SPATIOTEMPORAL MODELING OF

AGRICULTURAL YIELD MONITOR DATA[1]

by

Adela Nistor, Raymond J.G.M. Florax,
Jess Lowenberg-DeBoer, J. P. Brown

Working Paper # 08-01

January 2008

Department of Agricultural Economics

Purdue University

[1] We thank Paul Elhorst for his Matlab routines, suggestions and constructive comments. We gratefully acknowledge the United States Department of Agriculture-Cooperative State Research and Extension Service Grant 2004-51130-03111 entitled "Drainage Water Management Impacts on Watershed Nitrate Load, Soil Quality and Farm Profitability" for supporting this research.

SPATIOTEMPORAL MODELING OF AGRICULTURAL YIELD MONITOR DATA

by
Adela Nistor[*,1], adelanistor@alumni.purdue.edu,
Raymond J.G.M. Florax[1,2], rflorax@purdue.edu
Jess Lowenberg-DeBoer[1], lowenbej@purdue.edu
Jason P. Brown[1],brown151@purdue.edu

[1] Purdue University, Dept. of Agricultural Economics,
403 W. State St., West Lafayette, IN 47907-2056, USA

[2] VU University, Dept. of Spatial Economics,
De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

Working Paper #08-01
January 2008

Abstract

This paper shows that spatial panel data models can be successfully applied to an econometric analysis of farm-scale precision agriculture data. The application focuses on the estimation of the effect of controlled drainage water management equipment on corn yields. Using field-level precision agriculture data and spatial panel techniques, the yield response equation is estimated using the spatial autoregressive error random effects model with temporal heterogeneity, incorporating spatial dependence in the error term, while controlling for the topography, weather and the controlled drainage treatment. Controlling for random effects allows for the disentanglement of the effects of spatial dependence from spatial heterogeneity and omitted variables, and thus, to properly investigate the yield response. The results show that controlled drainage has a statistically significant effect on corn yields. The effect is generally positive but varies widely from year to year and field-to-field. For the two years of data controlled drainage was linked to a 2.2% increase in field average yield, but that varied from a -2.6% to a +6.5%. Evaluated at mean elevation and slope in the east part of the field, controlled drainage is associated with 10 bu/a increase and a 0.6 bu/a decrease in yields in 2005 and 2006, respectively. In the West part of the field, controlled drainage is associated with a 11 bu/a increase in 2006 and 2.81 bu/a decrease in 2005.

Keywords: corn, drainage, precision agriculture, spatial panel model

JEL classification: O18, Q18, R15, R38, R58

1. Introduction

This paper applies econometric spatial panel models developed by Anselin (1988), Elhorst (2003) and others to agricultural yield monitor data. Specifically, we investigate an experiment using controlled drainage technology and assess its impact on corn yields at the farm level in Indiana. We analyze yield monitor data over time and space by using Geographical Information Systems (GIS) and spatial panel econometric methods, in particular the spatial fixed and random effects models with spatial error autocorrelation. The use of panel data methods controlling for spatial and temporal heterogeneity and dependence as well as potential omitted variable bias provides precision agriculture researchers with a powerful framework to model crop sensor data over space and time. A specification that conforms to the agronomic requirements of yield response is the spatial autoregressive error random effects model with spatial and temporal heterogeneity. The development and use of spatio-temporal models in precision agriculture research enhances the array of spatial cross-sectional evaluation tools available to measure the impact of alternative management practices on crop yields, and aids to a better understanding of the complex agronomic phenomena underlying yield response.

In terms of application we focus on the impact of using drainage water management on corn yields. Apart from the potentially beneficial effect of drainage water management practice on yields, the use of the controlled drainage technology is also motivated by environmental concerns. Excess nutrients from anthropogenic sources increase algal production, causing eutrophication of coastal ecosystems. For instance, in the Midwest of the United States too much nitrate (N) load in surface waters from drained agricultural land creates negative environmental impacts in the Gulf of Mexico (Burkhart and James 1999; Gilliam et al. 1999; Rablais et al.

2002). In the future, farmers may therefore be required to adopt technologies that have been demonstrated to reduce N loads to surface water, such as controlled drainage, also referred to as drainage water management. Controlled drainage restricts outflow during periods of the year when equipment operations are not required in the field (i.e., winter and midsummer). This may increase water available to crops in midsummer and thereby increase yields (Evans and Skaggs 1996). Drainage trials in small plots are difficult, as they require major investment in barriers to prevent water movement between plots, thus creating an unnatural situation that may not be representative of field conditions. For drainage trials, landscape experimental designs works well and the most cost effective way to collect yields from landscape designs is with yield monitors. The drainage water cases studied in this paper are motivated by the recognition that voluntary adoption of drainage water management by growers depends on the size of the yield increase (Evans and Skaggs 1996). In addition, existing incentive programs such as the Environmental Quality Incentives Program (EQIP) require quantitative information on practice efficacy and on private benefits.

2. Literature review

Recent spatial panel data applications in economics include the analysis of household level survey data from villages observed over time to study nutrition (Case 1991), per capita expenditures on police to study their effect on reducing crime across counties (Kelejian and Robinson 1992), the productivity of public capital like roads and highways in the private sector across U.S. states (Holtz-Eakin 1994), hedonic pricing equations using residential sales (Bell and Bockstael 2000), unemployment clustering with respect to different social and economic metrics

(Conley and Topa 2002), spatial price competition in wholesale gasoline markets (Pinkse et al. 2002) and regional growth modeling in Italy (Arbia and Piras 2004).

There have been only a small number of studies that employed spatio-temporal regression analysis in the study of yield monitor data (Bongiovanni and Lowenberg-DeBoer 2002; Lambert et al. 2006; Liu et al. 2006; Nistor 2007). Prediction in spatio-temporal domains has drawn significant attention in the data analysis community (Pace 1988) and can contribute to a better understanding of complex phenomena studied in precision agriculture. Bullock and Lowenberg-DeBoer (2007) provide a recent review of studies using spatial econometric analysis techniques applied to precision agriculture data.

There exist only a limited number of studies of the effect of drainage management on average crop yields, and none of those addresses conditions in the Midwest of the U.S. Sipp et al. (1986), Cooper et al. (1991, 1992), Drury et al. (1997) and Fisher et al. (1999) documented yield increases with subirrigation, while Tan et al. (1988) measured yield changes with managed drainage as opposed to conventional drainage. Trials by Tan et al. (1998) in Southwestern Ontario showed a slight soybean yield benefit for managed drainage under conventional tillage and a small yield decline with no-till, but neither of these yield differences was statistically significant at conventional levels. Nine out of 15 farmers involved in a central Illinois drainage management project said that they had higher yields with drainage management (Pitts 2003). All the above studies estimating the effect of controlled drainage on yields use small plot or whole-field data with the harvest from the combine transferred to a weigh wagon, and subsequent analysis based on comparing treatment trials or performing an analysis of variance. In both cases, however, spatial econometric or spatial statistical techniques have not been used. Effectively, it is a priori assumed that the distribution of yields across the field is homogenous and independent of location. Brown (2006) applies spatial econometric techniques to cross-section yield monitor data in 2005 for four farms located in White, Montgomery and Randolph County in Indiana in order to study the economic feasibility of controlled drainage in the Cornbelt. Using spatial error regression models for the estimation of yields as a function of linear, quadratic and interaction terms including elevation, slope, distance to the nearest tile line and infrared soil color, Brown (2006) found that controlled drainage impacts yield in the range of 8 bu/acre to 29 bu/acre. Nistor (2007) proposes a framework to model crop sensor data over time by using the spatial fixed and random effects models, with an application focused on estimating the controlled drainage impact on farm profitability in the Cornbelt. Nistor (2007) found the decision to invest in controlled drainage technology to be supported for three of the four experimental farms, both with and without subsidy.

3. Methods and data

3.1 Data and specification

The empirical example in this paper is concerned with yield monitor data sampled from the farm located at Davis Purdue Agricultural Center (DPAC), field W, located in Randolph County, Indiana. The yield data were collected with an AgLeader yield monitor linked to a global positioning system (GPS). The yield monitor is located on the combine and records crop yields on the go. Yield files include data-point information about yields (bu/a), latitude, longitude and grain moisture, which is used to generate a geopositioned database and site-specific yield maps. The yield measurement samples collected have been taken from the field surface with the locations considered as points or very small areas (see Griffin et al. 2005a, for a more elaborate discussion). The design of the controlled and conventional drainage experiments are created via digitization using the tile line maps.

Because the spatial layout of the raw data is such that it included points located closer together within the row than between the rows, the dataset was constructed as follows: yield monitor data were aggregated into average combine pass width squares in order to provide data that are spatially balanced in all directions. Previous applications of this methodology can be found in Malzer et al. (1996), Mamo et al. (2003) and Anselin et al. (2004). The square grid with cells thus created was overlaid on the yield points and the grids were rotated by the corresponding field angle. Each cell value, expressed in bushels per acre, represented the average of all points contained within that square so that a yield map was created with a finite number of color scales easily identifiable to the viewer from many thousands of individual yield point values. This process was performed using the same grid each year, so that the grids are coincident, which permits the comparison of yields for different years in the "same" location.

The balanced design thus obtained allows for a spatial econometric approach using a weighting design (Anselin et al. 2004). Moreover, since the prediction error for the average values of yields within grids is smaller than the prediction error for any yield point prediction, the precision of the average yield estimator is higher than that of point estimator (Haining 2003), although this procedure also introduces heteroskedasticity to a certain extent.[1] Elevation point data with reference to the sea level, collected by topographic surveys performed by contractors for the farm, were interpolated using the Inverse Distance Weighted (IDW) power 1 method,[2] so that a point data set was obtained with elevation across the whole field. Each cell value was assigned the average of the elevation points that completely fell inside each cell and was converted with reference to the lowest elevation level in the field. This implies that the elevation in each grid cell equals the difference between the average elevation with respect to the sea level and the minimum average elevation. Slope data expressed in percents were derived from the elevation data in the same manner, using the Toolbox in ArcGIS9.

The measures for the average combine pass width were 4.95m (1996), 5.08m (1998), 4.86m (2000), 4.93m (2001), 4.95m (2002), 5.03m (2003), 4.94m (2005) and 5.02m (2006), and the grid size was therefore rounded to 5 meters. The dataset was constructed as follows: yield point data were aggregated into squares of 5×5 m that were overlaid on the yield points and the grids were rotated by the corresponding field angle (357.7^0). The controlled and conventional drainage parts of the field were the northwest, southeast and the northeast, southwest parts of the fields, respectively (see Figure 1). Field W was cultivated under a corn-soybeans rotation for many years, but years when corn was planted was not the same for the East and West Sides of the field, hence the two sides of the field, East and West were analyzed separately. Controlled drainage was performed in 2005 and 2006 only; the years with corn rotation were 1996, 1998, 2000, 2002, 2005, 2006, and 1996, 1998, 2001, 2003, 2005, 2006 for the East and West part of the field respectively.

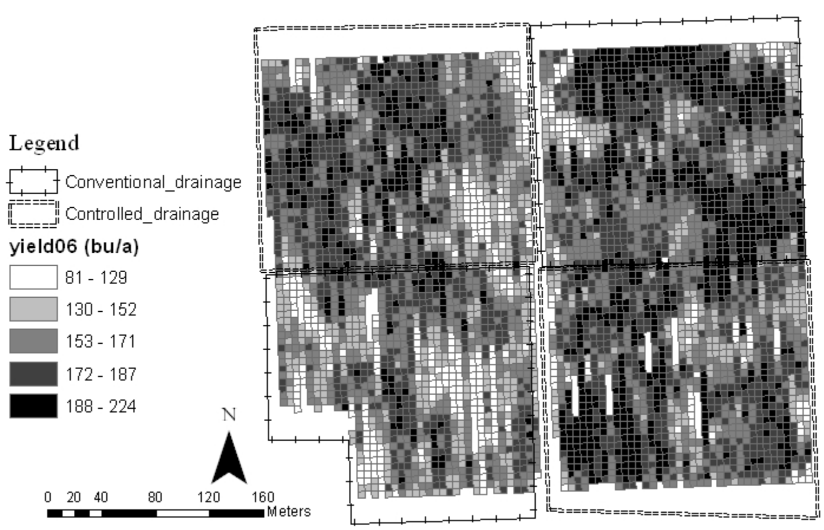

Figure 1. Yield map (Davis, Field W; 2006, corn)

Rainfall data over the growing season, taken as July to September, were obtained from the weather station located at 0.5-mile distance from Field W. The choice of the growing season period was determined by professional judgment of soil scientists and agricultural engineers involved in the project. Although this is unusual, in some years (2005, 2006) corn did not reach physiological maturity (i.e., the R6 growth stage when black layer forms at the tip of the kernels) before the end of September due to late planting (end of May, early June). This motivates the inclusion of the September rain data.

Heady and Dillon (1972) provide a review of algebraic functional forms for crop response estimation. The selection of variables and specification of the crop yield functional form are difficult because of lack of theoretical guidance in the agronomy and soil science literature, and the complexity of yield response (Swanson 1962; Florax et al. 2002; Anselin et al. 2004). Nistor (2007) provides an elaborate overview of different functional forms that have been used in agronomy and soil science. For this application a simple linear form with interaction variables is chosen, because of the limited availability of data. For on-farm yield trials slope, elevation and rainfall are the most commonly available variables. Data that varies in time and space (e.g. annual soil tests, remotely sensed biomass) is sometimes available on research farms, but rarely for commercial fields like those used for the drainage trials.

Since the yield monitor data is a sample rather than a population (Griffin et al. 2005), the random effects (RE) model is appropriate for the analysis of precision agriculture data. Nistor (2007) provides a discussion of the proper framework for precision agriculture data over time. For precision agriculture data the spatial error model is more appropriate than the spatial lag model, because spatial autocorrelation is due to omitted variables rather than to the effect of corn yield grid cells on each other (Anselin et al. 2004; Lowenberg-DeBoer et al. 2006). In addition, temporal heterogeneity is much more important than spatial heterogeneity and should also be taken into account, since the yield response and the controlled drainage impact vary across the years (Bongiovanni and Lowenberg-DeBoer 2002; Nistor and Lowenberg-DeBoer 2007). Therefore, the random effects spatial error model extended to account for temporal heterogeneity (SEM-RE model) was chosen for estimation. The lack of routines for the two-way random effects model extended to account for spatial error autocorrelation, led us to consider temporal heterogeneity in the SEM-RE model in the form of time dummy variables.

The drainage dummy was interacted with time dummies in the experimental years to account for the variability in the yield response to controlled drainage over years. The interaction terms between the drainage dummy, elevation and slope were included since impact of controlled drainage vary with topography and controlled drainage does not affect yields the same across the field.

The crop yield response to controlled drainage is different across years, with no yield benefit in years with insufficient rain, or a negative impact with very low field topography that would allow high enough water to have a detrimental effect (Nistor and Lowenberg–DeBoer 2007). Because of the relationship between topographic attributes, soil properties and available water, the precipitation in the growing season is interacted with the topographic attributes that may influence crop yields (Kaspar et al. 2003). With the inclusion of these interaction variables, the specification estimated reads as:

$$\begin{aligned} Y = {} & \beta_1 + \beta_2 Elevation + \beta_3 Slope + \beta_4 D \times Elevation + \beta_5 D \times Slope + \beta_6 Rain \times Elevation + \\ & \beta_7 Rain \times Slope + \beta_8 T_1 + \beta_9 T_2 + \beta_{10} T_3 + \beta_{11} T_4 + \beta_{12} T_5 + \beta_{13} D \times T_1 + \beta_{15} D \times T_2 \end{aligned} \qquad (1)$$

where D is the drainage dummy, Y refers to yields and $T_1,...,T_5$ dummy variables for the time periods.

The SEM-RE model specification in equation (1) with temporal heterogeneity offers a more comprehensive approach of yield response estimation that conforms to the requirements of yield response in the agronomy literature, while accounting for both spatial and temporal heterogeneity, and therefore offers a framework with most reliable results.

3.2 Spatial panel models

The traditional panel data models used in applied research are the fixed effects (FE) and the random effects (RE) model (Baltagi 2001). A panel data set consist of a sequence of observations repeated through time, on a set of units (e.g., individuals, firms, or countries). A panel data regression is different from a time-series or cross-section regression in that it considers both the temporal and the cross-sectional dimension. Panel data offer researchers extended modeling possibilities as compared to purely cross-sectional data or time-series data, because they contain more information, more variability, less collinearity among the variables, more degrees of freedom, and hence the estimators are likely to be more efficient. Panel data can reduce the effects of omitted variables bias by controlling for individual heterogeneity. Panel data also allow for the specification of more complicated behavioral hypotheses, including effects that cannot be addressed using pure cross-sectional or time-series data. For example, technical efficiency is better studied and modeled with panel data sets, because in cross-sectional models it cannot be identified, and in time series models it is assumed to be identical across cross-sectional units (Hsiao 1986; Baltagi 2001). An important advantage of panel data compared to time series or cross-sectional data sets is that it is better able to identify and measure effects that are simply not detectable in pure cross-section or pure time-series data (Ben-Porath 1973). Panel data can reduce the effects of omitted variables bias by controlling for individual heterogeneity. Time-series and cross-section studies not controlling for this heterogeneity run the risk of obtaining biased results (Moulton 1986, 1987).

Contemporaneous spatial dependence between observations at each point in time and spatial heterogeneity (i.e., parameter heterogeneity that varies with the spatial location) may arise when panel data include a location component (Anselin 1988; Elhorst 2003). Spatial dependence may be incorporated into the model as spatial error autocorrelation or as a spatially lagged dependent variable, or a combination of both (Anselin and Hudak 1992). These different specifications of spatial dependence have different implications for estimation and statistical inference. Estimating a model ignoring spatial error autocorrelation by means of Ordinary Least Squares (OLS) produces unbiased and consistent parameter estimates, but the OLS estimator loses the efficiency property. Erroneously omitting a spatially autocorrelated dependent variable from the explanatory variables causes the OLS estimator to be biased and inconsistent, except under special circumstances (Anselin 1988).

Anselin et al. (2006) provide an overview of specifications and estimators available for spatial panel data. The traditional spatial random effects model described in Anselin (1988) has recently been extended. Kapoor et al. (2007) allow for the same spatial error autocorrelation in both the individual effects and the remainder errors. Baltagi et al. (2006) extend the theoretical econometric specification of Kapoor et al. (2007) to assume different spatial error processes in the spatial and remainder error components, and test for their restricted counterparts. Regarding the software resources for estimating the panel spatial econometrics, the situation is still rather bleak (Anselin et al. 2006). For the family of dynamic spatial panel models, no straightforward estimation procedure is yet available (Elhorst 2001, 2005). The fact that the estimation of spatial panel data models is not very well documented in the literature may be due to each model having its own specific problems. This study applies the estimation framework as developed by Elhorst (2003), specifically the random effects spatial error models that incorporate spatial error autocorrelation in the context of maximum likelihood estimation procedures.

Following Elhorst (2003), if we stack the observations in one equation for each set of cross-sections over time (i.e., T spatial series with N observations over space), the traditional RE model extended to spatial error autocorrelation, SEM-RE for short, can be specified as:

$$Y_t = X_t\beta + v\,,\quad v = (\iota_T \otimes I_N)\alpha + [I_T \otimes (I_N - \delta W)^{-1}]\varepsilon, \tag{2}$$

where $i = 1,2,...,N$ refers to a spatial unit, $t = 1,2,...,T$ to a given time period, $Y_t = (Y_{1t},...,Y_{Nt})'$, $X_t = (X_{1t},...,X_{Nt})'$, $\varphi_t = (\varphi_{1t},...,\varphi_{Nt})'$, $\alpha = (\alpha_1,...,\alpha_N)'$, $\varepsilon_t = (\varepsilon_{1t},...,\varepsilon_{Nt})'$, and α is the variable intercept treated as random representing the effect of omitted variables that are specific to each spatial unit considered. The random effects model treats α_i as a random variable assumed to be $IIN \sim (0, \sigma_\alpha^2)$, and we have $E(\alpha_i, \alpha_j') = \sigma_\alpha^2$ if $i = j$ and zero otherwise. It is assumed that the random variables α_i and ε_{it} are independent of each other.

The weights matrix W is an $N \times N$ matrix describing the spatial arrangement of the spatial units, where w_{ij} is the (*i*,*j*)-th element of W with $w_{ij} = 1$ if *i* and *j* are neighbors, and $w_{ij} = 0$ otherwise. In equations (2) and (3), δ is called the spatial autoregressive coefficient. Estimation is by maximum likelihood (Elhorst 2003). Kapoor et al. (2007) provide an approach based on general moments estimation.

4. Data and results

4.1 Exploratory spatial data analysis

We can see from Table 1 that the mean of the corn yields is fairly stable over time, except for 1996 (weed problems) and 2002 (severe draught). There is an unstable pattern of the yields variance corresponding to the controlled drainage zones over the years. In the southeast part of the field, the variance in 2006 was statistically significantly lower than in 2000, but higher than the rest of the years; the variance in 2005 was statistically significantly lower than in 2000, 2002, and 2006, but higher in 1996 and 1998. In the northwest part of the field, the variance in 2005 was highest than in all the other years, with equality in 2003; the variance in 2006 was statistically significantly lower than in 2003 and 2005 only, but higher than in 1996, 1998 and 2001. For the west side of the field in 2006, the mean yields with controlled drainage were higher than the mean yields with free flowing drainage, but not for 2005 when controlled drainage yields were lower. For the east side of the field in 2005, the mean yields with controlled drainage were higher than the mean yields with free flowing drainage, but not for 2006, when controlled drainage yields were lower. The comparison based on average yield may be misleading because it does not take into account differences in topography, soils, microclimate and other factors between controlled drainage areas and those with free flowing drainage.

Table 1. Corn Yield (bu ac^{-1}) and Precipitation Descriptive Statistics, Davis, Field W

EAST (Controlled)	1996	1998	2000	2002	2005	2006	WEST (Controlled)	1996	1998	2001	2003	2005	2006
Minimum	60	97	102	10	104	102	Minimum	31	87	106	54	80	81
Maximum	131	184	240	105	239	212	Maximum	127	197	227	180	210	224
Mean	98	144	186	47	178	170	Mean	82	149	176	130	150	167
SD	13	15	25	19	17	20	SD	18	20	20	23	23	22
EAST (Uncontrolled)							WEST (Uncontrolled)						
Minimum	35	86	88	10	79	115	Minimum	33	83	101	51	91	81
Maximum	131	199	262	102	227	222	Maximum	121	199	232	183	208	209
Mean	98	147	189	50	160	177	Mean	89	135	175	117	156	154
SD	12	17	30	19	30	19	SD	14	22	19	27	19	25
EAST (Whole Field)							WEST (Whole Field)						
Minimum	35	86	88	10	79	102	Minimum	31	84	101	51	81	81
Maximum	131	199	263	105	239	223	Maximum	127	199	232	183	210	224
Mean	98	145	187	49	169	173	Mean	85	143	176	124	153	161
SD	13	16	28	19	25	20	SD	17	22	20	25	21	24
Rain (in)	3.6	4.03	4.89	2.53	5.67	3.78	Rain (in)	3.6	4.03	4.96	7.33	5.67	3.78

Descriptive statistics values for the topography of the field (see Table 2) show that for the east side of the field, mean elevation was higher in the controlled than in the uncontrolled part, while for the west part of the field, mean elevation was lower in the controlled than in the uncontrolled part.

Table 2.Elevation and slope descriptive statistics (Davis, Field W)

	Whole field		Controlled		Uncontrolled	
EAST	Slope (%)	Elevation (m)	Slope (%)	Elevation (m)	Slope (%)	Elevation (m)
Minimum	0.05	0.00	0.08	0.50	0.05	0.00
Maximum	1.87	1.84	1.87	1.84	1.44	1.29
Mean	0.57	0.91	0.57	1.08	0.58	0.74
SD	0.26	0.35	0.23	0.29	0.29	0.32
WEST						
Minimum	0.04	0.00	0.04	0.00	0.06	0.37
Maximum	2.36	2.30	1.53	1.26	2.36	2.30
Mean	0.60	0.87	0.56	0.59	0.64	1.20
SD	0.30	0.47	0.27	0.27	0.33	0.46

Figures 2a and 2b show an obvious clustering of similar attribute values: relatively high yields, very low yields and relatively low yields.

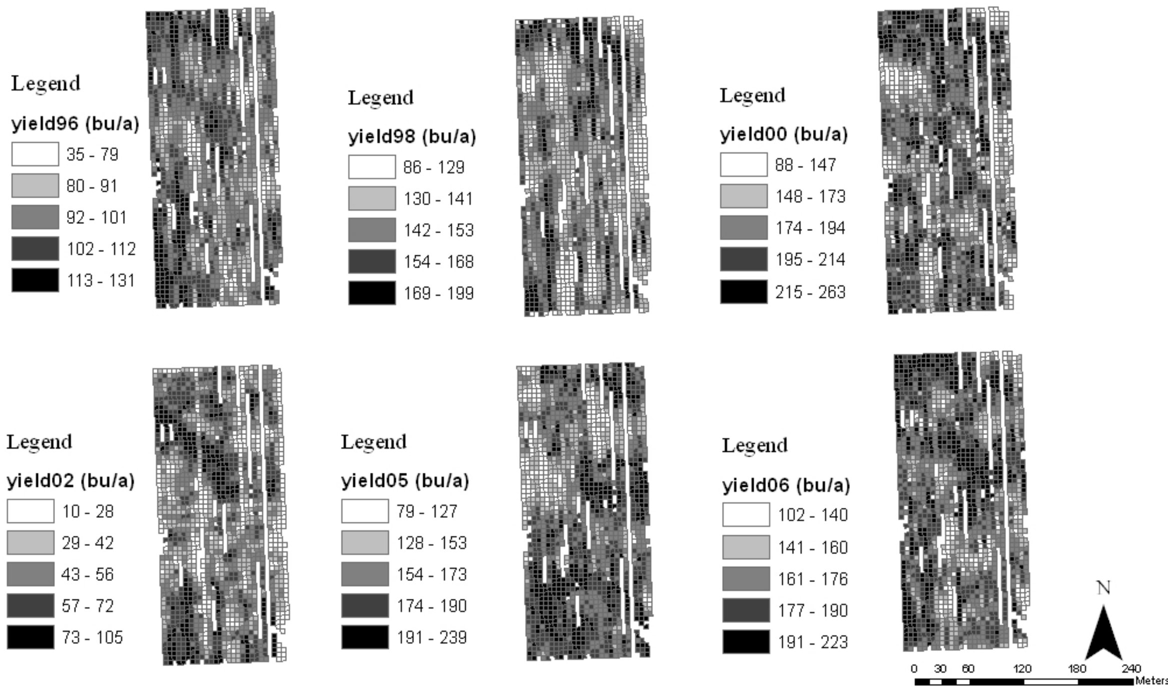

Figure 2a. Yield Map (Davis, Field W, East; 1996, 1998, 2000, 2002, 2005, 2006, Corn)

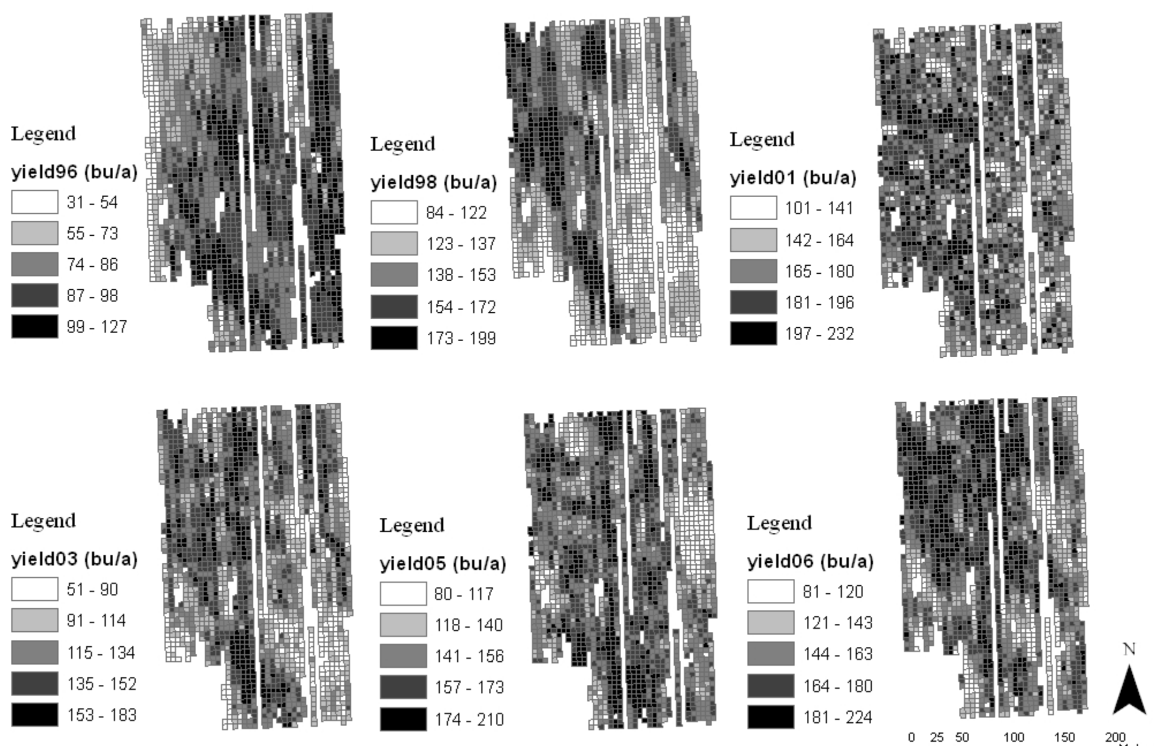

Figure 2b. Yield Map (Davis, Field W, West; 1996, 1998, 2001, 2003, 2005, 2006, Corn)

To evaluate the significance of the spatial clustering pattern by means of the Moran's *I* statistic, the spatial weights matrix was defined according to the queen criterion, implying that grid cells are neighbors if they have a common border in the horizontal or vertical dimension, or if they share a common vertex, up to the one "band" of neighbors. The feasibility of the regression models required a compromise in choosing the first order queen weights matrix, since spatial panel models cannot be estimated using a weights matrix with many neighbors. The spatial panel models estimated consider only contemporaneous spatial dependence, and hence the combined weights matrix for all years is block-diagonal, with W for each year as a submatrix on the diagonal. When the weights matrix is row standardized, the spatially lagged yield variable is the average of the yields in the neighboring grid cells. We can see from Table 3 that the sign of Moran's *I* statistic for yields is positive and highly significant so that high (low) values are surrounded by high (low) values in neighboring grids, indicating positive spatial correlation of yields.

Table 3. Moran's *I* (yields), Davis, Field W

	1996	1998	2000	2002	2005	2006
EAST	0.52***	0.53***	0.57***	0.63***	0.70***	0.56***
	1996	**1998**	**2001**	**2003**	**2005**	**2006**
WEST	0.68***	0.74***	0.39***	0.61***	0.64***	0.62***

*** denotes significance at 1% level (permutation assumption).

4.2 Regression results

4.2.1 Davis, Field W, East

Table 4 presents the results of the a-spatial random effects (RE model, column a) and spatial error random effects model (SEM-RE model, column b). Table 4 shows that the controlled drainage impact varies across the years and with topography.

Table 4. Pooled estimates of corn yields, Davis, Field W, East
Dependent variable: yields, T=6 and N=1592*

Dependent variable: yields	RE (a)	SEM-RE (b)
Constant	108.036*** (1.270)	108.338*** (1.499)
Elevation	-19.449*** (2.493)	-18.704*** (4.699)
Slope	17.829*** (3.212)	1.255 (3.803)
Drainage × Elevation	-4.001** (1.951)	-2.111 (3.246)
Drainage × Slope	8.906*** (2.406)	1.090 (2.552)
Rain × Elevation	2.827*** (0.595)	2.850** (1.168)
Rain × Slope	-5.780*** (0.745)	-1.519* (0.904)
2006	77.576*** (0.833)	78.262*** (1.771)
2005	63.099*** (1.505)	61.693*** (2.767)
2002	-50.114*** (0.955)	-49.768*** (1.931)
2000	90.289*** (1.091)	84.638*** (2.110)
1998	47.713*** (0.722)	47.614*** (1.548)
Drainage × 2006	-4.360 (2.344)	0.654 (3.939)
Drainage × 2005	17.850*** (2.274)	11.377** (3.826)
Spatial autocorrelation	-	0.806*** (0.009)
R-squared	0.879	0.962
LIK	-41899	-41899

* Standard errors are in parentheses

The controlled drainage impact is $\Delta Y / \Delta D = -2.11 Elevation + 1.09 Slope + 0.65$ and $\Delta Y / \Delta D = -2.11 Elevation + 1.09 Slope + 11.37$ for the 2006 and 2005 years, respectively. Table 5 shows the controlled drainage impact on yields and the associated confidence intervals (C.I.). Nistor (2007) provides a detailed explanation on the C.I. computation used. Evaluated at mean topological values in the field, the impact of controlled drainage on yields is small in 2006 (-0.6

bu/a) but substantial in 2005 (10 bu/a). In both years controlled drainage has a significant impact on yields (at the 1% level), with a corresponding Likelihood Ratio (LR) test of 28 and 340 for 2006 and 2005 respectively, under the $\chi^2(3)$ distribution.

Table 5. Controlled drainage impact (bu ac^{-1}) on yields, Davis, Field W*

	EAST		WEST	
	2006	2005	2006	2005
Mean elevation, slope	-0.64 (-4.41; 3.11)	10.08 (6.23; 13.91)	11.27 (7.52; 15.05)	-2.81 (-6.52; 0.96)
Minimum elevation, slope	0.71 (-6.96; 8.38)	11.43 (3.99; 18.87)	8.93 (3.66; 14.18)	-5.14 (-10.34; 0.05)
Maximum elevation, slope	-1.19 (-9.67; 7.27)	9.53 (0.77; 18.27)	11.16 (-0.87; 23.23)	-2.91 (-14.96; 9.19)
Controlled drainage area	-1.02 (-4.67; 2.65)	9.70 (5.87; 13.55)	9.87 (6.29; 13.42)	-4.20 (-7.72; -0.70)

* Confidence intervals are in parentheses

Table 5 shows that the overall estimate of the yield effect for the controlled drainage area is negative and negligible in 2006 (-1.02 bu/a, -0.6% of the average whole field yield) and positive in 2005 (+9.70 bu/a, +5.6% of the average whole field yield). The greatest impact of controlled drainage on yields is in the year 2005 with a negligible negative impact in 2006 (see Figure 3a). The overall yield estimate for the controlled treatment area was calculated by summing over per cell yield effects in that part of the field.

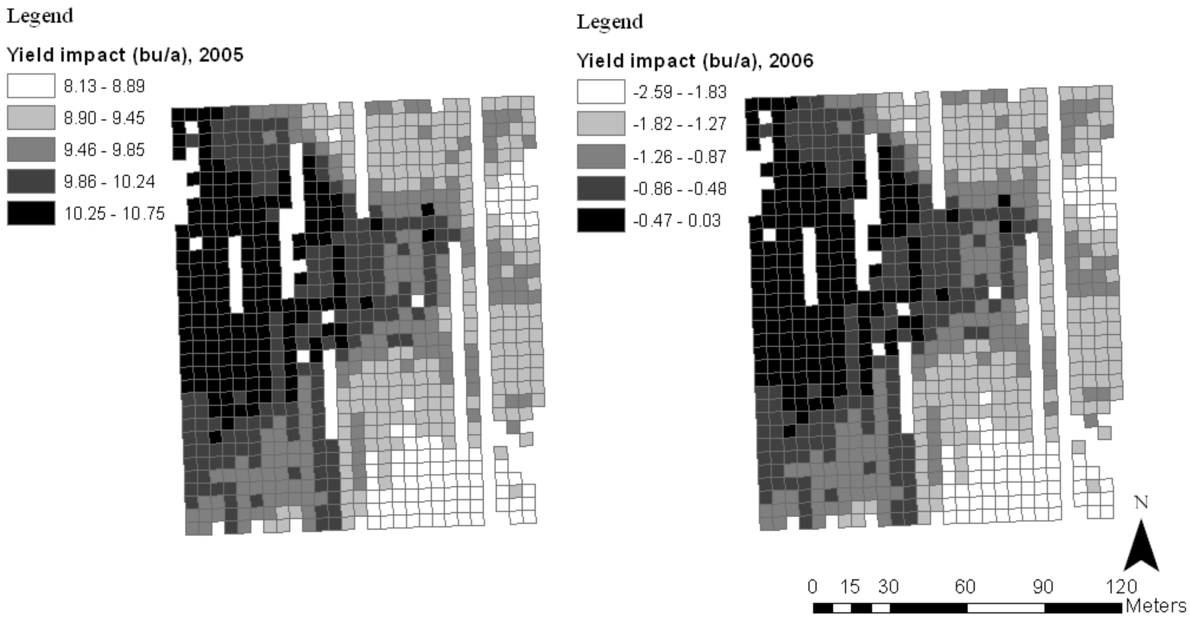

Figure 3a. Controlled Drainage (SEM-RE Model) Impact (bu ac^{-1}) on Yields (Davis, Field W, North East Quadrant – Controlled Drainage Treatment Area)

4.2.2 Davis, Field W, West

Table 6 presents the results of the a-spatial random effects (RE model, column a) and spatial error random effects model (SEM-RE model, column b). Table 6 shows that the controlled drainage impact on yields is $\Delta Y / \Delta D = 5.70 Elevation - 4.69 Slope + 9.12$ and $\Delta Y / \Delta D = 5.70 Elevation - 4.69 Slope - 4.95$ for the 2006 and 2005 years, respectively. Evaluated at mean topological values in the field, controlled drainage is negatively associated with corn yields in 2005 (-2.81 bu/a) and positively in 2006 (11 bu/a). Controlled drainage has a significant impact on yields (at the 1% level) in both 2005 and 2006 years, with a corresponding LR test of 154 and 167 for 2006 and 2005 respectively, under the $\chi^2(3)$ distribution.

Table 6. Pooled estimates of corn yields, Davis, Field W, West
Dependent variable: yields. T=6 and N=1953*

Dependent variable: yields	RE	SEM-RE
Constant	81.977*** (1.004)	85.579*** (1.394)
Elevation	10.033*** (1.685)	7.944*** (2.819)
Slope	1.911 (2.582)	-1.702 (3.241)
Drainage × Elevation	-4.636*** (1.762)	5.704** (2.837)
Drainage × Slope	-6.654*** (1.793)	-4.694** (2.382)
Rain × Elevation	-2.628*** (0.310)	-2.151*** (0.548)
Rain × Slope	0.594 (0.470)	1.162* (0.618)
2006	70.386*** (0.794)	67.411*** (1.644)
2005	77.841*** (1.070)	74.055*** (2.033)
2003	46.650*** (1.354)	38.276*** (2.446)
2001	93.513*** (0.756)	89.727*** (1.536)
1998	58.749*** (0.627)	55.577*** (1.367)
Drainage × 2006	17.515*** (1.642)	9.124*** (2.719)
Drainage × 2005	-4.606*** (1.625)	-4.954* (2.692)
Spatial autocorrelation	-	0.822*** (0.007)
R-squared	0.71	0.95
LIK	-52051	-52051

*Standard errors are in parentheses

Table 5 shows that the overall estimate of the yield effect for the controlled drainage area is negative in 2005 (-4.20 bu/a, -2.6% of the average whole field yield) and positive in 2006 (+9.87 bu/a, +6.5% of the average whole field yield). Figure 3b visualized the impact of controlled drainage on yields which is positive throughout the field in 2006 but not in 2005 when it is negative.

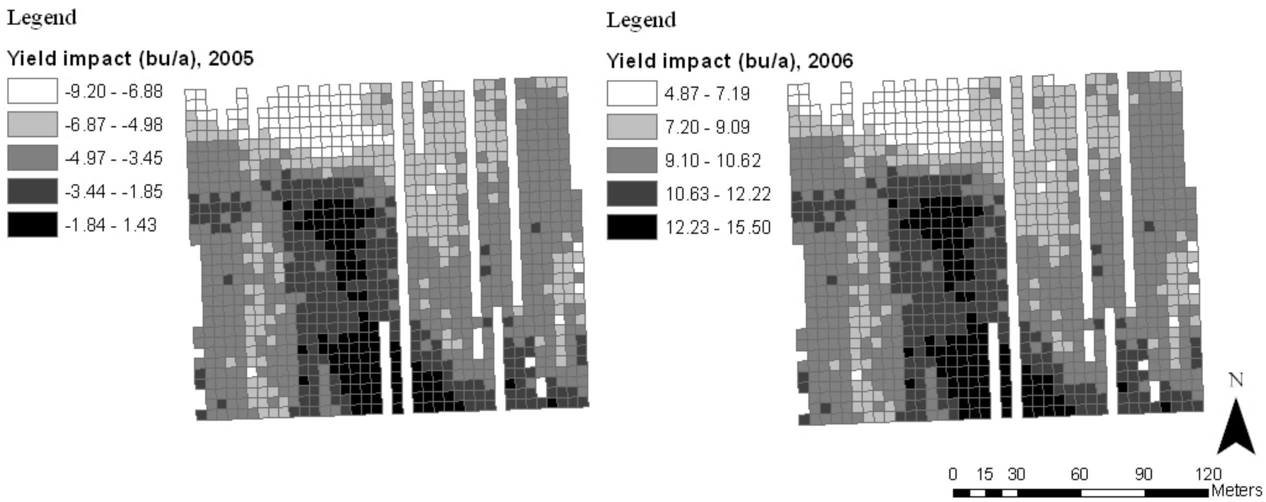

Figure 3b. Controlled Drainage (SEM-RE Model) Impact (bu ac^{-1}) on Yields
(Davis, Field W, South West Quadrant – Controlled Drainage Treatment Area)

5. Conclusions

This study shows that spatial panel data models can be applied to an econometric analysis of farm-scale precision agriculture information in data rich environments with independent variables that vary over time and space. The application deals with the assessment of the impact of controlled drainage technology on corn yields for two sides of one field in Indiana. Using field-level yield monitor data, the yield response equation is estimated using spatial panel econometric models, namely the spatial autoregressive error random effects model with both spatial and temporal heterogeneity incorporating spatial dependence in the error term, while controlling for the topography, weather and the controlled drainage treatment. The use of random effects allows for the disentanglement of the effects of spatial dependence from spatial heterogeneity and omitted variables, and thus, is necessary to properly investigate the yield response. The results show that the relationship between controlled drainage and corn yields is quite variable across years and fields. The effect is generally positive, but varies widely from year to year and field-to-field. Evaluated at mean elevation and slope in the field, controlled drainage is associated with 10 bu/a increase and a 0.6 bu/a decrease in yields in 2005 and 2006 respectively for the East part of the field. In the west part of the field, controlled drainage is associated with a 11 bu/a increase in 2006 and 2.81 bu/a decrease in 2005. The overall estimates of the yield effect for the controlled drainage area show that controlled drainage is associated with a decrease in yields in 2005 (-4.20 bu/a , -2.6%, West) and 2006 (-1.02 bu/a, -0.6% East) and an increase in yields in both 2005 (9.70 bu/a, +5.6%, East) and 2006 (9.87 bu/a, +6.5%, West). The overall yield impact over the two years and two fields averaged 2.2% of average whole field yield.

This paper shows both results regarding controlled drainage impact on corn yields and a method of how to analyze precision agriculture data over time, by using GIS and spatial panel methods. Precision agriculture researchers can use the applied frameworks for modeling crop sensor data over time, to better evaluate the effect of various management practices and better understand the complex crop growth phenomena studied in precision agriculture. Regarding the implications for drainage management, the results have to be interpreted cautiously, due to drainage management issues. The experimental field was not under controlled drainage over the winter period, as environmental best practices would require (Frankenberger et al. 2007). More data is needed for more precise results. Inferences cannot be generalized to all the fields in the Midwest or beyond, since the analysis focuses on within field variations. Future research incorporating spatial correlation in the random effects may be a useful extension of the approach adopted here.

Notes

[1] The procedure of averaging the yields in the grids induce heteroskedasticity because the variance will generally depend on the number of points per cell. This is difficult to incorporate in the regression models, because some grid cells only contain one observation

[2] The inverse weighted distance (IDW) method assignes values to unknown points by using values from known points. For *p*, any positive real number called the power parameter, the value of the interpolated point is $\sum_{i=1}^{N} \frac{Z_i}{d_i^p} \Big/ \sum_{i=1}^{N} \frac{1}{d_i^p}$ where Z_i is a known value at each point i, N the total number of known points used in interpolation, and d the distance from the known value to the unknown value.

References

Anselin, L. (1988). “Spatial econometrics: methods and models”. Dordrecht: Kluwer Academic Press

Anselin, L., Bongiovanni, R., Lowenberg DeBoer, J. (2004). A spatial econometric approach to the economics of site-specific nitrogen management in corn production. American Jornal of Agricultural Economics 86: 675–687

Anselin, L., Hudak, S. (1992). Spatial econometrics in practice: a review of software options. Regional science and urban economics 22: 509–536

Ariba G., Piras, G. (2004). Convergence in per-capita GDP across European regions using panel data models extended to spatial autocorrelation effects. Paper presented at the European Regional Science Association, conference Porto

Baltagi, B. (2001). Econometric analysis of panel data. J. Wiley & Sons Ltd. England

Baltagi, B., Egger., Pfaffermayr, M. (2006). A generalized spatial panel model with random effects. ADRES International Conference 2006, September 14–15. Saint-Etienne, France. “Networks of Innovation and Spatial Analysis of Knowledge Diffusion”

Bell, K., Bockstael, N. (2000). Applying the generalized-moments estimation approach to spatial problems involving microlevel data. Review of Economics and Statistics 82: 72–82

Ben-Porath, Y. (1973). Labor force participation rates and the supply of labor. Journal of Political Economy 81: 697–704

Bongiovanni, R., Lowenberg-Deboer, J. (2002). Economics of nitrogen response variability over space and time: results from the 1999–2001 field trials in Argentina. Proceedings of the 6th International Precision Agriculture Conference, Minneapolis, MN, 2002

Brown, J. P. (2006). Methodology for determining the economic feasibility of controlled drainage in the eastern corn belt. Unpublished manuscript, MS Thesis, Purdue University

Bullock, D., Lowenberg-DeBoer, J. (2007). Using spatial analysis to study the values of variable rate technology and information. Journal of Agricultural Economics 58: 517–535

Burkhart, M., James, D. (1999). Agricultural-nitrogen contributions to hypoxia in the. Gulf of Mexico. J. Environ Qual. 28: 850–859

Case, A. (1991). Spatial patterns in household demand. Econometrica 59: 953–965

Conley, T., Topa, G. (2002). Socio-economic distance and spatial patterns in unemployment. Journal of Applied Econometrics 17: 303–327

Cooper, R., Fausey, N., Streeter, J. (1991). Yield potential of soybean grown under a subirrigation/drainage water management system. Agronomy Journal 83: 884–887

Cooper, R., Fausey, N., Streeter, J. (1992). Effect of water table level on the yield on soybean grown under subirrigation/drainage, Journal of Production Agriculture 5: 180–184

Drury, C., Tan, C., Gaynor, J., Oloya, T., Vab Wesenbeeck I., McKeeney, D. (1997). Optimizing corn production and reducing nitrate losses with water table control-subirrigation, Soil Science Society of America Journal, 61: 889–895

Elhorst, P. (2001). Dynamic models in space and time. Geographical Analysis 33: 119–140

Elhorst, P. (2003) Specification and estimation of spatial panel models, International Regional Science Review 26: 244–268

Elhorst, P. (2005). Unconditional maximum likelihood estimation of dynamic models for spatial panels. Geographical analysis 37: 62–83

Evans, R., Skaggs, W. (1996). Operating controlled drainage and subirrigation systems. North Carolina Cooperative Extension Service: AG 356

Fisher, M., Fausey, N., Subler, S., Brown, L., Bierman, P. (1999). Water table management, nitrogen dynamics and yields of corn and soybeans, Soil Science Society of America Journal 63: 1786–1795

Florax, R., Voortman, R., Brouwer, J. (2002). Spatial dimensions of precision agriculture: a spatial econometric analysis of millet yield on Sahelian coversands. Agricultural Economics 27: 425–443

Frankenberger J., Kladivko, E., Sands, G., Jaynes, D., Fausey, N., Helmers, M., Cooke, R., Strock, J., Nelson, K., Brown, L. (2007). Questions and answers about drainage water management in the Midwest. Purdue Extension document WQ 44, available at http://www.ces.purdue.edu/extmedia/WQ/WQ-44.pdf

Gilliam, J., Baker, J., Reddy, K. (1999). Water quality effects of drainage in humid regions, Cap. 24 in R. W. Skaggs and J. Van Schilfgaarde (eds.). Agricultural Drainage, Agronomy Monograph 38, American Society of Agronomy, Madison, WI

Haining, R. (2003). Spatial data analysis: theory and practice. Cambridge, Cambridge University Press

Heady, E., Dillon, J. (1972). Agricultural production functions. Iowa State Univ. Press

Holtz-Eakin, D. (1994). Public-sector capital and the productivity puzzle. Review of Economics and Statistics 76: 12–21

Hsiao, C. (1986). Analysis of panel data. Cambridge University Press, Cambridge

Kapoor, M., Kelejian, H., Prucha, I. (2007). Panel data models with spatially correlated error components. Journal of Econometrics 140(1): 97–130

Kaspar, T., Colvin, T., Jaynes, D., Karlen, D., James, D., Meek, D. (2003). Relationship between six years of corn yields and terrain attributes. Precision Agriculture 4: 87–101

Kelejian H., Robinson, D. (1992). Spatial autocorrelation: a new computationally simple test with an application to per capita county police expenditures. Regional Science and Urban Economics 22: 317–331

Lambert, D.M., Lowenberg-DeBoer, J., Malzer, G.L. (2006). Economic analysis of spatial-temporal patterns in corn and soybean response to nitrogen and phosphorus. Agronomy Journal 98: 43–54

Liu, Y., Swinton, S., Miller, N. (2006). Is site-specific yield response consistent over time? Does it pay? American Journal of Agricultural Economics 88: 471–483

Lowenberg-DeBoer, J., Griffin, T., Florax, R. (2006). Use of cross regression to model local spatial autocorrelation in precision agriculture. Site specific management center, Department of agricultural economics, Purdue University.

Mamo, M., Malzer, D., Mulla, D., Huggins, D., Strock, J. (2003). Spatial and temporal variation in economically optimal nitrogen rate for corn. Agronomy Journal 95: 958–964

Malzer, G., Copeland, J., Davis, J., Lamb, J., Robert, P. (1996). Spatial variability of profitability in site-specific N management. *In* P. Robert, R. Rust, W. Larson, eds. Proceedings of the third international conference on precision agriculture. American society of agronomy, crop science society of America, Soil science society of America, Madison, WI.

Moulton, B. (1986). Random group effects and the precision of regression estimates. Journal of Econometrics 32: 385–397

Moulton, B. (1987). Diagnostics for group effects in regression analysis. Journal of Business and Economic Statistics 5: 275–282

Nistor, A. (2007). The impact of controlled drainage adoption: a spatial panel model using yield monitor data. PhD dissertation, Purdue University

Nistor, A., Lowenberg-DeBoer, J. (2007). Drainage water management impact on farm profitability. Journal of soil and water conservation 62: 443–446

Pace, K., (1988). Spatiotemporal autoregressive models of neighborhood effects. Journal of Real Estate Economics 17: 15–33

Pinkse, J., Slade M., Brett, C. (2002). Spatial price competition: a semiparametric approach. Econometrica 70: 1111–1153

Pitts, D. (2003). Illinois drainage water management demonstration project, presentation at the Indiana Drainage Water Management Meeting, West Lafayette, IN

Rablais, N., Turner, R., Scavia, D. (2002). Beyond science into policy: Gulf of Mexico hypoxia and the Mississippi River. Bio Science 52: 129–142

Sipp, S., Lembke, W., Boast, C., Peverly, J., Thorne, M., Walker, P. (1986). Water management of corn and soybeans on a claypan soil. Transactions of the ASAE 29: 780–784

Swanson, E. (1963). The static theory of the firm and the three laws of plant growth. Soil Science 88: 338–343

Tan, C., Drury, C., Soultani, M., Van Wesenbeeck, I., Ng, H, Gaynor, J., Welacky, T. (1988). Effect on controlled drainage and tillage in soil structure and tile drainage nitrate loss at the field scale. Water Science and Technology 38: 103–110