either at the base of existing nutsedge plants in the quadrat or at random points in the quadrat if no nutsedge plants were present. The location of each grid on an (x, y) coordinate plane was recorded along with the sample counts. The soil samples were processed by elutriation in order to estimate the number of SRKN juveniles present in the soil (Ou et al., 2008).

The experiment was overall successful. As the alfalfa crop grew, the pest-complex generally fell apart. With a few exceptions, the data shows a decreasing number of PNS, YNS, & SRKN counts during the course of the experiment (Ou et al., 2008).

Chapter 2 - Previous Statistical Analysis

Because of the difficulty and expense of obtaining SRKN counts, the analysis of the data (Ou et al., 2008) from the experiment sought to discover if YNS and PNS counts could be reliable predictors of SRKN counts. Due to the count nature of the data, a Generalized Linear Model approach was used. A separate model was fit for each month's worth of data using the Poisson probability distribution

$$f(y, \theta) = \frac{\theta^{y} e^{-\theta}}{y!}$$

with the log link-function relating the linear predictor with the SRKN counts. Explanatory variables included in the linear predictor were YNS and PNS counts, and their squares and crossproduct:

$$\ln(\theta) = \beta 0 + \beta 1(YNS) + \beta 2(PNS) + \beta 3(YNS*PNS) + \beta 4(YNS2) + \beta 5(PNS2)$$

All six fitted models had problems of over-dispersion. To handle the issue of over-dispersion, a re-scaling approach was also examined (Ou et al., 2008, McCullagh and Nelder 1989).

The results of the data analysis were varied (Table 2.1). In the first year of the experiment the infestation of the pest complex was denser. In May of the first year, the linear PNS variable was the only significant predictor. It was speculated that perhaps it was too early in the growing season for YNS to establish itself in the field. In July and September 2005 the linear variables, YNS and PNS, were significant positive predictors for SRKN counts while the YNS*PNS term was negative. The following year, as the pest complex was diminished, the nutsedge counts were not significant predictors for SRKN counts in May and July. In September of the second year, PNS was a significant predictor. Perhaps, the linear variable PNS had regained some of its prior losses as the year continued and perhaps provided an adequate host to the SRKN by the end of the growing season. During the second year, the sample size of 80 might not have been large enough to pick up the relationship with the diminished pest-complex (Ou et. al, 2008).

Table 2.1: Regression Models obtained by Ou et al. (2008)

Month	Predictors(not including X and Y coordinates)
May 2005	PNS
July 2005	YNS, PNS, YNS*PNS(interaction)
September 2005	YNS, PNS, YNS*PNS(interaction)
May 2006	Intercept only[1]
July 2006	Intercept only[1]
September 2006	PNS
Full Model	$\ln(\theta) = \beta_0 + \beta_1(YNS) + \beta_2(PNS) + \beta_3(YNS*PNS) + \beta_4(YNS^2) + \beta_5(PNS^2)$

1. YNS and PNS predictors were re-added to this model for the spatial data study.

Chapter 3 - Current Statistical Analysis

The objectives of this paper is to re-examine the Generalized Linear Model (proposed by Ou et al., 2008) to see if using the available spatial information in the model helped alleviate the problem of over-dispersion. The Poisson model is over-dispersed when its variance exceeds its mean. Over-dispersion is caused by missing explanatory variables to explain the variability in the response (McCullagh and Nedler, 1989). Including the spatial information might remove some of the over-dispersion.

The field was flood irrigated along the y axis (Figure 3.1). It is suggested that the SRKN might move along the y-axis because of the flood irrigation but the movement along the x-axis would be minimal (Murray et al., 2012). For example, if there was high SRKN count in location A in Figure 3.1, it was expected that there would more likely be another high count of SRKN somewhere further along the y-axis, perhaps in location B but not necessarily in location C.

The spatial information was applied in two ways. First the x and y coordinates were included with YNS and PNS as fixed-effect predictors in the Poisson model. Thus an additional two parameters were added to the model. For example, for July 2005, the model fitted here was:
$\ln(\theta) = \beta_0 + \beta_1(\text{YNS}) + \beta_2(\text{PNS}) + \beta_3(\text{YNS*PNS}) + \boldsymbol{\beta_4}(\text{X-Coordinate}) + \boldsymbol{\beta_5}(\text{Y-Coordinate})$.
The exception was made for the intercept only models (May06 and July06). Both YNS and PNS were included to test if the addition of the x and y coordinates made a difference. Table 2.1 contains the models from Ou et al. (2008) and the models used in this study for each particular month.

Including the spatial data as fixed effects allows for a simple interpretation of the results. For example, if the estimate for β_5 is positive in the model above, the expected number of SRKN would increase as the value of the y-coordinate increased, with all other factors held constant. If the flood irrigation theory is correct, a positive estimate for the y-coordinate parameter would be expected and the x-coordinate parameter would be insignificant (close to zero).

Models with spatial information modeled as fixed effects were fitted using the GLIMMIX procedure of SAS (version 9.2, http://support.sas.com/documentation/onlinedoc/stat). See Appendix A for the code for May05.

The second approach included the spatial data as a random component by modifying the residual variance matrix using various spatial formats. For example, a basic model in linear regression for two plots is as follows:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 x_1 \\ \beta_0 + \beta_1 x_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

Then $E = \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$ the error or residual vector and is generally assumed to have variance-covariance matrix $\Sigma = \sigma^2 I$, where I is the 2x2 identity matrix. Thus each of the two e_i has a variance of σ^2 and a covariance between e_1 and e_2 of zero. By modifying the matrix I, the residuals can include a non-zero covariance based on distance. This means that observations (e.g. plots) closer to each are generally more highly correlated than observations that are farther apart. Let d_{ij} be the distance between observations i and j and $f(d_{ij})$ be a function of the distance. Replace the identity matrix I with $f(d_{ij})$ and put on a constraint that $f(d_{ij})$ equals 1 when i=j. The new residual variance-covariance matrix that includes the spatial data becomes:

$$\Sigma = \begin{bmatrix} \sigma^2 & f(d_{ij}) \\ f(d_{ij}) & \sigma^2 \end{bmatrix}$$

The variance remains the same but the covariance becomes a function of the distance. Three different types of spatial covariance structures were used in the analysis (Table 3.1). As explained in the results, semi-variograms were used to determine the best spatial structure when spatial information was treated as a random effect (Cressie 1991). Semi-variograms were obtained using the SAS VARIOGRAM procedure, then regression models were fitted to the semi-variograms using the NLIN procedure (see Appendix B). Finally, mixed models were fitted to the data using the GLIMMIX procedure (see Appendix C). Yellow nutsedge and purple nutsedge counts were still treated as fixed predictor variables while x and y coordinates were used in the spatial covariance structure. All analyses were conducted using SAS version 9.2 (http://support.sas.com/documentation/onlinedoc/stat).

Results

Summaries of all results are included in Appendix D.

Spatial coordinates as fixed effects

Because of the nature of the experiment, the results of the analysis change from month to month (Ou et. al 2008; Murray et. al 2012). At the beginning of the data collection (May 2005), the field had a high infestation of the YNS, PNS and SRKN pest-complex. By the end of the experiment (September 2006), the YNS, PNS, and SRKN counts were considerably decreased. Thus, the experimental field was in much different condition at the beginning of the experiment then it was at the end (Ou et. al 2008; Murray et. al 2012).

In opposition to a-priori belief, the x-coordinate was generally a significant predictor during the first year (2005). It was significant at the .05 level for May and September and nearly significant at the .10 level for July (Table 3.2). It was also consistently a negative predictor all three months. This suggest as one moved from point C (Figure 3.1) to point A, the expected number of SRKN present would increase if all other variables were held constant. While this was an interesting discovery, there was no reasonably explanation about why this was occurring in this particular field. The y-coordinate was not significant in the 2005 months.

The following year (2006), the pest-complex was decreasing because of the competitive alfalfa. The relationship between the x-coordinate and the SRKN count disappeared. In other words, the x-coordinate was not significant in the 2006 months. However as time increased in 2006, the y-coordinate significance increased. The y-coordinate's p-value in May 2006 was .7796, in July 2006 .1605, and in September 2006 .0013 (Table 3.2). The y-coordinate estimate in September 2006 was positive. This suggests that the flood irrigation was having the effect that was expected on SRKN counts. However, this is the only instance of the y-coordinate being significant and is not strong evidence that the flood irrigation was associated with the SRKN counts in any way. It would have been interesting to see if this trend had continued if the experiment was extended another year in this field.

The magnitude of parameter estimates and standard errors for the YNS, PNS and interaction terms from the original model were not affected substantially by the inclusion of the spatial data as fixed effects (Table 3.2). The parameter estimates both decreased and increased during certain months with no observable pattern. The same held true for the standard errors. The parameters in general maintained the same level of significance as in Ou et. al (2008). However, the interaction term in July 05 was affected enough to raise its p-value from .0892 to .1164 making its inclusion in the July 05 spatial model more questionable. Similarly, the PNS parameter in the September 05 model p-value jumped from .0309 to .0868 which is no longer significant at the .05 level.

Overall, including the x and y coordinates as fixed effects did not improve the original model consistently enough to justify including them. When comparing the original models' AIC to the spatial models' AIC for each month, three out of the six months (May 05, Sept 05, Sept 06) the model fit improves (Table 3.3). For the other three months the AIC increases when including the spatial information. As expected, the three months that have decreased AIC are also the same three months in which either the x or y coordinate are significant at the .05 level.

The x and y coordinates also did not consistently help alleviate the problem of over dispersion in the Poisson model (Table 3.3). For fixed effects generalized linear models, Pearson's chi square statistic divided by the degrees of freedom should be close to one if the Poisson model is a good fit. A value greater than one indicates over-dispersion, and a value less than one indicates under-dispersion. With the spatial parameters included in the model, Pearson's statistic divided by the degrees of freedom had similar results to the AIC. The same three months decreased (May 05 , Sept 05, Sept 06) indicating only a slight decrease in overdispersion. The other three months increased indicating only a slight increase in over-dispersion.

Spatial coordinates as random effects

Examining the spatial coordinates as random effects highlighted the effect the alfalfa was having on the pest-complex. Semi-variograms were used to determine the type of spatial structure that might best fit the data (Cressie 1991). Semi-variograms were plotted for each month to visualize the correlation between distance and SRKN count (Figure 3.2). The months that have the most reasonable spatial structure are May 2005 and May 2006. The other months have little to no correlation between distance and SRKN count. This might suggest that the alfalfa was effectively destroying the pest-complex relationship as early as July 2005.

Based on the May 05 and May 06 semi-variograms , the three spatial structures (in Table 3.1) were chosen because their theoretical semi-variograms appeared most likely to fit the spatial structure (Cressie 1991). Regression lines were fit to the May 05 and May 06 semi-variogram for the three spatial structures to see which one would be expected to perform the best (Figure 3.3). Of the three regression lines, the exponetial spatial structure fit slightly better in May 05. In May 2006, all three spatial structures essentially degenerated into linear approximations. The power and exponential spatial structures fit slightly better than the Gaussian. In both months, none of the spatial structures appeared to be a bad fit.

The estimates for the covariance parameters were generally low. The exponential structure in particular was often estimated to be approximately zero (Table 3.4). The Gaussian and power spatial structure usually estimated a higher covariance than the exponential.

Overall, including the spatial coordinates as random effects unnecessarily complicated the model. All of the estimates for the YNS, PNS, and interaction in the random effects models were very similar if not exactly the same as the original models. The standard errors increased across the board for the YNS, PNS, and interaction regression coefficients estimates (The May 05 estimates are given as in example in Table 3.5). In some cases, the model never converged. None of the spatial structures stood out as being more effective than the others.

There was difficulty comparing model fit for the original model to the random effects model. The AIC to pseudo-AIC comparison is included in the table but should not be used as real evidence in support of one model over the other (Schabenberger 2005). In fact, the only month that had a decreased pseudo-AIC (for all spatial structures) compared to original model AIC was May 05 (Table 3.6). This also is the month that showed the best spatial correlation according to the semi-variogram.

Conclusions

Using the available spatial information did not consistently improve the original models. For the fixed effects models, the expectation that the y-coordinate would be a more significant predictor than the x-coordinate because of the flood irrigation was not met except in the final month of the experiment. For the random effects models, there was no improvement on the YNS, PNS, & YNS*PNS predictors, and there was not a clear approach to determine if the model fit improved.

A possible reason for the lack of success could be the small sample size. Especially as time increased and the YNS, PNS, and SRKN counts decreased, the sample size of 80 out of 2750 grids may not have been large enough to pick up any kind of relationship that distance and SRKN counts might have had. Because of the logistics of collecting soil samples and measuring SRKN counts, sample sizes much larger than 80 weren't feasible in the original experiment. As mentioned earlier, the nature of the experiment could be another reason for the lack of success.

The alfalfa may have effectively destroyed any kind of spatial relationship early in the experiment.

It would be interesting to collect the same data from a field that the pest-complex competes more aggressively (i.e. chile pepper or cotton field) to see if a spatial relationship would be more prominent in such a situation. As with most situations, any increase in sample size would also be beneficial to establishing more concrete results.

Figure 3.1: Flood Irrigation

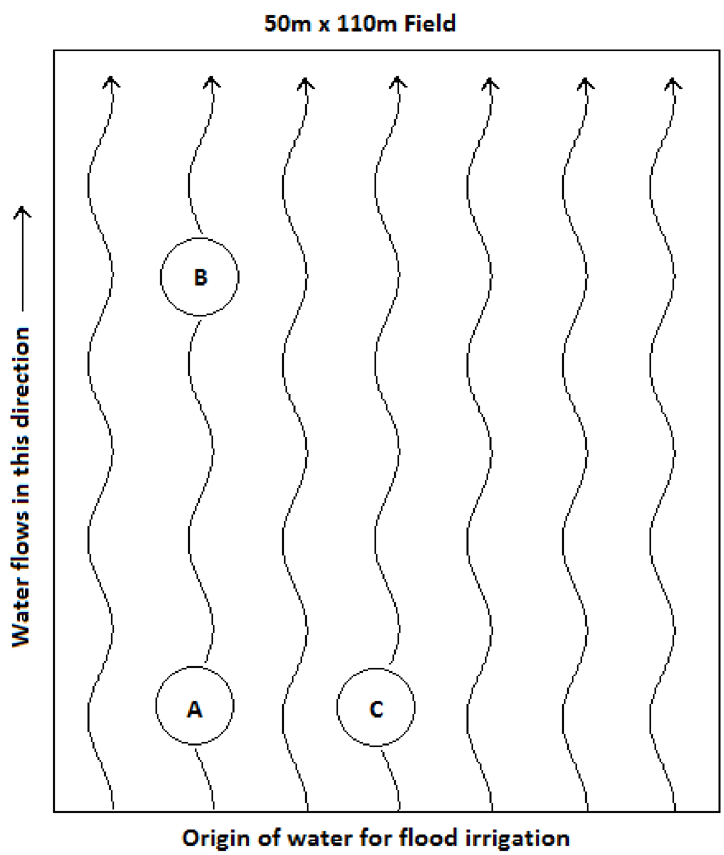


Figure 3.2: Empirical semi-variograms

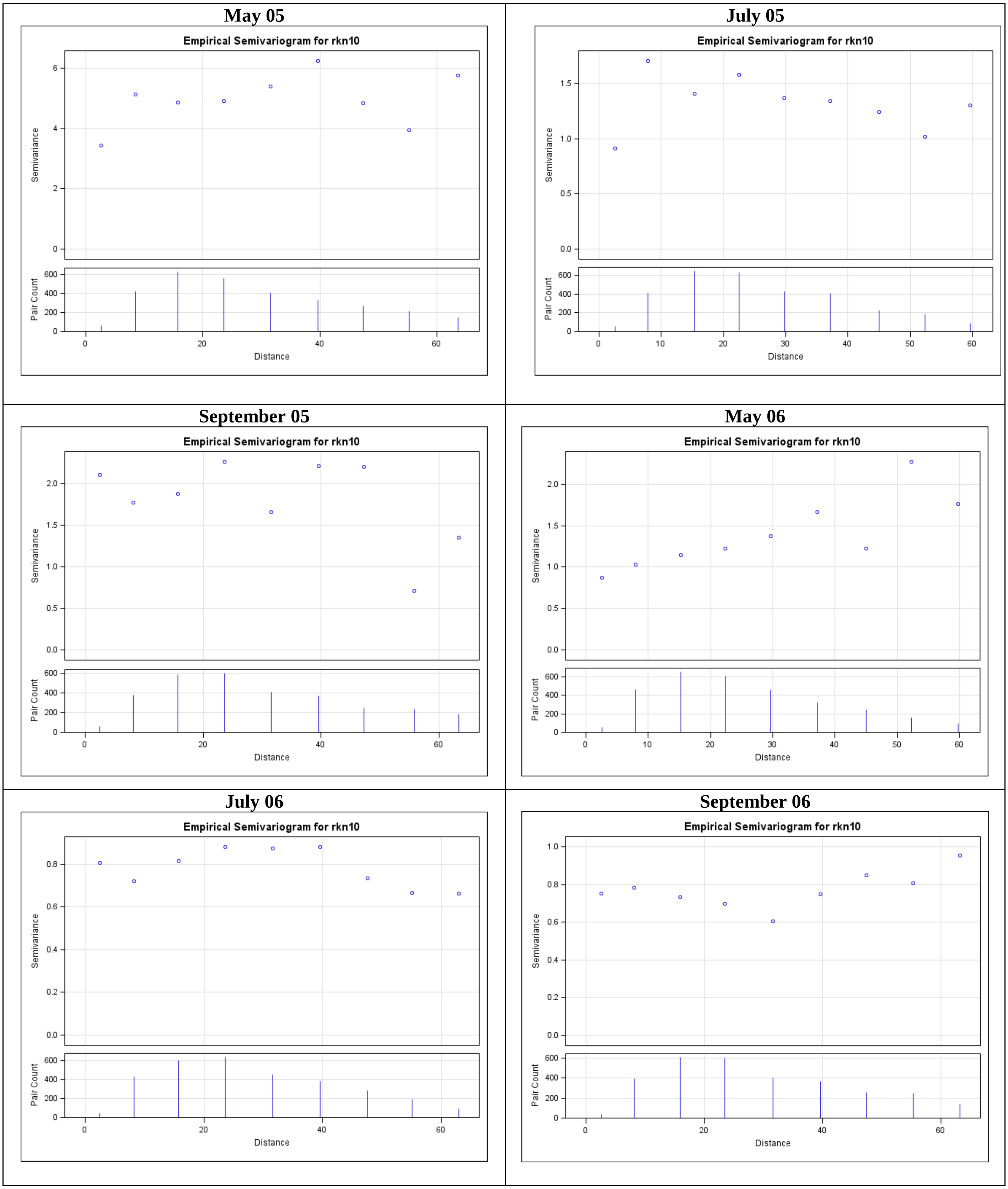
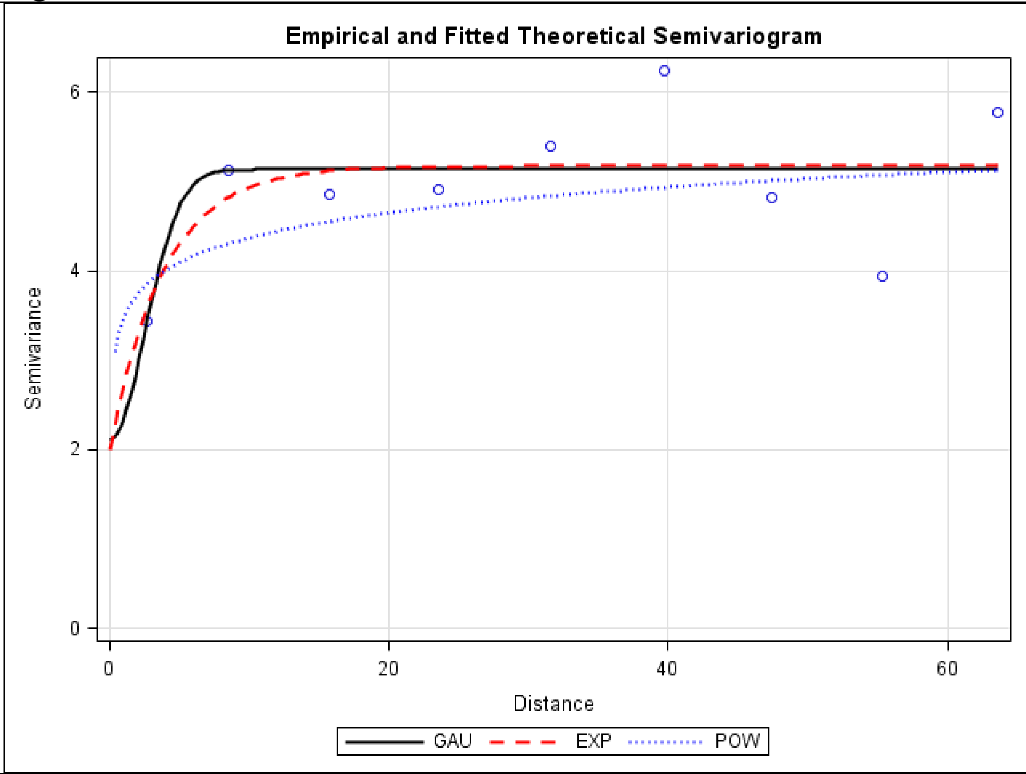
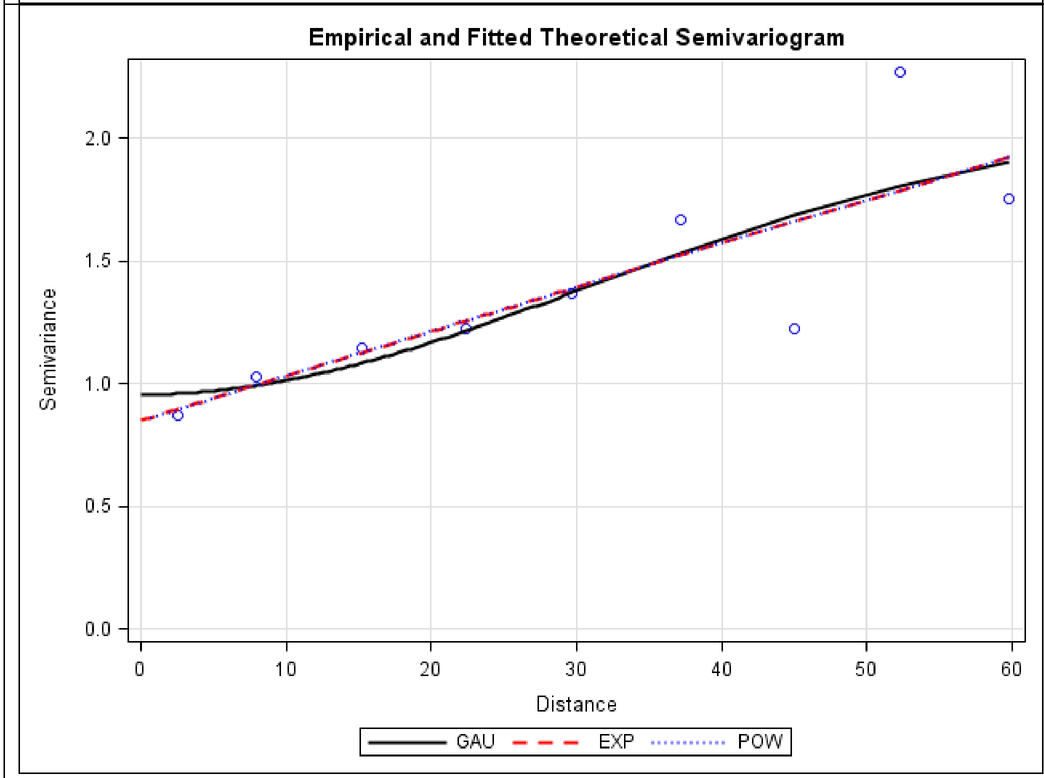


Figure 3.3: May 05 and May 06, Empirical semi-variogram with fitted theoretical regression lines



May 05



May 06

Table 3.1: Spatial Variance-Covariance Structure

Spatial Structure*	$f(d_{ij})$
Exponential	$\sigma^2 \exp\{-d_{ij}/\alpha\}$, α is constrained to be positive
Gaussian	$\sigma^2 \exp\{-d_{ij}^2/\alpha^2\}$, α is constrained to be positive
Power	$\sigma^2 \rho^{d_{ij}}$, where $\rho \geq 0$

d_{ij} is the Euclidean distance between the x and y coordinates of observations i and j.
*All spatial structures are parameterized using SAS Documentation v9.2

Table 3.2: Regression Parameter Estimates and (p-values) for Fixed Effects Models

Month	YNS	PNS	YNS*PNS	X-Coord	Y-Coord
May 05 (p-value)	 -	0.4359 (.0008)	 -	-0.0312 (.0053)	0.0038 (.2817)
July 05 (p-value)	0.081 (.0082)	0.1555 (.0237)	-0.0149 (.1164)	-0.0288 (.1006)	0.0054 (.3508)
Sept 05 (p-value)	0.4596 (.0055)	0.1242 (.0868)	-0.1242 (.0334)	-0.038 (.0279)	0.0067 (.2585)
May 06 (p-value)	0.4328 (.1545)	0.5072 (.6259)	 -	0.0035 (.8579)	-0.0019 (.7796)
July 06 (p-value)	-0.0755 (.4480)	0.0672 (.3535)	 -	0.0142 (.4434)	-0.0104 (.1605)
Sept 06 (p-value)	 -	0.2658 (.0099)	 -	0.0037 (.8695)	0.0216 (0.013)

Table 3.3: Model Fit Statistics for Fixed Effects Models

Month	**AIC Original Model**	**AIC w/Spatial Data**	**P-Chi/df Original Model**	**P-Chi/df w/Spatial Data**
May 05	329.60	324.05	2.27	2.19
July 05	219.27	219.81	1.42	1.43
September 05	213.88	211.61	2.01	1.94
May 06	200.61	204.49	2.06	2.15
July 06	170.61	172.10	1.53	1.65
September 06	141.58	138.32	1.80	1.48

Table 3.4: Covariance Parameter Estimates

Month	EXP(α)	POW(ρ)	GAU(α)
May 05	≈ 0	0.5111	1.917
July 05	≈ 0	0.4753	0.8853
Sept 05	0.0509	DNC	0.216
May 06	≈ 0	0.4183	1.3539
July 06	DNC	DNC	DNC
Sept 06	0.0427	-0.0237	0.2473

Table 3.5: May 05 PNS parameter information

	May 2005		
Model	**PNS Estimate**	**PNS Standard Error**	**PNS P-value**
Original Model	0.5099	0.1253	0.0001
Exponential	0.5099	0.1886	0.0084
Power	0.4598	0.1894	0.0175
Gaussian	0.5033	0.1825	0.0073

Table 3.6: Model Fit Statistics for Random Effects Models

	AIC	Pseudo-AIC for given Spatial Structure		
Month	**Original Model**	**Exponential**	**Power**	**Gaussian**
May 05	329.6	247.13	242.26	242.34
July 05	219.27	290.13	292.07	292.45
September 05	213.88	327.78	DNC	327.78
May 06	200.61	321.16	323.91	326.60
July 06	170.61	DNC	DNC	DNC
September 06	141.58	365.16	365.28	365.16

References

Cressie, N. A. C. 1991. *Statistics for Spatial Data.* New York: John Wiley & Sons, Inc.

McCullagh P, Nelder JA. New York: Chapman and Hall; 1989. Generalized linear models.

Murray L., Thomas, S. H., Schroeder, J., Ou, Z., Trojan J.M., and Fiore C. 2012. Modeling the Root-Knot Nematode/Nutsedge Pest Complex: Perspectives from Weed Science, Nematology, and Statistics. *KSU Conference in Applied Statistics in 2012.*

Ou, Z., Murray, L., Thomas, S. H., Schroeder, J., and Libbin, J. 2008. Nutsedge Counts Predict *Meloidogyne incognita* Juvenile Counts in an Integrated Management System. *Journal of Nematology* 40(2):99-108.

Schabenberger, O. 2005. Introducing the GLIMMIX Procedure for Generalized Linear Mixed Models. SAS Global Users Group: Paper 196-30.

Schroeder, J., Kenney, M. J., Thomas, S. H., and Murray, L. W. 1994. Yellow nutsedge response to southern root-knot nematode, chile pepper, and metolachlor. Weed Science 42:534–540.

Schroeder, J., Thomas, S. H., and Murray, L. W. 2004. Root-knot nematodes affect annual and perennial weed interactions with chile pepper. Weed Science 52:28–46.

Schroeder, J., Thomas, S. H., and Murray, L. W. 2005. Impacts of crop pests on weeds and weed-crop interactions. Weed Science 53:918–922

Thomas, S. H., Schroeder, J., Kenney, M. J., and Murray, L. W. 1997. *Meloidogyne incognita* inoculum source affects host suitability and growth of yellow nutsedge and chile pepper. Journal of Nematology 29:404–410.

Thomas, S. H., Schroeder, J., and Murray, L. W. 2004. *Cyperus* tubers protect *Meloidogyne incognita* from 1,3-dichloropropene. Journal of Nematology 36:131–136.

Thomas, S. H., Schroeder, J., and Murray, L. W. 2005. The role of weeds in nematode management. Weed Science 53:923–928.

Appendix A - SAS Code for Fixed Effects Spatial Models for May 05

```
data May2005;
input sample  xcoord  ycoord   yns     pns     rkn100    rkn10;
ycoord = ycoord*2;
cards;
   1       2       14      0       0       10      1
   1       2       31      0       0       30      3
.
.
.
;

*Original model No Spatial Data;
title 'No Spatial Data';
proc glimmix data = May2005;
      model rkn10 = pns            / dist = poisson
                  link = log solution;
run;

*Fixed effects model;
title 'Fixed Effects (X,Y)';
proc glimmix data = May2005;
      model rkn10 = pns xcoord ycoord           / dist = poisson
                  link = log solution;
run;
```

Appendix B - SAS Code for May 05 Semi-variogram and Theoretical Regression Lines

```
data May2005;
input sample  xc  yc   yns     pns     rkn100    rkn10;
ycc = yc*2;
cards;
   1       2      14      0       0       10      1
   1       2      31      0       0       30      3
.
.
.
;

data May2005;
      set May2005;
      yc = yc*2;

ods graphics on;

*Initial analysis to determine approriate lag distance and max lags;
proc variogram data=May2005;
coordinates xcoord=xc ycoord=yc;
compute novariogram;
var rkn10;
run;

*Fitting the Empirical semi-variogram;
proc variogram data=May2005;
coordinates xcoord=xc ycoord=yc;
compute lagdist=8 maxlags=8 ;
ods output SemivariogramTable=sv;
var rkn10;
run;

data pv;
do Distance = 0 to 60 by 0.3;
Semivariance = .;
output;
end;
run;

data sv; set sv pv; by Distance;
run;

*Fitting Gaussian theoretical semi-variogram to the empirical;
proc nlin data=sv;
      parms Range=5
            Sill=2.5
            Nugget=2;
      model Semivariance =
            Nugget + Sill*(1-exp(-(Distance*Distance)/(Range*Range)));
```

```
      output out=GAU p=GAU;
   run;

*Fitting Exponential theoretical semi-variogram to the empirical;
   proc nlin data=sv;
      parms Range=5
            Sill=2.5
            Nugget=2;
      model Semivariance =
            Nugget + Sill*(1-exp(-Distance/Range));
      output out=EXP p=EXP;
   run;

*Fitting Power theoretical semi-variogram to the empirical;
proc nlin data=sv;
      parms Range=0
            Sill=0
            Nugget=0;
      model Semivariance =
            Nugget + Sill*(Distance**Range);
     output out=POW p=POW;
   run;

   data pv;
      merge GAU EXP POW;
   run;

*Plotting results;
 proc sgplot data=pv;
      title "Empirical and Fitted Theoretical Semivariogram";
      xaxis label = "Distance" grid;
      yaxis label = "Semivariance" grid;
      scatter y=Semivariance x=Distance /
              markerattrs = GraphData1(symbol=circle)
              name = 'SemiVarClassical';
      series x=Distance y=GAU /
             lineattrs = (thickness=2px color=black)
             name = 'Gaussian';
        series x=Distance y=EXP /
             lineattrs = (thickness=2px color=red
                          pattern=Dash)
             name = 'Exponential';
      series x=Distance y=POW /
             lineattrs = (thickness=2px color=blue
                          pattern=Dot)
             name = 'Power';
      discretelegend 'SemiVar' 'Gaussian'
                     'Exponential' 'Power';
   run;

   ods graphics off;
```

Appendix C - SAS Code for Random Effects Spatial Models for May 05

```
data May2005;
input sample  xcoord  ycoord   yns     pns     rkn100    rkn10
;
cards;
   1       2       14      0       0       10       1
   1       2       31      0       0       30       3
.
.
.
;

data May2005;
      set May2005;
      ycoord = ycoord*2;
      ynspns = yns*pns;

*Original model No Spatial Data;
      title 'No Spatial Data';
proc glimmix data = May2005 ic=pq asycov ;
      model rkn10 = pns            / dist = poisson
            link = log solution;
run;

*Random Effects - Exponential;
      title 'Random Effects, X,Y, EXP';
proc glimmix data = May2005 ic=pq asycov;
      model rkn10 = pns            / dist = poisson
                  link = log solution;
      random resid /type=SP(EXP)(xcoord ycoord);

run;

*Random Effects - Power;
      title 'Random Effects, X,Y, POW';
proc glimmix data = May2005 ic=pq asycov;
      model rkn10 = pns            / dist = poisson
                  link = log solution;
      random resid /type=SP(POW)(xcoord ycoord);

run;

*Random Effects - Gaussian;
title 'Random Effects, X,Y, GAU';
proc glimmix data = May2005 ic=pq asycov;
      model rkn10 = pns            / dist = poisson
                  link = log solution;
      random resid /type=SP(GAU)(xcoord ycoord);

run;
```

Appendix D - Complete Tables of Results

May '05
Fixed Effects

Model	YNS	YNS(SE)	PNS	PNS(SE)	YNS*PNS	YNS*PNS (SE)	X-Coord	X(SE)	Y-Coord	Y(SE)
No Spatial Data	-	-	0.5099 (.0001)	0.1253	-	-	-	-	-	-
X,Y Coordinate	-	-	0.4359 (.0008)	0.1251			-0.0312 (.0053)	0.0109	0.0038 (.2817)	0.0035

July '05
Fixed Effects

Model	YNS	YNS(SE)	PNS	PNS(SE)	YNS*PNS	YNS*PNS (SE)	X-Coord	X(SE)	Y-Coord	Y(SE)
No Spatial Data	0.089 (.0031)	0.0292	0.1266 (.0436)	0.0617	-0.0148 (.0892)	0.0086	-	-	-	-
X,Y Coordinate	0.081 (.0082)	0.0298	0.1555 (.0237)	0.0673	-0.0149 (.1164)	0.0094	-0.0288 (.1006)	0.0173	0.0054 (.3508)	0.0058

Sept '05
Fixed Effects

Model	YNS	YNS(SE)	PNS	PNS(SE)	YNS*PNS	YNS*PNS (SE)	X-Coord	X(SE)	Y-Coord	Y(SE)
No Spatial Data	0.3879 (.0132)	0.1529	0.1604 (.0309)	0.073	-0.1184 (.0417)	0.0572	-	-	-	-
X,Y Coordinate	0.4596 (.0055)	0.1608	0.1242 (.0868)	0.0715	-0.1242 (.0334)	0.0573	-0.038 (.0279)	0.0169	0.0067 (.2585)	0.0059

May '06
Fixed Effects

Model	YNS	YNS(SE)	PNS	PNS(SE)	YNS*PNS	YNS*PNS (SE)	X-Coord	X(SE)	Y-Coord	Y(SE)
No Spatial Data	0.4286 (.1289)	0.2792	0.4459 (.6602)	1.01	-	-	-	-	-	-
X,Y Coordinate	0.4328 (.1535)	0.3002	0.5072 (.6259)	1.036	-	-	0.0035 (.8579)	0.0192	-0.0019 (.7796)	0.0069

July '06
Fixed Effects

Model	YNS	YNS(SE)	PNS	PNS(SE)	YNS*PNS	YNS*PNS (SE)	X-Coord	X(SE)	Y-Coord	Y(SE)
No Spatial Data	- 0.0795 (.3947)	0.0929	0.0536 (.4530)	0.0711	-	-	-	-	-	-
X,Y Coordinate	- 0.0755 (.4480)	0.099	0.0672 (.3535)	0.072	-	-	0.0142 (.4434)	0.0185	-0.0104 (.1605)	0.0074

Sept '06
Fixed Effects

Model	YNS	YNS(SE)	PNS	PNS(SE)	YNS*PNS	YNS*PNS (SE)	X-Coord	X(SE)	Y-Coord	Y(SE)
No Spatial Data	-	-	0.2617 (.0073)	0.095	-	-	-	-	-	-
X,Y Coordinate	-	-	0.2658 (.0099)	0.1005	-	-	0.0037 (.8695)	0.0224	0.0216 (.0130)	0.0085

May '05
Random Effects

Model	YNS	YNS(SE)	PNS	PNS(SE)	YNS*PNS	YNS*PNS(SE)
No Spatial Data	-	-	0.5099 (.0001)	0.1253	-	-
EXP	-	-	0.5099 (.0084)	0.1886		
POW	-	-	0.4598 (.0175)	0.1894		
GAU	-	-	0.5033 (.0073)	0.1825		

July '05
Random Effects

Model	YNS	YNS(SE)	PNS	PNS(SE)	YNS*PNS	YNS*PNS(SE)
No Spatial Data	0.089 (.0031)	0.0292	0.1266 (.0436)	0.0617	-0.0148 (.0892)	0.0086
EXP	0.0889 (.0124)	0.0347	0.1266 (.0888)	0.0734	-0.0148 (.1523)	0.0102
POW	0.0872 (.0168)	0.0357	0.1413 (.0600)	0.074	-0.0168 (.1131)	0.0105
GAU	0.0898 (.0130)	0.0353	0.1335 (.0762)	0.0742	-0.0161 (.1288)	0.0105

Sept '05
Random Effects

Model	YNS	YNS(SE)	PNS	PNS(SE)	YNS*PNS	YNS*PNS(SE)
No Spatial Data	0.3879 (.0132)	0.1529	0.1604 (.0309)	0.073	-0.1184	0.0572
EXP	0.3879 (.0775)	0.2167	0.1604 (.1250)	0.1034	-0.1184 (.1480)	0.081
POW	Did not converge					
GAU	0.3879 (.0775)	0.2167	0.1604 (.1250)	0.1034	-0.1184 (.1480)	0.081

May '06
Random Effects

Model	YNS	YNS(SE)	PNS	PNS(SE)	YNS*PNS	YNS*PNS(SE)
No Spatial Data	0.4286 (.1289)	0.2792	0.4459 (.6602)	1.01	-	-
EXP	0.4286 (.2877)	0.4003	0.4459 (.7590)	1.4483		
POW	0.4314 (.2674)	0.3862	0.6372 (.6309)	1.3208	-	-
GAU	0.4724 (.1909)	0.358	0.5282 (.7134)	1.4332	-	-

July '06
Random Effects

Model	YNS	YNS(SE)	PNS	PNS(SE)	YNS*PNS	YNS*PNS(SE)
No Spatial Data	-0.0795 (.3947)	0.0929	0.0536 (.4530)	0.0711	-	-
EXP	Did not Converge					
POW						
GAU						

Sept '06
Random Effects

Model	YNS	YNS(SE)	PNS	PNS(SE)	YNS*PNS	YNS*PNS(SE)
No Spatial Data	-	-	0.2617 (.0073)	0.095	-	-
EXP	-	-	0.2617 (.0431)	0.1273		
POW	-	-	0.2611 (.0438)	0.1274	-	-
GAU	-	-	0.2617 (.0431)	0.1273	-	-