Using Spatial Analysis to Study the Values of

Variable Rate Technology and Information

David S. Bullock
James Lowenburg-DeBoer

**Invited paper prepared for presentation at the
International Association of Agricultural Economists Conference,
Gold Coast, Australia, August 12-18, 2006**

Using Spatial Analysis to Study the Values of Variable Rate Technology and Information

David S. Bullock, University of Illinois, Urbana-Champaign

James Lowenburg-DeBoer, Purdue University

Abstract. We present a review of the last few years' literature on the economic feasibility of variable rate technology in agriculture. Much of the research on this topic has involved the estimation of site-specific yield response functions. Data used for such estimations most often inherently lend themselves to spatial analysis. We discuss the different types of spatial analyses that may be appropriate in estimating various types yield response functions. Then, we present a taxonomy for the discussion of the economics of precision agriculture technology and information. We argue that precision agriculture technology and information must be studied together since they are by nature economic complements. We contend that longer-term, multi-location agronomic experiments are needed for the estimation of ex ante optimal variable input rates and the expected profitability of variable rate technology and information gathering. We use our taxonomy to review the literature and its results with consistency and rigor.

Introduction and Background

Bullock, Lowenberg-DeBoer, and Swinton (2002) emphasized that for farmers' demand for precision agriculture (PA) technology to grow, future research would have to produce two specific types of information. The first type of research is run long-run, wide-ranging agronomic experiments to estimate a yield-response "meta-function" revealing the relationships between yield, managed inputs, field characteristics, and weather variables. The second type of research is long-run agronomic experiments on particular farms. Over the past five years, a number of research projects have been begun to produce both these types of information, principally using spatial econometric techniques to estimate crop yield response-to-fertilizer functions. Current research issues include the estimation of the values of information with and without precision agricultural technology, the modeling of temporal correlation in spatial data sets, experimental design (classical strip trials vs. replicated large block design using spatial statistical analysis), and the comparison of spatial lag versus spatial error models. We review, discuss, and critique the current state of the research, emphasizing that precision agriculture technology itself greatly lowers the cost of on-farm experimentation, thus shifting out the supply-of-information curve, which will result in increased adoption of precision agriculture technology.

Spatial Regression Alternatives

Classical statistics applied to agronomic and on-farm experiments assume that observations are independent. But in the case of PA data this assumption of independence is untenable. Data from crop yield sensors and soil sensors is almost always spatially correlated. Kessler and Lowenberg-DeBoer (1998) showed that spatial

correlation is an issue even when 2.5 acre grid soil sample points (separated by roughly 330 feet) are used. When spatial correlation is present field heterogeneity may be underestimated, and inferences about crop response to variable fertilizer rates may be misleading.

Spatial dependence occurs when the dependent variable or error term at any location is correlated with observations of the dependent variable or error terms at other locations (Anselin, 1992). Regression methods that model spatial correlation have been developed in a variety of disciplines (e.g., geography, agronomy, regional economics, and geology). A key difference in these methods is whether spatial relationships between observations are best described as *discrete* or *continuous* (for details on this distinction, see Anselin (1988)).

Other differences revolve around estimation methods (e.g., OLS, maximum likelihood, general method of moments) and how data at different spatial scales are combined for analysis. For example, raw yield data is often available at 5 to 10 feet between observations within the pass, and at a header width between passes (10 to 40 feet) , while soil sample information is often only available from grid sampling (often over 300 feet between samples). In some disciplines interpolated data is used to estimate the sparse data layer at the resolution of the dense data layer. Other disciplines are concerned about the spatial patterns that might be introduced by interpolation and prefer to aggregate the dense data layers to the resolution of the least dense layer. This might be done by a simple or weighted average of observations in the more dense layers over a certain search distance around each data point in the least dense layer.

The most common spatial regression techniques are: (i) a restricted maximum likelihood (REML) geostatistical approach (GEO) (Cressie, 1993; Schabenberger and Pierce, 2002); (ii) a spatial regression approach using polygons as discrete units of observation (or spatial autoregression, SAR) (Anselin, 1988); (iii) a polynomial trend (PTR) approach (Tamura et al., 1988); and (iv) a classical nearest neighbor (NN) approach first suggested by Papadakis (1937).

Nearest Neighbor

Nearest neighbor is the oldest approach to spatial statistical analysis. Fisher introduced randomized complete block designs in the 1920s as a way to deal with spatial heterogeneity. Papadakis (1937) responded to Fisher's blocking methodology with the nearest-neighbor approach (NN). Brownie et al. (1993) describe the NN model proposed by Papadakis as:

(1) $$Y_{ij} = \mu + t_{ij} + ?z_{ij} + e_{ij},$$

where Y is yield, μ is the overall mean yield, t_{ij} is the treatment effect, z_{ij} is the set of nearest neighbor residuals perpendicular to y_{ij}, and $?$ is a slope coefficient of the covariance between the residual errors of yield y_{ij} and its z_{ij} neighbors. The residual error differences are expressed as $r_{ij} = y_{ij} - \hat{Y}_k$, where $\hat{Y}_k$ is the overall mean for treatment k. The average of the NN residuals for y_{ij} is determined as $z_{ij} = (r_{i,j-1} + r_{i,j+1} + r_{i-1,j} + r_{i+1,j})/4$. The structure of the NN model as expressed in (1) is that of the familiar ANOVA model commonly used to test for treatment differences for on-farm trials. Equation (1) can be generalized into the familiar regression model by inserting the z_{ij} into an n x k matrix of explanatory variables, **X**. This re-specification is important since the primary interest of this study is to estimate site-specific yield response to nitrogen (N). The NN model

becomes $\mathbf{y} = \mathbf{X\beta} + ?\mathbf{z} + \mathbf{e}$, where the covariance parameter $?$ is an averaging parameter

for the neighborhood of residual errors perpendicular to observation y_{ij}. In this modified

NN model, $?$ explains the residual error caused by spatial structure. Equation (1) is

estimated with OLS.

Polynomial Trend Regression

Tamura, et al. (1988) proposed another alternative to modeling spatial dependence by

inserting a polynomial trend variable (T_{ij}) into the familiar ANOVA model $Y_{ij} = \mu + T_{ij} +$

e_{ij}. This approach is somewhat related to the spatial expansion regression methodology

that has received attention in urban and regional geography (Anselin, 1988). A trend

surface is introduced into the model specification to capture spatial relationships between

observations. This approach assumes that omission of spatial dependence is analogous to

the omitted variable problem in the econometric literature. The omitted variable problem

is handled by inclusion of trend variables in ANOVA models. Like the NN method

proposed by Papadakis, Tamura et al.'s PTR model was developed to account for

spatially structured error processes not dealt with by conventional blocking techniques.

The simultaneous estimation of a polynomial response surface with the regression model

separates systematic error components caused by spatial dependence from the

unsystematic portion of e_{ij} (Kirk, et al., 1980). Parameter estimates are derived only with

respect to remaining random components, the e_{ij}. In effect, addition of a system of

coordinates relating observation *i* to *j* into the familiar regression model $y = X\beta + e$

expressed in terms of polynomials eliminates the omitted variable problem, assuming the

trend surface specified by the polynomial expression is the correct specification. The

omitted variable(s) in question would be those that explain spatial structure in error residuals.

The PTR model is specified as

(2) $$Y_{ij} = \mu + t_{k(ij)} + T_{ij} + e_{ij},$$

where Y is the yield, μ is the overall mean, t_k is the treatment effect, T is a polynomial trend, and e is an independent and identically distributed (i.i.d.) random error component. The quadratic trend term is estimated as $T_{ij} = f_1 x + f_2 y + f_3 x^2 + f_4 y^2 + f_5 xy$, where f_i is a slope coefficient for the Cartesian (x, y) coordinate of observation y_{ij}. The (x, y) coordinates are expressed as row/column pairs. That is, the first observation in the first row $i = 1,\ldots,n$ and column $j = 1, \ldots,m$ is identified with the coordinates $(1,1)$, and the last observation is identified with coordinates (m, n). Like the NN approach, the PTR method was developed for ANOVA of treatment effects controlling for spatial dependence.

Geostatistical Approach

Many agronomists have used geostatistical tools to model crop and soil spatial relationships. Perhaps this is because of the disciplinary links between soil science and geology. Originally, geostatistics was developed to produce maps by interpolation between observations. To facilitate mapping, geostatistics assumes that spatial variability is a continuous function of distances modeled by a semivariogram. Within the geostatistical framework inferential testing of the relationships between variables (for example, layers in the crop GIS) at a given point has been developed relatively recently. Cressie (1993) introduced the REML-geostatistical approach. Little, et al. (1996) and Schabenberger and Pierce (2002) elaborated upon this approach, which entails estimating

empirical semivariograms, and then using semivariogram parameter estimates as priors in a regression model to characterize spatial correlation between observations.

The semivariogram is the backbone of the REML-geostatistical regression model. The semivariogram parameters (range, nugget, and sill) are estimated and then used as priors to model the regression covariance matrix. The regression model is estimated with the familiar model $\mathbf{y} = \mathbf{X\beta} + \mathbf{e}$, but spatial covariance ($\mathbf{R}$) is modeled through $\mathbf{R} = \mathrm{Var}(\mathbf{e})$, where $\mathrm{Var}(\mathbf{e}) = \mathbf{I}\, s_n^2 + s_s^2 \mathbf{F}$, $\mathbf{F}$ is an $N \times N$ matrix whose i,j-th element is characterized by a distance decay function, and s_n^2 and s_s^2 are nugget and sill semivariogram estimates (Little, et al., 1996). The REML parameter estimates are estimated generalized least squares (EGLS) estimates adjusted for spatial autocorrelation.

Discrete Spatial Regression

The discrete spatial regression approach assumes that spatial dependence is a relationship among discrete observations, or polygons. Spatial structure may be found in either the dependent variable (e.g., yield) or in regression residuals. Spatial structure is modeled assuming that the dependent variable or residuals are a function of a weighted average of neighboring observations. This approach has been used extensively in epidemiology, geography, and regional economics. In agriculture the structure of the data is similar, but the polygons are often soil types or management zones instead of states, counties, districts, or neighborhoods. This approach uses polygon data, enabling the simultaneous maximum likelihood estimation of the spatial structure and the relationships between GIS layers.

A spatial weights matrix is constructed to identify neighbors in a dataset. The matrices are designed to incorporate processes such as gravity, entropy, or decay into

regression models (Anselin, 1988). Data arranged in regular rectangular lattices are defined using three criteria: bishop, rook, or queen. These classes describe the level of contiguity, or common boundaries, between polygons. In spatial terms, contiguity is defined as a function of the distance that separates one cell from another. Blocks belonging to the same neighborhood share the same weight, and the composite of neighborhoods covering the entire grid defines the spatial weights matrix. This matrix ($\mathbf{W}$) is an $N \times N$, positive definite matrix with elements w_{ij}, with zeroes along the diagonal. Before spatial weights matrices are used to estimate spatial effects in regression models, they are row-standardized. This facilitates comparison of spatial characteristics across neighborhoods. Each element in a row is divided by the row sum.

Anselin (1988) identifies two general patterns whereby spatial dependence may manifest itself in regression analysis: spatial *lag* and spatial *error*. If spatial error processes are ignored, OLS estimates are inefficient, but remain unbiased. If spatial lag processes are ignored, then OLS estimates are inconsistent and biased. For lag processes, the modified regression model becomes $\mathbf{y} = ?\,\mathbf{Wy} + \mathbf{X\beta} + \mathbf{e}$; with ? as the autoregressive moving average parameter for neighboring y_j's. The spatial error model is specified as $\mathbf{y} = \mathbf{X\beta} + \mathbf{e}$ with $\mathbf{e} = ?\,\mathbf{W\,e} + \mathbf{u}$, where $\mathbf{u}$ represents well-behaved, non-heteroskedastic, uncorrelated errors. The β_i's are EGLS estimates corrected for spatial autocorrelation.

A Taxonomy

Ruffo et al. (2006) present a taxonomy for the discussion of the economics of precision agriculture and information, which builds upon the ideas presented in Bullock and

Bullock (2000) and Bullock, et al. (2002) by modeling information. Here we briefly present that taxonomy. We refer the reader to Ruffo et al. (2006) for details.

Response Functions

Model crop yield y on any small, uniform piece of ground as a function of a managed inputs $\mathbf{x} = (x_1, \ldots, x_J)$, unmanaged spatially dependent field characteristics $\mathbf{c} = (c_1, \ldots, c_K)$, and unmanaged time-dependent variables $(\mathbf{z} = (z_1, \ldots, z_L)$:

(3) $$y = f(\mathbf{x}, \mathbf{c}, \mathbf{z}).$$

We refer to $\mathbf{z}$ as "weather," though pest infestations and other variables may be included. Function $f(\mathbf{x}, \mathbf{c}, \mathbf{z})$ is *nature's meta-response function*.

Subdivide a producer's field into *sites*, indexed $s = 1, \ldots, S$. We assume that each site is small enough so that the characteristics vector $\mathbf{c}$ takes on the same value everywhere on a site. Let c_k^s be the level of characteristic $k \in \{1, \ldots, K\}$ at site $s \in \{1, \ldots, S\}$. Then $\mathbf{c}^s = (c_1^s, \ldots, c_K^s)$ be the vector of characteristic levels of site s, and $\mathbf{c}' = (\mathbf{c}^1, \ldots, \mathbf{c}^S)$ is the entire field's *characteristics map*. Let $\mathbf{x}^s = (x_1^s, \ldots, x_J^s)$ be the vector of management inputs used on site s. Then $\mathbf{x}' = (\mathbf{x}^1, \ldots, \mathbf{x}^S)$ is the entire field's *management map*.

The *site-specific response function* at site s is a function of managed inputs and weather:

(3) $$f^s(\mathbf{x}, \mathbf{z}) \equiv f(\mathbf{x}, \mathbf{c}^s, \mathbf{z}), \, s \in \{1, \ldots, S\}.$$

The *map of site-specific response functions* is $\left(f^1(\mathbf{x}, \mathbf{z}), \ldots, f^S(\mathbf{x}, \mathbf{z})\right)$. Should $\mathbf{c}^s$ be the same at every site in the field, then there exists a field-specific response function $f^{field}(\mathbf{x}, \mathbf{z}) \equiv f^1(\mathbf{x}, \mathbf{z}) \equiv \ldots \equiv f^S(\mathbf{x}, \mathbf{z})$.

Assume that weather is not spatially variable within a field in a given year t, and let $\mathbf{z}^t$ denote the field's weather in year t. The *site-and-year-specific response function* at site s in year t is a function of managed inputs:

(4) $$f^{s,t}(\mathbf{x}) \equiv f\left(\mathbf{x}, \mathbf{c}^s, \mathbf{z}^t\right), \; s \in \{1,...,S\}.$$

The *map of site-and-year-specific response functions* in year t is $\left(f^{1,t}(\mathbf{x}),...,f^{S,t}(\mathbf{x})\right)$.

The term *climate* and the function $h(\mathbf{z})$ denote nature's joint probability density function of weather on the field. Let H denote the support of $h(\mathbf{z})$. Assume that the producer knows the climate of his field, though before he makes decisions he may not know the weather, which is the draw to be taken from the climate.

States of Nature and Information-and-Technology Structures without Noise

In general, when he makes decisions in year t, the producer does not know his site-and-year specific response function map with certainty. Nor does he know with certainty what the weather will end up being during the growing season. Rather, he lacks information. We refer to Laffont (1990, chapter 4) to formalize what we mean by information. (Our presentation here will be very brief. Details are provided in Ruffo, et al. (2006).) Let $\Omega = \{w_1, w_2,...,w_N\}$ be our model's *set of states of nature*, defined as the set of all conceivable site-and-year specific response function maps. For simplicity we assume that there are a finite number N states of nature, though N may be very large. Following Laffont (1990), any partition of Ω is an *information structure without noise*. For example, if the farmer faces the partition $\{\{w_1, w_3\}, \{w_2, w_4,...,w_N\}\}$, it is as if he has hired an expert who will give exactly one of two responses: when the true site-and-year-specific response function map is in $\{w_1, w_3\}$ the expert will tell him so, and if the true site-and-year-specific response

11

function map is in $\{w_2, w_4, ..., w_N\}$ the expert will tell him so. Another expert might

partition Ω more finely, say providing information structure $\{\{w_1, w_3\}, \{w_2\}, \{w_4, ..., w_N\}\}$,

meaning that if the true production map is in $\{w_1, w_3\}$, the producer will be told, if it is w_2

the producer will be told, and if it is in $\{w_4, ..., w_N\}$, the producer will be told.

We define an *information-and-technology structure* as an ordered pair, with the first

element being an information structure, and the second element being a technology level.

There are several information-and-technology structures assumed in the literature. They

can be efficiently denoted using the form $(\Phi_{d,h,s}, t)$, with $d \in \{ante, post\}$, $h \in \{\exp,$

noexp$\}$, $s \in \{$STAN, REML, SAR-ERR, SAR-LAG, SAR-GEN, PTR, NN, NULL$\}$, and t

$\in \{URT, VRT\}$. Here d denotes whether production decisions are made *ex ante* or *ex post*,

h denotes whether a site-specific agronomic experiment is run to gather data, s represents

the technique that was used to estimate the response function (either the standard OLS

approach (STAN), the restricted maximum likelihood geostatistical approach (REML), the

spatial autoregressive approach that uses spatial errors (SAR-ERR), etc.), and t denotes the

technology, either variable rate or uniform rate, that the producer uses. For example,

$\Phi_{post,exp,sar-lag}$, is the set of possible site-specific response function maps when the weather is

known before management decisions are made (so decisions are made *ex post*), an

agronomic experiment has been run, and the data from the experiment have been analyzed

using spatial lag techniques. Similarly, $\Phi_{ante,noexp,null}$, is the set of possible site-specific

response function maps when the weather is not known before management decisions are

made (so decisions are made ex *ante*), no agronomic experiment has been run (so perhaps

the producer follows university-recommended managed input rates), and since there is no

experiment, then there is no data from the experiment to analyze, so s = null. The

economically optimal management map will depend on the information and technology

structure, and so is denoted $\mathbf{x}'^{*}(\Phi_{d,h,s}, t) = (\mathbf{x}^{1*}(\Phi_{d,h,s}, t), \dots, \mathbf{x}^{S*}(\Phi_{d,h,s}, t))$. The studies in

the literature generally assume that the producer's objective is to maximize expected

profits. (If decisions are made ex post, then there is no uncertainty in the model, and

expected profits are simply profits.) Letting p denote the output price, total optimal

expected revenues on the entire field are $R^{*}(\Phi_{d,h,s}, t) = \int_{H} \sum_{s=1}^{S} pf^{s,s} \left(\mathbf{x}^{s*}\left(\Phi_{d,h,s}, t\right), \mathbf{z} \right) h(\mathbf{z}) d\mathbf{z}$.

Letting $\mathbf{w} = (w_1, \dots, w_J)$ denote managed input prices, optimal managed input costs for the

entire field are $C_x^{*}(\Phi_{d,h,s}, t) = \sum_{s=1}^{S} \mathbf{w}\mathbf{x}^{s*}\left(\Phi_{d,h,s}, t\right)$. In some studies, the costs of employing

the technology employed, C_t, and the costs of running the experiment and analyzing the

data, C_s, are modeled. Optimized expected profits are $P^{*}(\Phi_{d,h,s}, t) = R^{*}(\Phi_{d,h,s}, t) -$

$C_x^{*}(\Phi_{d,h,s}, t) - C_t, - C_s$. Optimized expected gross margins over application costs of a

particular input, say input j, are $GM_j^{*}(\Phi_{d,h,s}, t) = R^{*}(\Phi_{d,h,s}, t) - \sum_{s=1}^{S} w_j x_j^{s*}\left(\Phi_{d,h,s}, t\right)$.

Using the Taxonomy to Review Recent Literature on the Value of Variable Rate

Technology and Information

Peone, et al. (2004) have provided a detailed review of the literature on the profitability of

precision agriculture up through the year 2003. In the following, we will update that

review, using our taxonomy freely for brevity, clarity, and rigor.

Anselin, Bongiovanni, and Lowenberg-DeBoer (2004)

Anselin, Bongiovanni, and Lowenberg-DeBoer (ABL) (2004) wanted to find a low-cost technique providing evidence of spatial heterogeneity in a field that can lead to more efficient estimation of site-specific response functions. As an inexpensive alternative to plugging estimated site characteristics $\mathbf{c}^s$ and the year's weather $\mathbf{z}^t$ into an estimate of nature's meta-response function $f(\mathbf{x}, \mathbf{c}, \mathbf{z})$ to obtain estimated site-and-year-specific response functions $f(\mathbf{x}, \mathbf{c}^s, \mathbf{z}^t) \equiv f^{s,t}(\mathbf{x})$, they used different landscape positions to define sites for which they directly estimated site-and-year specific yield response functions $f^{s,t}(\mathbf{x})$ from experimental data.

ABL examined the spatial error model and the spatial lag model, and deemed the former superior for their data. They argued that many characteristics that affect yield response to managed inputs, such as subsoil characteristics, vary spatially. But when such characteristics are unobserved, their contribution to yield variation is subsumed in the error term, and therefore the error term should display a spatial pattern. They described their spatial error model (SAR) as a "spatial regimes model with spatial autoregressive error term as well as groupwise heteroskedasticity" (p. 682).

ABL assumed throughout their welfare analyses that the SAR model perfectly estimates site-specific response functions. They employ three information structures. All three are *ex post* information structures (that is, the producer is assumed to have no uncertainty about what the yield response functions are, even though he may be wrong about them). Under information structure $\Phi_{post,exp,sar\text{-}err}$ the producer knows the estimated parameters from the SAR model, and believes that the SAR model perfectly estimates site-specific response functions. Under information structure $\Phi_{post,exp,stan}$ the producer believes that the standard model perfectly estimates the site-specific response functions. Under

information structure $\Phi_{post,noexp,null}$ the producer believes that the economically optimal N

rate is the university-recommended rate.

ABL stated (p. 675) their specific objectives to be the following.

1) "To analyze the potential value of spatial econometrics in the

estimation of site-specific crop N responses from yield monitor data."

That is, in terms of our taxonomy, they wished to estimate $[\boldsymbol{P}^{*}(\Phi_{post,exp,sar\text{-}err,}, URT)$ -

$\boldsymbol{P}^{*}(\Phi_{post,exp,stan}, URT)]$ and $[\boldsymbol{P}^{*}(\Phi_{post,exp,sar\text{-}err,}, VRT)$ - $\boldsymbol{P}^{*}(\Phi_{post,exp,stan}, VRT)]$.

2) "[t]o estimate the profits from site-specific N management using the

yield responses estimated from both spatial and non-spatial models."

That is, in terms of our taxonomy, they wished to estimate $[\boldsymbol{P}^{*}(\Phi_{post,exp,sar\text{-}err,}, VRT)$ -

$\boldsymbol{P}^{*}(\Phi_{post,exp,sar\text{-}err}, URT)]$ and $[\boldsymbol{P}^{*}(\Phi_{post,exp,stan}, VRT)$ - $\boldsymbol{P}^{*}(\Phi_{post,exp,stan}, URT)]$.

3) "[t]o compare profits from site-specific N management using crop

response functions with uniform rate management and VRT management

strategies."

That is, in terms of our taxonomy, they wished to estimate $[\boldsymbol{P}^{*}(\Phi_{post,exp,sar\text{-}err,}, VRT)$ -

$\boldsymbol{P}^{*}(\Phi_{post,noexp,null}, URT)]$ and $[\boldsymbol{P}^{*}(\Phi_{post,exp,sar\text{-}err,}, VRT)$ - $\boldsymbol{P}^{*}(\Phi_{post,noexp,null}, VRT)]$.

To pursue these objectives, ABL conducted a one-year, one-field agronomic

experiment in Córdoba, Argentina, gathering data on corn yield response to nitrogen

fertilizer. They varied nitrogen fertilizer rates in their experiment, and applied all other

managed inputs at conventional levels. Their experiment was set up with three repetitions

of six N rates in each of four landscape positions (sites). For each of various models and

econometric methods, they estimated four site-and-year specific response functions,

assuming a quadratic functional form: $f^{s,\boldsymbol{s}}\left(N_{s,j}\right) = \boldsymbol{a}_{s,\boldsymbol{s}} + \boldsymbol{b}_{s,\boldsymbol{s}} N_{s,j} + \boldsymbol{g}_{s,\boldsymbol{s}} N_{s,j}^{2} + \boldsymbol{e}_{s,\boldsymbol{s}}$, where s

is the site index (s = Slope W, Hilltop, Slope E, Low E, or s = 1, 2, 3, 4), j is the repetition

index, $N_{s,j}$ is the amount of nitrogen fertilizer applied in replication j on site s, $\boldsymbol{s}$ is the

index of the model/econometric method, and $\boldsymbol{e}_{s,j}$ is the error term for that observation. For

the various models/econometric techniques $\boldsymbol{s}$, they calculated *ex-post* site-specific

economically optimal N rate maps, $N^*(\Phi_{post,exp,\boldsymbol{s}}, VRT)$, $N^*(\Phi_{post,exp,\boldsymbol{s}}, URT)$. They used a

university-recommended rate for $N^*(\Phi_{post,noexp,null,}, URT)$. Principally, they estimated the

model using standard (OLS) and SAR specifications (their results are shown in their table

1, p. 682). They also conducted sensitivity analysis by using twelve different combinations

of model specification, spatial weights, and estimation techniques. In all cases, they

obtained the same main result that SAR models estimated higher returns from VRT greater

than the $6 ha^{-1} that the authors used as a benchmark for VRT application costs.

They used net revenues per ha for the producer utility function. In their table 2 (p.

684), they reported the difference in net revenues per ha when a profit-maximizing

producer using uniform rate technology uses the spatial error model instead of the standard

model to estimate the response function map. Put in terms of our taxonomy, they reported

$[\boldsymbol{P}^*(\Phi_{post,exp,sar-err,}, URT) - \boldsymbol{P}^*(\Phi_{post,exp,stan}, URT)] = \$2.62$. Similarly, they reported the

difference in net revenues per ha when a profit-maximizing producer using variable rate

technology uses the spatial error model instead of the standard model to estimate the

response function map: $[\boldsymbol{P}^*(\Phi_{post,exp,sar-err,}, VRT) - \boldsymbol{P}^*(\Phi_{post,exp,stan}, VRT)] = \$4.78$.

Since spatial regression provides better estimations of response functions that do

standard OLS methods, the results above illustrate the argument made in Bullock and

Bullock (2000) and in Bullock, Lowenberg-DeBoer, and Swinton (2002) that information

and variable rate technology are economic complements: switching from uniform to

variable rate technology increases the value of obtaining better information from $2.62 ha^{-1}

to $4.78 ha^{-1}. The implication is that when we lower the marginal costs of obtaining

information (shift the information supply curve out), we in turn shift the schedule of

marginal value product of variable rate technology—we shift the VRT demand curve out.

This implies that spatial econometrics may play a role in increasing the adoption of

variable rate technology.

Ruffo, et al. (forthcoming)

The objectives of the Ruffo, et al. (forthcoming) were to develop site-and-year-specific

corn yield response functions for variable fertilization, and to determine the characteristic

variables affecting corn response to N fertilizer. The authors conducted agronomic

experiments on eight different commercial production fields (four in 2002 and four in

2003). Fields were divided into 13 to 20 sites, each composed of five plots. Each plot

received one N fertilizer rate. Five fertilizer rates were applied, two below the University

of Illinois recommended rate, one at that rate, and two above that rate.

Site-and-year-specific response functions were estimated by multiple regression

maximum likelihood procedures. Model development started with a second-degree

polynomial for N fertilizer rate, and then the characteristic variables were added starting

with the variable that had the largest correlation with yield. Once a characteristic was

allowed into the model, its interaction with the linear term for N fertilizer rate was tested.

If significant, its interaction with the quadratic term for N fertilizer was also tested.

Characteristic variables included in the regressions were primary and secondary

terrain attributes, and the Illinois Soil Nitrogen Test (ISNT). Nitrogen fertilizer

significantly increased corn yield. The ISNT characteristic variable interacted with N

fertilizer in most fields. In general, parameter estimates indicated that well drained areas

(i.e. small specific catchment area, moderate slopes) had higher yields and responded less

to N fertilizer than did poorly drained areas. The authors stated that these results indicate

that terrain attributes as proxies for soil water content and the ISNT as a measure of soil

mineralizable nitrogen are site-specific characteristics that affect corn yield and its

response to N fertilizer.

The spatial structure of the regression model was tested with a likelihood-ratio test

between the spatial covariance model and an independent error model, with the same

variables included in the fixed effect. Normality and homoskedasticity of the errors were

assessed.

In both years and in all fields, the response of corn yield to N fertilizer was

quadratic and N fertilizer interacted with at least one site-specific characteristic.

The authors state that their results indicate that ISNT has a spatial structure that will

allow mapping with a relatively sparse grid (approximately a 1 ha grid), that the soil

samples for ISNT would not need to be collected every year but rather every 4 or 5 and

that therefore ISNT is a suitable soil test for VRN.

Ruffo, et al. (2006)

Ruffo et al. (2006) used the estimated response functions reported in Ruffo et al.

(forthcoming) to conduct economic analysis. Specifically, their aims were to estimate for

each field i) the ex-ante optimal uniform application rate with site-specific experimental

information, $N^{field*}(\Phi_{ante,esp}, sar\text{-}err, URT)$; ii) the ex-ante optimal uniform application rate

without site-specific experimental information, $N^{field*}(\Phi_{ante,noexp,null}, URT)$; iii) the ex-ante

optimal variable application map with site-specific experimental information,

$\mathbf{N}^{\mathbf{field´*}}(\Phi_{ante,esp}, sar\text{-}err, VRT)$; iv) the value of site-specific information with URT,

$[\boldsymbol{P}^{*}(\Phi_{ante,exp,sar\text{-}err,}, URT) - \boldsymbol{P}^{*}(\Phi_{ante,noexp,null}, URT)]$; v) the ex-ante value of VRT given

site-specific information, $[\boldsymbol{P}^{*}(\Phi_{ante,exp,sar\text{-}err,}, VRT) - \boldsymbol{P}^{*}(\Phi_{ante,exp,sar\text{-}err,}, VRT)]$; and vi) the

ex-ante combined value of site-specific information and of VRT, $[\boldsymbol{P}^{*}(\Phi_{ante,exp,sar\text{-}err,}, VRT)$

$- \boldsymbol{P}^{*}(\Phi_{ante,noexp,null}, URT)]$.

      Ruffo et al. (2006) estimated *ex ante* values by including weather variables in

their response function estimates, and then using historical weather data in a simulation.

Their data came from farms in two different counties over two years, and they used it to

estimate what they called "county meta-response functions," $f^1(\mathbf{x}, \mathbf{c}, \mathbf{z})$ and $f^2(\mathbf{x}, \mathbf{c}, \mathbf{z})$. In

their regressions the vector of managed variables $\mathbf{x}$ had a single element, N. Vector $\mathbf{c}$ had

as elements the topographical attributes and the *ISNT* variable. Initially, precipitation

amounts in each month of the calendar year made up the elements of $\mathbf{z}$. Then for each

field they obtained site-specific response functions by inserting the site-specific

characteristics $\mathbf{c}$ into the estimated county meta-response function, as in equation (3).

They state that because of the small number of years in their experiment, they assumed

the quadratic functional form in their regressions.

      Ruffo, et al. (2006) reported that their experiments and analysis implied an

average (across the eight fields) value of site-specific information under uniform rate

technology was $[\boldsymbol{P}^{*}(\Phi_{ante,exp,sar\text{-}err,}, URT) - \boldsymbol{P}^{*}(\Phi_{ante,noexp,null}, URT)] = \$0.28$ ha^{-1}. The

average value of site-specific information under variable rate technology was

$[\boldsymbol{P}^{*}(\Phi_{ante,exp,sar\text{-}err,}, VRT) - \boldsymbol{P}^{*}(\Phi_{ante,exp,sar\text{-}err,}, VRT)] = \$ 1.90$ ha^{-1} and ranged from \$0.57

ha^{-1} to \$4.86 ha^{-1} among fields Finally, the average ex-ante value of site-specific

information and of VRT was [$\boldsymbol{P}^{*}$($\Phi_{ante,exp,sar-err,}$, *VRT*) - $\boldsymbol{P}^{*}$($\Phi_{ante,noexp,null}$, *URT*)] = \$2.16

ha^{-1}, and ranged from \$0.67 ha^{-1} to \$5.11 ha^{-1}.

Ruffo, et al. (2006) state that their results predict that while VRT would not be

profitable in every field, there was enough variability in most fields for VRT to be

profitable. Most of the value of site-specific information and of VRT was provided by

VRT, since the value of site-specific information was very small for all their fields. This

is the case even though VRT would be useless without the site-specific information and

the meta-response functions that relate site-specific characteristics, N rate weather, and

corn yield.

Liu, Swinton, and Miller (2006)

Liu, Swinton, and Miller (LSM) (2006) ran agronomic experiments to study corn yield

response to N fertilizer rates in Michigan. Their experiments were typically run on forty

acres, and took up eight fields in 1999, nine fields in 2000, and seven fields in 2001. Their

unbalanced panel data set represented "a longer time series and wider cross section than

any research published to date" (p. 472). They included both site characteristics and

weather variables in their data, providing the temporal and spatial variability necessary for

simulating profitability of *ex ante* site-specific application rates. (Almost all other studies

have only examined *ex post* rates, though *ex ante* rates are more appropriate for decision

makers.) The characteristic variables they attempted to measure were organic matter,

cation exchange capacity, electrical conductivity (a proxy for potential soil moisture), water

availability (with a soil wetness index variable as its proxy), and sunlight reception (with an index of topological variables as its proxy).

LSM tried three functional forms for their response function estimates: the quadratic, the von Liebig linear-plateau, and the von Liebig quadratic-plateau. Conducting statistical inference, they could not reject the quadratic functional form, and so conducted their economic analysis under the assumption that the yield response functions were quadratic. They ran their regressions using OLS, the spatial error model, the spatial lag model, and the general spatial model. Similar to ABL, LSM found that coefficient estimates under OLS "did not differ significantly from the spatial models estimated" (p. 477), but that the coefficient standard errors were significantly reduced by the spatial models when compared to OLS (p. 478).

LSM (2006) statistically examined three questions: (1) whether corn yield responds to N fertilizer, and whether any such response varies by site; (2) whether corn yield response to N fertilizer varies across seasons; and (3) whether, if a predictive model of yield response can be developed, the site-specific N fertilizer management is likely to be profitable. They addressed these questions within a framework of expected profit maximization, taking into account the costs of gathering information via experimentation, and the costs of hiring variable rate technology equipment. They rejected the hypothesis that corn yield did not respond to N fertilizer, they rejected the hypothesis that site characteristics did not interact with N fertilizer (these results were especially strong in years in which water was not a limiting factor), and they concluded that yield response does vary by site. However, they concluded that while yield response varies by site in some years, it does not do so in a consistent manner across years, and nor is the effect of

individual variables consistent across years.   But they also stated that their results

supported the conjecture when water is not a limiting factor, yield response to N did not

change over seasons.

LSM tested the hypothesis that site-specific research (under various econometric

techniques) and variable rate N fertilizer application together are more profitable than is

uniform rate application. To do this, they ran Monte Carlo simulations, assuming that the

coefficients of their estimated response functions followed a multivariate normal

probability distribution with their estimated mean values and the corresponding covariance

matrix. Then they bootstrapped confidence intervals for the simulated profits. With their

simulations they were able to estimate *ex ante* economically optimal N fertilizer rates for

each field. They examined five N management strategies in particular: (1) running

experiments to gather site-specific data, analyzing the data with the general SAR methods,

and using this information to obtain the site-specific *ex ante* optimal N application map

$\mathbf{N}^{*}(\Phi_{ante,exp,sar\text{-}gen,}, VRT)$; (2) running experiments to gather site-specific data, and

analyzing the data with the general SAR methods to obtain the uniform rate *ex ante* optimal

N application map $\mathbf{N}^{*}(\Phi_{ante,exp,sar\text{-}gen,}, URT)$; (3) forgoing experimentation, and instead

relying on recommended area-wide N fertilization rates to derive the *ex ante* optimal

uniform rate application map given no site-specific information $\mathbf{N}^{*}(\Phi_{ante,noexp,null,}, URT)$;

(4) running experiments to gather site-specific data, analyzing the data with standard OLS

methods, and using this information to obtain the site-specific *ex ante* optimal N

application map $\mathbf{N}^{*}(\Phi_{ante,exp,stan}, VRT)$; and (5) running experiments to gather site-specific

data, and analyzing the data with standard OLS methods to obtain the uniform rate *ex ante*

optimal N application map $\mathbf{N}^{*}(\Phi_{ante,exp,stan,}, URT)$. They also calculated the corresponding

expected gross margins over N application costs: $GM_N^*(\Phi_{ante,exp,sar\text{-}gen,}, VRT)$,

$GM_N^*(\Phi_{ante,exp,sar\text{-}gen,}, URT)$, $GM_N^*(\Phi_{ante,noexp,null,}, URT)$, $GM_N^*(\Phi_{ante,exp,stan,}, VRT)$, and

$GM_N^*(\Phi_{ante,exp,stan,}, URT)$.

In table 6 of p. 481, for each of their eight fields LSM list the bootstrapped 80%

confidence intervals for the following differences in gross expected margins over N

application costs: $[GM_N^*(\Phi_{ante,exp,stan,}, VRT) - GM_N^*(\Phi_{ante,exp,stan,}, URT)]$, $[GM_N^*(\Phi_{ante,exp,sar\text{-}}$

$_{gen,}, VRT) - GM_N^*(\Phi_{ante,exp,sar\text{-}gen,}, URT)]$, $[GM_N^*(\Phi_{ante,exp,stan,}, VRT) - GM_N^*(\Phi_{ante,noexp,null,},$

$URT)]$, and $[GM_N^*(\Phi_{ante,exp,sar\text{-}gen,}, VRT) - GM_N^*(\Phi_{ante,noexp,null,}, URT)]$. They conclude that

in no instances does the difference in gross margins exceed the $5 per acre cost of variable

rate application of nitrogen fertilizer. That is, even if running the experiments and

analyzing the data were costless so that nitrogen fertilizer could be applied optimally, it still

would not pay to adopt variable rate technology on any of the eight field-years in the study.

Moreover, one could gather no more information, but rather rely on the tri-state

recommended algorithm with URT, and still make more money than if one were to gather

the information and pay for VRT.

Kahabka, et al. (2004)

Kahabka, et al. (2004) conducted agronomic experiments on a 12 ha field in central New

York in 1998, 1999, and 2000. Nitrogen fertilizer, tillage method, and rotation were the

managed inputs varied in the experiment. Maize was the output. Other management inputs

were applied at conventional levels. The pre-sidedress soil nitrate test (PSNT) was used to

measure the characteristic variable. Each site received a base application of 50 kg ha^{-1}, and

then later sidedress applications of 0, 60, 110, and 170 kg ha^{-1} were applied. The authors

did not attempt to estimate yield response functions, but instead averaged the observations in each year-tillage-rotation-N-rate, and then used ANOVA to test whether profits differed significantly among N-rates in each year-tillage-rotation. They assumed a nitrogen price to maize price ratio of 0.2. The value of the *ex post* economically optimal N rate varied greatly among years, and rotation affected the optimal N rate only in the dry year 1999. They concluded that "the lack of year-to-year consistency in N response, in terms of the field average and spatial structure, indicate that appropriate site-specific rates cannot be predicted with reasonable accuracy" (p. 473), and "The poorly spatial structures of both PSNT and profit response to N make the site-specific application of N impractical" (p. 474). They also concluded that the PSNT results could not be simply applied to determined site-specific management practices, and that yield potential in itself was not a strong predictor of *ex post* optimal N rates.

Ehlert, Schmerler, and Voelker (2004)

Ehlert, Schmerler, and Voelker (ESV) (2004) attempted to address the high costs that can come with measuring characteristics variables **c**. They emphasized that the high cost of soil sampling may make the value of gathering information and hiring VRT to be less profitable than simply following the conventional uniform rate. (In terms of our notation, for any data analysis method $\boldsymbol{s}$, $[GM_N^*(\Phi_{post,exp,\boldsymbol{s}}, VRT) - GM_N^*(\Phi_{post,noexp,null}, URT)]$ is less than the costs of hiring VRT and obtaining the site-specific information.) To lower the cost of gathering necessary information, the authors searched for an inexpensively obtainable proxy for the characteristic variables obtained from soil sampling. In particular, they studied whether using a pendulum-meter as an indirect sensor of plant mass could

profitably provide sufficient information about nitrogen fertilizer application rates. Soil-based nitrogen was the characteristic variable for which they attempt to use plant mass as a proxy. Knowing the soil-based nitrogen map gives the producer important information about how to more profitably apply nitrogen fertilizer. The authors theorized that in parts of the field with low plant growth there was stress from lack of water, and therefore plant roots could not absorb well the fertilizer applied.

ESV (2004) ran agronomic experiments to examine the response of winter wheat to nitrogen fertilizer in three fields: one field in 2000, and two other fields in 2001. The value of the plant mass index obtained ranged from 18 to 58. They report having used 7 kg N ha^{-1} of calcium-amonium-nitrate (CAN) fertilizer on the parts of the fields with the sparsest plant mass, and they used 68 kg N ha^{-1} of CAN on the parts of the fields with the densest plant mass.[1] Their results were that in two of the three site-years their site-specific management algorithm resulted in increased yields and decreased N fertilizer application than did the conventional application of nitrogen fertilizer at a uniform rate. In another site-year their site-specific management algorithm led to decreased yields and decreased N fertilizer application. Clearly, in the two fields in which yields rose and N application fell, the gross margin of the authors' method exceeded that of the conventional practice: $GM_N^*(\Phi_{post,exp,\mathbf{s}}, VRT) > GM_N^*(\Phi_{post,noexp,null,}, URT)$. Unfortunately, the authors do not attempt to report whether this increase in the gross margin was greater than or less than the costs of hiring VRT and obtaining the site-specific information. Therefore the implications

---

[1] The authors do not state the specific formula used to determine the nitrogen application rate for each plant mass index value. They do mention that it was "stepwise linear adjusted according to measured pendulum angle. The maximum and minimum rates were determined by the agronomist of the farm" (p. 265)." The authors do not offer an explanation for why this formula was used.

of their research for the economic feasibility of variable rate technology remains unclear.

Nor is it clear that that the nitrogen rates given by their algorithm were economically

optimal. Additionally, the authors point out that the years of their study were very dry.

Since their analysis is ex post, the results can only be used to interpret the yield and profit

effects of their methods in unusually dry years. Since no attempts are made to deal with

the weather variables **z**, their results cannot be used to make general inferences about

nitrogen fertilizer management.

Berntsen, et al. (2006)

Berntsen, et al. (2006) ran a total of nine one-year agronomic experiments in nine fields in

Denmark and used the data to estimate nine field-and-year-specific winter wheat response

functions, $f^i(\mathbf{x}, \mathbf{c})$ for i = 1, . . . , 9. Some of the experiments were run in 2001, some in

2002, and the rest in 2003. Nitrogen fertilizer was the managed input they varied in the

experiments; other managed inputs were applied at conventional rates uniformly

throughout the sites. Three functional forms were examined with each being the sum of

three functions: (i) a function of N fertilizer applied (assumed to be either quadratic,

linear-plus-exponential, and Mischerlich; (ii) a quadratic function of one or two of the

characteristics variables; and (iii) a polynomial function of the interactions between N

fertilizer applied and the characteristics variables. After examination of the estimations

using each of the three forms in (i), the authors settled on using the quadratic, implying that

yield response in field *i* took the functional form:

(5) $$f^i(N, c) \equiv \mathbf{a}N^2 + \left(\mathbf{b} + \mathbf{f}c + \mathbf{g}c^2 + \mathbf{h}c^3\right)N + \mathbf{c}_i + \mathbf{d}_i c + \mathbf{e}_i c^2.$$

Coefficients a, b, f, g, and h were constrained to be equal for all fields, and thus are not subscripted by the field index i in (5). Coefficients c_i, d_i, and e_i were estimated for each field. The authors estimated (5) seven times for each field, letting c be each of the following in turn: the ratio vegetation index (RVI), the Yara Biomass, and the soil apparent electrical conductivity (EC_a), which were characteristics measured by sensors; and topographic characteristic variables height, slope, solar insolation index (SII), and aspect. All estimations were conducted using the general linear model (GLM). Spatial regression methods were not used.

Using their field-specific response functions, the authors obtained estimates of site-specific response functions by measuring the value of the characteristics variable c in site s of field i, and inserting it into the field-specific response function:

$$(6) \quad f^{field,s}(N) \equiv f^{field}\left(N, c^{field,s}\right) \equiv aN^2 + \left(b + fc^{field,s} + g\left[c^{field,s}\right]^2 + h\left[c^{field,s}\right]^3\right)N +$$

$$c_i + d_i\left[c^{field,s}\right] + e_i\left[c^{field,s}\right]^2, \text{ for } field = 1,...,9, \text{ and for } s = 1,...,S..$$

Using (6), the authors used OLS to estimate $N^*(\Phi_{post,exp,stan,}, VRT)$, the *ex post* economically optimal site-specific N-rate for each site on each field, as that rate that maximized revenues minus nitrogen fertilizer costs. They also addressed the question of how to maximize yield by redistributing N fertilizer among a field's sites without increasing the total amount of N fertilizer used in the original experiment.

The authors concluded that benefits derived from using their algorithms were not large enough to offset the extra expenses of running the experiments and hiring the variable rate technology. They write, "The . . . benefits resulting from using the algorithms described were too small . . . to warrant their application in practice. The

limited benefits are due to the limited capability of the currently available sensors to predict yield at harvest" (p. 80). In terms of our notation, this can be written

$$\boldsymbol{P}^*(\Phi_{post,exp,\boldsymbol{s}}, \boldsymbol{t}) = R^*(\Phi_{post,exp,\boldsymbol{s}}, VRT) - C_N^*(\Phi_{post,exp,\boldsymbol{s}}, VRT) - C_{VRT,} - C_{\boldsymbol{s}.} <$$

$$\boldsymbol{P}^*(\Phi_{post,noexp,null}, URT) = R^*(\Phi_{post,noexp,null}, URT) - C_N^*(\Phi_{post,noexp,null}, URT) - C_{URT,} - C_{null},$$

which implies $GM_N^*(\Phi_{post,exp,\boldsymbol{s}}, VRT) - GM_N^*(\Phi_{post,noexp,null}, URT) < [C_{VRT,} + C_{\boldsymbol{s}.}] - [C_{URT,} + C_{null}]$.

Hurley, Oishi, and Malzer (2005)

Hurley, Oishi, and Malzer (HOM) (2005) used data from agronomic experiments on corn yield response to N fertilizer conducted on 4.5 ha areas in two fields in Minnesota in 1995. They estimate site-and-year-specific response functions $f^{s,t}(x)$ by imposing a rectangular grid on each field, and delineating the grid cells as sites. They made no attempt to incorporate characteristic or weather variables ($\boldsymbol{c}$ or $\boldsymbol{z}$) in their estimations. Within the 4.5 ha areas were six replications with six treatments each in a randomized complete block design. The treatments ran the length of the 4.5 ha areas, and were randomized between but not within strips. HOM noted that Lambert, Bongiovanni, and Lowenberg-DeBoer (2002), used both a geostatistical model to deal with spatial correlation and a spatial econometric model to deal with spatial correlation and site heteroskedasticity. Lambert, Bongiovanni, and Lowenberg-DeBoer (2002) concluded that the spatial model best represented the data from their agronomic experiment in Argentina. The purpose of the Huley, Oishi, and Malzer (2005) paper was to present a spatial autoregressive error (sar-err) model incorporating site, treatment, and strip dependent heteroskedasticity and

correlation, to estimate site-specific response functions, and then to compare their results to earlier results from application of a geostatistical (geo) model to the same data.

Hurley, Oishi, and Malzer (2005) assumed quadratic functional forms for their yield response functions. They investigated the importance of incorporating site, treatment, and strip effects in their response function estimation by estimating five nested models and comparing them using a likelihood ratio statistic. They concluded that yield response to N fertilizer varied considerably among sites in each field, and that the group-wise heteroskedastic SAR model performed better than the regular SAR model and the GEO model.

Using both the SAR and the GEO models, Hurley, Oishi, and Malzer (2005) estimated site-specific *ex post* optimal variable and uniform application rates, and estimated the *ex post* value of variable rate technology given site-specific information: $\boldsymbol{P}^*(\Phi_{post,exp,\boldsymbol{s}}, VRT) - \boldsymbol{P}^*(\Phi_{post,noexp,null}, VRT)$ for $\boldsymbol{s} =$ SAR, GEO, and the *ex post* value of site-specific information under uniform rate technology: $\boldsymbol{P}^*(\Phi_{post,exp,\boldsymbol{s}}, URT) - \boldsymbol{P}^*(\Phi_{post,noexp,null}, URT)$ for $\boldsymbol{s} =$ SAR, GEO. They also report the total value of variable rate technology and site-specific information combined: $\boldsymbol{P}^*(\Phi_{post,exp,\boldsymbol{s}}, VRT) -- \boldsymbol{P}^*(\Phi_{post,noexp,null}, URT)$. (They assumed that without site-specific information, the producer would apply N at the standard recommended rate.) They concluded that in both of their fields the estimated value of $\boldsymbol{P}^*(\Phi_{post,exp,\boldsymbol{s}}, VRT) - \boldsymbol{P}^*(\Phi_{post,noexp,null}, URT),$ was less than the costs of hiring variable rate technology. Their GEO model implied that in one of their fields, they could not reject at a 10% level of statistical significance the hypothesis that the value of $\boldsymbol{P}^*(\Phi_{post,exp,geo}, URT) - \boldsymbol{P}^*(\Phi_{post,noexp,null}, URT),$ was greater than the cost of obtaining the site-specific information. Their SAR model did not lead to the same conclusion.

Shanahan, et al. (2004)

Shanahan, et al. (2004) conducted agronomic experiments in Colorado in 1997, 1998, and 1999 and estimated eighteen field-and-year-and-hybrid-and-yield-potential-specific yield response functions of corn to plant density. They also pooled their data, then estimated two hybrid-specific response functions of corn yield to plant density. Their regression method was standard OLS, and they assumed quadratic functional forms throughout. In conducting their agronomic experiments, Shanahan et al. (2004) used as treatments a factorial combination of two hybrids (early maturing and late maturing) and four plant densities. To identity management zones, they sorted their data into low- medium- and high-yield areas. They estimated plant density response curves for each management zone and each hybrid, assuming a quadratic functional form.

In addition to estimating corn response to plant density, the authors incorporated in their analysis nine characteristic variables or proxies of characteristics variables, including elevation, slope, soil brightness (red, green and blue bands of images), electrical conductivity (shallow and deep readings), pH, and soil organic matter. They write that because, unlike Bullock and Bullock (1998), they used inexpensively obtainable proxies ("indirect measures") for characteristics variables, that it is important to examine whether variable rate plant density might be economically feasible in the Great Plains. They did not model in their regressions yield as a function of characteristic variables or weather variables, but rather examined the effects that characteristics variables had on yield differences between field-years.

To examine the importance of the impacts of the various characteristics variables on yield variation between field-years, they used step-wise regression analysis, retaining only significant variables in their final prediction models. They concluded that the elevation variable was the most important landscape attribute for explaining spatial variation in yield, and that pH, soil organic matter, and soil brightness were also important. In general, they observed negative correlation between soil organic matter and elevation, and positive correlation between pH and elevation.

In an attempt to judge which inexpensively obtainable proxies could be used for more expensively-obtainable characteristics variables such as pH and soil organic matter, Shanahan, et al. (2004) calculated linear correlation matrices for yield and the characteristics variables and proxies. They found electrical conductivity to be correlated with the important characteristics pH and soil organic matter, and that measuring soil brightness by aerial photography has potential for delineating areas of important characteristics, in particular soil organic matter. They concluded that "general landscape attributes like elevation may provide an indirect means of assessing spatial variation in soil properties that have direct impact on crop productivity" (p. 223).

Shanahan et al. (2004) used their estimated response functions to estimate *ex post* economically optimal plant density rates, and to chart out the implied derived demand curve for the managed input. They found that these rates differed by as much as 5000 plants ha^{-1} between high- and low-yield field areas. They stated that this finding implies a potential savings in seed costs of \$6.25 ha^{-1}, implying that the benefits of variable rate plant density may be greater than the costs. But in their approach they do not attempt to account for weather. Since their experiments are run on a different field every year, then it may be

that their results confound the effects of weather with the effects of "management zone" characteristics, and that their estimates of economically optimal plant densities suffer from omitted variable bias.

Khosa, Inman, and Westfall (2005)

In estimating site-and-year specific response functions of corn to N fertilizer, Khosa, Inman, and Westfall (KIW) (2005) conducted on one field one two-year agronomic experiment, and on another field one one-year agronomic experiment. The experiments were located in northeast Colorado, on irrigated fields in which corn is grown continuously. They delineated sites using a commercial a commercial software program using aerial imagery of bare soil, the farmer's perception of field topography, and the farmer's past sol and crop management practices. They set three N rates: the recommended rate (as determined from an N rate algorithm), approximately half the recommended rate, and a control treatment of 0 kg ha^{-1}. Treatments were replicated once, nested within sites, and randomly allocated to experimental strips running the length of the field. Response curves were estimated with standard OLS, assuming a curvilinear functional form. KIW found that N response functions and yields differed significantly across the "management zone" sites delineated in their study. They conducted no economic analysis with their estimated response functions.

Whelan and Taylor (2005)

Whelan and Taylor (2005) report the results of a 2003 agronomic experiment in South Australia measuring wheat yield to N fertilizer on a 50-ha field they call "Bill's Field," and then in the same field measuring barley yield to N fertilizer in 2004. They also report the

results of N-rate experiments conducted on "Field 44," a 130-ha field in Victoria, Australia. In 2003 the Field 44 experiment measured canola response to N fertilizer, and in 2004 the Field 44 experiment measured wheat response. Whelan and Taylor divided each field into three sites ("management zones.") In each field in each year, they used their data to estimate site-and-year-specific response functions $f^{s,t}(N)$, assuming quadratic functional forms and using standard OLS. They use the estimated response functions to estimate *ex post* site-and-year specific economically optimal N rates $N^{s,t*}(\Phi_{post,exp,stan}, VRT)$, and site-and-year specific optimal gross margins $GM_N^{s,t*}(\Phi_{post,exp,stan}, VRT)$. They compared this to an estimate of the gross margin that would have been earned at that site-year had it been fertilized at the farmer's customary uniform rate $GM_N^{s,t*}(\Phi_{post,noexp,null}, URT)$. They report a difference in gross margin of A\$23.38 ha^{-1} in 2003 in Field 44:

$$\sum_{s=1}^{3}\left[GM_N^{s,2003,Field44*}\left(\Phi_{post,exp,stan},VRT\right) - GM_N^{s,2003,Field44*}\left(\Phi_{post,noexp,null},VRT\right)\right] = A\$12.08\ ha^{-1}.$$

Similarly, in Bill's field in 2004, they report a difference in gross margin of A\$1.94 ha^{-1}:

$$\sum_{s=1}^{3}\left[GM_N^{s,2004,Bill*}\left(\Phi_{post,exp,stan},VRT\right) - GM_N^{s,2004,Bill*}\left(\Phi_{post,noexp,null},VRT\right)\right] = A\$1.94\ ha^{-1}.$$

Whelan and Taylor (2005) conclude that their methods could be used to aid decision-making anywhere: "all fields on all farms can provide the information relevant for individual management… Input response data from individually fields may then e used directly or as a replacement for generic models in crop simulation programs" (p. 871). The authors state explicitly that, "there is potential for a more sophisticated spatial analysis of the N response data" (p. 871).

Lambert, Lowenberg-DeBoer, and Malzer (2006)

Lambert, Lowenberg-DeBoer, and Malzer (LLM) (2006) conducted a five-year agronomic experiment on a 12.2 ha field in Minnesota to analyze corn response to N and P and soybean response to N. Their stated objectives were to test whether corn response to N and P and soybean response to N are spatially and/or temporally stable, and to evaluate ex-post net present value profitability of using variable rate technology over a 5-year corn-soy rotation.

LLM designed their experiment with 3 replications of 13 treatments, in a split-plot arrangement of a randomized block design, where P was the main plot and N treatments were randomized within the P treatments. Every year, they divided the experimental area year into 69 "sub-blocks" (sites). Assuming quadratic functional forms, they estimated site-and-year-specific response functions for all 69 sites in every one of the five years. When spatial correlation among observations was significant (as it was for corn in 1999 and 2001 (p. 46)), they used a geostatistically weighted generalized method of moments procedure to estimate their response functions. Because their yield observations were not evenly dispersed, they claimed that it was not appropriate to use a spatial weighted design based on a lattice contiguity matrix.

By statistically comparing the estimated linear N and P coefficients, LLM tested the spatial and temporal stability of corn and soy crop response to N and P. Their results indicated that spatial variation of crop response to N and P was significant, and that the responses of corn and soybean to P was temporally table in parts of the field but not in all of the field. Crop response to N was not temporally stable.

LLM compared the *ex post* profitabilities of URT and VRT using partial budget analysis. They used net present values of five-year profit streams on the whole field as

measures of profitability. The solved for economically optimal input application rates using standard analytical methods. They included in their calculations estimates of site-specific management and information gathering. They compared URT and VRT management strategies in four *ex post* scenarios. In all scenarios, the uniform rate was obtained by following the University of Minnesota recommended fertilizer rate algorithms.

In the first scenario, LLM estimated the difference between the (1) the profitability of experimentation, analysis, and use of variable rate technology, (the per ha net present value of the five-year stream $\sum_{t=1}^{5}(1+r)^{-t}\Pi^{*}\left(\Phi_{post,exp,s_t},VRT\right)$, where r was a discount rate and s_t was either GEO or STAN, depending on which method was most appropriate in year t), and (2) the profitability of following University of Minnesota recommended rates using URT (the per ha net present value of the five year stream $\sum_{t=1}^{5}(1+r)^{-t}\Pi^{*}\left(\Phi_{post,noexp,null},URT\right)$. LLM estimated that the net present value of the plan using experimentation, analysis, and VRT was \$28.38 ha^{-1} greater than the net present value of following the University of Minnesota recommended rates using URT:

$$\sum_{t=1}^{5}(1+r)^{-t}\left[\Pi_t^{*}\left(\Phi_{post,exp,s_t},VRT\right)-\Pi_t^{*}\left(\Phi_{post,noexp,null},URT\right)\right]=\$28.38 \text{ ha}^{-1}.$$

This estimate was deemed significant at the 1% level using a paired *t* test.

In the second scenario, LLM developed an estimate that is comparable to those of several other published studies, in that they ignored the costs of information gathering, analysis, and the use of the uniform or variable rate technology. That is, they compared the difference between gross margins, and found

$$\sum_{t=1}^{5}(1+r)^{-t}\left[GM_{NP}^{*}\left(\Phi_{post,exp,s_t},VRT\right)-GM_{NP}^{*}\left(\Phi_{post,noexp,null},URT\right)\right]=\$68.47 \text{ ha}^{-1}.$$

This estimate was also deemed significant at the 1% level using a paired *t* test.

In the third scenario, LLM compare the net present value streams derived from (1) conducting experiments, analyzing data, and spreading P at the informed variable rate while following the recommended rate for uniform N application, and (2) following the recommended rates for uniform N and P application. They estimated that the net present value of management plan (1) to be $0.25 ha^{-1} greater than that of management plan (2). However this estimate was not deemed significant at the 1% level using a paired t test.

Finally, in their fourth scenario, LLM compare the net present value streams derived from (3) conducting experiments, analyzing data, and spreading N at the informed variable rate while following the recommended rate for uniform P application, and (2) following the recommended rates for uniform N and P application. They estimated that the net present value of management plan (3) to be $19.81 ha^{-1} less than that of management plan (4). This estimate was deemed significant at the 1% level using a paired t test.

Conclusions

It is clear from our review of recent literature on the profitability of variable rate technology that much has been accomplished in that literature over the past two years. Even recently published studies have claimed that not much applied work combing agronomic experiments and economic theory has been done to study the economic feasibility of variable rate technology. For example, Lambert, Lowenberg-DeBoer, and Malzer (2006) wrote, "Until recently, mulit-year on-farm production data from controlled variable rate technology (VRT) experiments hve been limited" (p. 43). Similarly, Liu,

Swinton, and Millier (2006) stated, "[t]he published research to date has not demonstrated how inter-year moisture variability affects [site-specific] yield crop response to N" (p. 472). Along these same lines, Ruffo, et al. (forthcoming) wrote, "the estimation of site-specific response functions will allow the estimation of the economically optimal NF rate site-specifically at a management scale and detail required for VRN application. Unfortunately, very little research has attempted to explain the causes of this variability or to develop site-specific response functions." As the current literature shows, agricultural economists and agronomists are energetically pursuing the shortcomings in the literature mentioned in the three articles quoted above. A good deal of research has been reported that has used agronomic experimentation to estimate economically optimal variable rate application of managed inputs, and to compare the profitability of seeking information and using variable rate technology to that of following "recommended rate" algorithms and using uniform rate technology.

The empirical results of the research on the profitability of generating information and using variable rate technology are mixed. Some studies find information seeking and VRT to be profitability, and others find just the opposite. There may be three reasons for these inconsistent results: insufficient use of spatial analysis, the need for longer-term data, and the need for ex ante analysis. Significant progress has been made in all three of these research areas over the past few years.

Of the papers reviewed in this article, Lambert, Lowenberg-DeBoer, and Malzer (2006), Liu, Swinton, and Miller (2006), Ruffo et al. (forthcoming), Ruffo, et al. (2006), Anselin, Bongiovanni, and Lowenberg-DeBoer (2004), Hurley, Oishi, and Malzer (2005), use spatial econometrics in the estimation of response functions. While it remains an open

question how much practical difference using spatial econometrics makes, some evidence is beginning to be brought forth. The result of Anselin, Bongiovanni, and Lowenberg-DeBoer that using spatial methods increased profits in their experiment by over $2.00 ha^{-1} is very pertinent.

Several authors have begun to emphasize the growing need for longer-term experimental data (Lambert, Lowenberg-DeBoerr, and Malzer 2006, p. 43; Liu, Swinton, and Miller 2006, p. 472; Ruffo, et al. forthcoming). A principal reason that longer-term, multi-location data is needed is to estimate *ex ante* optimal variable management rates. Except for Liu, Swinton, and Millier (2006) and Ruffo et al. (2006), all current research on the economic feasibility of variable rate technology and its information have been *ex post* in nature. Thus, the inference space provided by research to date is too small. Allmost all studies report results for specific site-years. When longer-term experimental data is generated, and weather variables are included in estimations of yield response functions, it will be possible to begin examining the more pertinent question of whether using variable rate technology and generating the information required are economically feasibility *ex ante.*

****As our discussion in the section on Spatial Regression Alternatives makes clear, there remain many issues within spatial econometrics that still need to be examined in the context of the estimation of yield response functions. …*** ***Emphasize shifting out the supply curve.***

References

Anselin., L, Bongiovanni, R., Lowenberg-DeBoer, J., 2004. A spatial econometric approach to the economics of site-specific nitrogen management in corn production. Amer. J. Agric. Econ. 86, 675-687.

Berntsen, J., Thomsen A., Schelde, K., Hansen, O.M., Knudsen, L., Broge, N., Hougaard, H., Hørfarter, R., 2006. Algorithms for sensor-based redistribution of nitrogen fertilizer in winter wheat. Precision Agric. 7, 65-83.

Bullock, D.S., Bullock, D.G., 2000. From agronomic research to farm management guidelines: a primer on the economics of information and precision agriculture. Precision Agric. 2, 71-101.

Bullock, D.S., Lowenberg-DeBoer, J., Swinton, S., 2002. Adding value to spatially managed inputs by understanding site-specific yield response. Agric. Econ. 27, 233-245.

Bullock, D.S., Bullock, D.G., 2000. From agronomic research to farm management guidelines: a primer on the economics of information and precision agriculture. Prec. Agric. 2, 71-101.

Cressie, N.A.C., 1993. Statistics for Spatial Data. John Wiley & Sons, New York.

Hurley, T.M., Oishi, K., Malzer, G.L., 2005. Estimating the potential value of variable rate nitrogen applications: a comparison of spatial econometric and geostatistical models. J. Agric. and Resource Econ. 30, 231-249.

Kessler, M.C., Lowenberg-DeBoer, J., 1998. Regression analysis of yield monitor data and its use in fine-tuning crop decisions. In: Precision Agriculture: Proceedings of the 4th International Conference, July 19-22, Madison, WI, p. 821-828.

Khosa, R., Inman, D., Westfall, D.G., 2005. Optimum N management using site-specific management zones. In: Precision Agriculture ´05. J.V. Stafford (Ed.), Wageningen, The Netherlands: Wageningen Academic Publishers, pp. 827-834.

Kirk, H.J., Haynes, F.L., Monroe, R.J., 1980. Application of trend analysis to horticultural field trials. J. Amer. Hort. Soc. 105, 189-193.

Lambert, D., Bongiovani, R., Lowenberg-deBoer, J., 2002. Spatial regression, an alternative statistical analysis for landscape scale on-farm trials: case study of soil density trials in central Illinois. 6th International Congress on Precision Agriculture and Other Precision Resource Management. July 15, Minneapolis, MN.

Lambert, D., Lowenberg-DeBoer, J., Malzer, G.L., 2006. Economic analysis of spatial-temporal patterns in corn and soybean response to nitrogen and phosphorous. Agron. J. 98, 43-54.

Lark, R.M., Wheeler, H.C., 2003. A method to investigate within-field variation of the response of combinable crops to an input. Prec. Agric. 95, 1093-1104.

Little, R.C., Milliken, G.A., Stroup, W., Wolfinger, R.D., 1996. SAS System for Mixed Models. SAS Institute, Inc., Cary, NC.

Liu, Y., Swinton, S.M., Miller, N.R., 2006. Is site-specific yield response consistent over time? Does it pay? Amer. J. Agric. Econ. 88, 471-483.

Papadakis, J.S., 1937. Méthode Statistique Pour Des Experiences Sur Champs. Bulletin de l'Institut de l'Amelioration des Plantes, Thessalonique, 23.

Peone, J., Lowenberg-DeBoer, J., Lambert, D.M., Griffin, and T.W., 2004. Precision agriculture profitability review - part 2. Site-Specific Management Center, Purdue University.

Ruffo, M., Bollero, G., Bullock, D.S., Bullock, D.G., forthcoming. Site-specific production functions for variable rate corn nitrogen fertilization. Precision Agriculture.

Ruffo, M., Bullock, D.S., Bullock, D.G., Bollero, G., 2006. Using precision technology to gather information necessary to make precision agriculture profitable: an on-farm demonstration. Working paper, University of Illinois Department of Crop Sciences.

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

Shanahan, J.F., Doerge, T.A., Johnson, J.J., Vigil, M.F., 2004. Feasibility of site-specific management of corn hybrids and plant densities in the Great Plains. Prec. Agric. 5, 207-225.

Tamura, R.N., Nelson, L.A., Naderman, G.C., 1988. An investigation of the validity and usefulness of trend analysis for field plot data. Agron. J. 80, 712-718.

Whelan, B.M., Taylor, J.A., 2005. Local response to nitrogen inputs: advancing SSCM within Australia. In: Precision Agriculture ´05. J.V. Stafford (Ed.), Wageningen, The Netherlands: Wageningen Academic Publishers, pp. 827-834.