Estimating Site-Specific Nitrogen Crop Response Functions: A Conceptual Framework and Geostatistical Model

Authors

Terrance M. Hurley
thurley@dept.agecon.umn.edu
Dept. of Applied Economics
Univ. of Minnesota
1994 Buford Ave.
St. Paul, MN 55108-6040

Gary L. Malzer
Dept. of Soil, Water, and Climate
Univ. of Minnesota
1991 Upper Bufird Circle
St. Paul, MN 55108-6040

Bernard Kilian
University Hohenheim, Germany

Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Montreal, Canada, July 27-30, 2003

Introduction

The precision agriculture (PA) hypothesis asserts that varying management activities between or within fields can benefit farmers or the environment. A necessary condition for PA is that the productivity of management activities must vary between or within fields due to factors typically not managed by farmers. Validating this necessary condition is a challenge due to the inherent difficulties of collecting, analyzing, and interpreting appropriate data.

An approach that has emerged to test the PA hypothesis for variable rate nitrogen applications is the estimation and comparison of site-specific crop response functions (SSCRFs) using multiple regression analysis (e.g. Davis et al., 1996; Malzer et al., 1996; Bongiovanni and Lowenberg-DeBoer, 2000 and 2001; Lambert et al., 2002; Hurley et al., 2001 and 2002a,b). The earliest applications of this approach rely on ordinary least squares (OLS), which does not account for heteroscedastic or autocorrelated errors. While OLS estimates may remain unbiased even with heteroscedasticity and autocorrelation, they are typically not efficient and can convey a false sense of precision (Schabenbeger and Pierce, 2001). After confirming the presence of spatial autocorrelation, recent applications use more sophisticated statistical models to address this problem. Still, the conceptual foundations used to justify these models are seldom explicit, making it difficult to judge the merit of the method.

The purpose of this paper is to provide a conceptual framework that illuminates how SSCRFs can be used to test the PA hypothesis. The framework is useful because it identifies an appropriate hypothesis and explains recent evidence of site and treatment dependent heteroscedasticity and spatial autocorrelation in regression errors (Hernandez and Mulla, 2002; Hurley et al., 2002a; and Lambert et al., 2002). The conceptual framework is used to guide the development of a heteroscedastic, fixed and random effects, geostatistical model for estimating

SSCRFs and testing the PA hypothesis using data from a randomized complete block design field experiment. The model is applied to 1995 field data from South Central Minnesota.

Conceptual Framework

The precision agriculture (PA) hypothesis asserts that farmers or the environment can benefit from varying management within or between fields. To better understand this hypothesis from a farmer's perspective (analogous arguments can be made from an environmental perspective), suppose crop yield y (e.g. corn kg ha^{-1}) depends on two types of inputs. The first, denoted by x, are variable inputs or inputs under the immediate control of a farmer (e.g. nitrogen kg ha^{-1}). The second, denoted by z, are fixed inputs or inputs not under the immediate control of a farmer (e.g. soil type, rainfall, and topography). In general, the relationship between yield, variable inputs, and fixed inputs can be described as $y = f(x, z)$, which is assumed continuously differentiable in x and z. For expositional convenience, y, x, and z are treated as scalars.

If a farmer's objective is to optimize the net return from a variable input, the classic rule from economic theory says to choose x^* such that $p_y \frac{\partial f(x^*, z)}{\partial x} = p_x$ where p_y is the price received per unit of crop yield and p_x is the price paid per unit of variable input.[1] In economic parlance, the rule states that the value of marginal product for an input should equal its marginal cost. The optimal amount of variable input depends on the yield price, variable input price, and most importantly for PA, the amount of fixed input. How the optimal amount of variable input depends on the amount of fixed input is found using the implicit function theorem: $\frac{\partial x^*}{\partial z} = -\frac{\partial^2 f(x^*, z)}{\partial x \partial z} \frac{\partial x^2}{\partial^2 f(x^*, z)}$. Note that the optimal amount of variable input does not change with

[1] There are a few caveats to this rule that can be found in any text treating microeconomic theory.

the amount of fixed input if $\frac{\partial^2 f(x^*,z)}{\partial x \partial z} = 0$, which means there is no interaction between variable and fixed inputs. For example, if the amount of soil organic matter does not influence how the crop responds to nitrogen applications, there is no value to varying nitrogen applications within a field in response to variation in soil organic matter.

Observational and experimental field data provide an opportunity to test the PA hypothesis, but the development of appropriate statistical models has proven challenging. To understand why, consider a set of data (y_i, x_i, z_i, e_i) for $i = 1, 2, \ldots, N$. An individual data point consists of y_i, an observed yield; x_i, an observed variable input; z_i, an observed or unobserved fixed input; and e_i, an unobserved error due to for example imperfect measurement or approximation. Using a Taylor series expansion,

$$(1) \qquad f(x,z) = \beta_{00} + \sum_{k_z=1}^{\infty} \beta_{0k_z} z^{k_z} + \sum_{k_x=1}^{\infty} \left(\beta_{k_x 0} + \sum_{k_z=1}^{\infty} \beta_{k_x k_z} z^{k_z} \right) x^{k_x} \quad \text{and}$$

$$(2) \qquad \frac{\partial^2 f(x,z)}{\partial x \partial z} = \sum_{k_x=1}^{\infty} \sum_{k_z=1}^{\infty} \beta_{k_x k_z} k_x k_z z^{k_z - 1} x^{k_x - 1}$$

where $\beta_{00} = f(0,0)$, $\beta_{0k_z} = \frac{\partial^{k_z} f(0,0)}{k_z! \partial z^{k_z}}$, $\beta_{k_x 0} = \frac{\partial^{k_x} f(0,0)}{k_x! \partial x^{k_x}}$, and $\beta_{k_x k_z} = \frac{\partial^{k_x + k_z} f(0,0)}{k_x! k_z! \partial x^{k_x} \partial z^{k_z}}$ for all k_x and k_z are real constants that indicate how variable and fixed inputs combine to influence crop yield. Equation (2) suggests the null hypothesis $\beta_{k_x k_z} = 0$ for all $k_x > 0$ and $k_z > 0$, which implies PA cannot be used to the benefit of a farmer or the environment because there is no interaction between the variable and fixed inputs. To test the null hypothesis, the constant coefficients in equation (1) must be estimated, a task that is generally not feasible.

The first obstacle is the dimensionality of the problem. Since the true relationship ($y = f(x, z)$) is seldom (if ever) known, some approximation is necessary. In addition to the need for

approximation, there is the potential for measurement error. However, neither of these two problems is idiosyncratic to PA. Truncating the Taylor series expansion while adding an error is a solution commonly employed (either explicitly or implicitly). The result is

$$(3) \qquad y_i = \beta_{00} + \sum_{k_z=1}^{K_z} \beta_{0k_z} z_i^{k_z} + \sum_{k_x=1}^{K_x} \left[\beta_{k_x 0} + \sum_{k_z=1}^{K_z} \beta_{k_x k_z} z_i^{k_z} \right] x_i^{k_x} + e_i$$

where K_x and K_z are integers. Equation (3) can be estimated using a variety of regression techniques. For example, if it is reasonable to assume e_i has zero mean and is independently and identically distributed for $i = 1, 2, \ldots, N$, ordinary least squares is appropriate. If heteroscedasticity or autocorrelation is of concern, then feasible generalized least squares or maximum likelihood incorporating heteroscedasticity and autocorrelation is appropriate. Depending on the estimation method, the null hypothesis can be tested using an F or likelihood ratio statistic.

Another obstacle more specific to PA is that z_i is often unobserved. A researcher or farmer may suspect some fixed inputs interact with the variable input, but not know precisely what fixed inputs are important. Confirming the PA hypothesis without knowledge of important fixed inputs is useful because it indicates whether searching for these fixed inputs is worth an effort. If the PA hypothesis cannot be confirmed generally or the value of discovering which fixed inputs are important is small, it may make sense to devote research effort elsewhere.

When z_i is not observed, it can be treated as another source of regression error, so equation (3) becomes

$$(4) \qquad y_i = \alpha_0 + \sum_{k_x=1}^{K_x} \alpha_{k_x} x_i^{k_x} + \xi_i$$

where $\alpha_0 = \beta_{00} + \sum_{k_z=1}^{K_z} \beta_{0k_z} Z_{k_z}$, $\alpha_{k_x} = \beta_{k_x 0} + \sum_{k_z=1}^{K_z} \beta_{k_x k_z} Z_{k_z}$, and Z_{k_z} are constants and

$\xi_i = \sum_{k_z=1}^{K_z} \left(\beta_{0k_z} + \sum_{k_x=1}^{K_x} \beta_{k_x k_z} x_i^{k_x} \right) \left(z_i^{k_z} - Z_{k_z} \right) + e_i$ is the regression error. There are two sources of error in equation (4). The first is attributable to unobserved fixed inputs. The second is attributable to approximation and measurement error. Under the traditional assumption that the expected value of the regression error is zero, Z_{k_z} reflects the sample average of $z_i^{k_z}$. This means that the estimation of the α parameters in equation (4) reflects the base yield and crop response to the variable input given the average of fixed inputs. The α parameters do not reflect how crop response varies within or between fields because the coefficients for testing the PA hypothesis ($\beta_{k_x k_z}$ for $k_x > 0$ and $k_z > 0$) are embedded within the α parameters, as well as the error.

Conceptually, it is possible to test the null hypothesis by estimating equation (4) and exploring the sensitivity of the regression error to the variable input. Note that under the alternative hypothesis the regression error depends on the variable input, so testing for this dependence presumably tests the null hypothesis. A practical problem with this strategy is the inherent difficulty of obtaining unbiased and efficient estimates of an unknown covariance structure. Another problem is that it assumes the approximation and measurement error is not sensitive to the variable input. Finally, if the PA hypothesis is not rejected, the results provide little insight into to the optimal amount of variable input.

The potential difficulty of testing the null hypothesis using equation (4) highlights the utility of estimating site-specific crop response functions (SSCRFs). Suppose the data is partitioned into *R* mutually exclusive subsets. The subsets can represent spatially adjacent

observations (e.g. individual fields or management zones within a field), though adjacency is not necessary. Separate parameters can be estimated for each partition by rewriting equation (4) as

$$(5) \qquad y_i = \alpha_{0r_i} + \sum_{k_x=1}^{K_x} \alpha_{k_x r_i} x_i^{k_x} + \xi_i$$

where $\alpha_{0r_i} = \beta_{00} + \sum_{k_z=1}^{K_z} \beta_{0k_z} Z_{k_z r_i}$, $\alpha_{k_x r_i} = \beta_{k_x 0} + \sum_{k_z=1}^{K_z} \beta_{k_x k_z} Z_{k_z r_i}$, and $Z_{k_z r_i}$ are constants;

$\xi_i = \sum_{k_z=1}^{K_z} \left(\beta_{0k_z} + \sum_{k_x=1}^{K_x} \beta_{k_x k_z} x_i^{k_x} \right) \left(z_i^{k_z} - Z_{k_z r_i} \right) + e_i$ is the regression error; and the subscript $r_i \in \{1,\ldots,R\}$ indexes the partition assigned to the *i*th observation. Under the assumption that the expected value of the regression error is zero, $Z_{k_z r_i}$ is the sample average of $z_i^{k_z}$ conditioned on $i \in r_i$. If the null hypothesis ($\beta_{k_x k_z} = 0$ for all $k_x > 0$ and $k_z > 0$) is correct, $\alpha_{k_x r_i} = \alpha_{k_x r_j}$ for all $k_x > 0$, r_i, and r_j. Therefore, SSCRFs allow the PA hypothesis to be tested based on the α or mean parameters— parameters for which efficient and unbiased estimates can usually be obtained even in the presence of heteroscedastic and autocorrelated errors.

Estimating equation (5) and testing the equality of $\alpha_{k_x r_i}$ for all $k_x > 0$ and r_i tests the PA hypothesis. What is important to realize is that this test does not imply the equality of α_{0r_i} for all r_i. When there is no interaction between variable and fixed inputs, check plot yields (with no variable input) can vary across sites, even though crop response to the variable input does not.

Estimating equation (5) to test the precision agriculture hypothesis is not trivial. In general, the covariance of ξ_i and ξ_j is

$$(6) \qquad \mathrm{E}(\xi_i \xi_j) = \sum_{k_z=1}^{K_z} \sum_{k_z'=1}^{K_z} \left(\beta_{0k_z} + \sum_{k_x=1}^{K_x} \beta_{k_x k_z} x_i^{k_x} \right) \left(\beta_{0k_z'} + \sum_{k_x=1}^{K_x} \beta_{k_x k_z'} x_j^{k_x} \right) \mathrm{E}\left[\left(z_i^{k_z} - Z_{k_z r_i} \right) \left(z_j^{k_z'} - Z_{k_z' r_j} \right) \right]$$

$$+ \sum_{k_z'=1}^{K_z}\left(\beta_{0k_z'} + \sum_{k_x=1}^{K_x}\beta_{k_xk_z'}x_j^{k_x}\right)\mathrm{E}\left[\left(z_j^{k_z'} - Z_{k_z'r_j}\right)e_i\right]$$

$$+ \sum_{k_z=1}^{K_z}\left(\beta_{0k_z} + \sum_{k_x=1}^{K_x}\beta_{k_xk_z}x_i^{k_x}\right)\mathrm{E}\left[\left(z_i^{k_z} - Z_{k_zr_i}\right)e_j\right] + \mathrm{E}[e_ie_j].$$

Equation (6) provides an explanation for three phenomena reported in the literature. The first and most common is spatial autocorrelation. If fixed inputs such as soil organic matter, soil type, and drainage are spatially correlated, $z_i^{k_z} - Z_{k_zr_i}$ and $z_j^{k_z'} - Z_{k_z'r_j}$ will be spatially correlated. Hernandez and Mulla (2002) report that semi-variogram estimates of regression errors vary by treatment, a result explained by the dependence of equation (6) on the variable input, $x_i^{k_x}$ and $x_j^{k_x}$, when the PA hypothesis is true. Hurley et al. (2002a) and Lambert et al. (2002) report group-wise or site-specific heteroscedasticity in the regression error, a result consistent with the dependence of equation (6) on the conditional average of fixed inputs, $Z_{k_zr_i}$ and $Z_{k_zr_j}$. Each of these phenomena implies crop response estimates from equation (5) using OLS will be inefficient. It also implies the precision of these estimates may be overstated.

A variety of methods have been proposed to deal with the estimation problems posed by these three phenomena. Spatial econometric and geostatistical models have been estimated to address problems arising from spatial autocorrelation. Hernandez and Mulla (2002) estimate treatment specific semi-variograms to deal with treatment dependent spatial correlation. Hurley et al. (2002a) and Lambert et al. (2002) incorporate site dependent heteroscedasticity in OLS, spatial econometric, and geostatistical models. Hurley et al. (2002a) also tries to define the partitions small enough to make $z_i^{k_z} = Z_{k_zr_i}$ for all r_i and k_z so the first three terms in equation (6)

disappear. None of these models or any others we are aware of successfully address both site and treatment dependent spatial autocorrelation.

The practical relevance of these problems is now explored using data from a randomized complete block design field experiment. The experiment was constructed to test within field variation in corn response to nitrogen. After discussing the experimental design, a new statistical model is specified using insights gleaned from this design and the conceptual framework.

Methods

Experimental

The data was collected in 1995 from two production fields near Hanska and Morgan (Brown and Redwood Co. in South Central Minnesota). These sites are located on a higher elevation of glacial till lowland plain that comprises the majority of the Counties. Most soils at these locations belong to the Clarion-Nicollet-Webster association or similar soil series/ associations. The area is nearly level to gently sloping, and the soils range from poorly to moderately well drained. All soils were mollisols, ranging from fine-loamy, mixed, mesic Typic Haplaquolls (the Webster clay loam) to fine-loamy, mixed, mesic typic Hapludolls (the Clarion loam). The climate is interior continental with cold winters and moderately hot summers with occasional cool periods. Total annual precipitation ranges from 711 to 635 mm, which is normally adequate for corn, since 80 percent falls during the growing season. The 1994 crop was soybean and no manure applications had occurred in the last five growing seasons.

Each location was 164 m wide and 274 m in length (4.5 ha). Within this area six replications of six treatments were established in a randomized complete block design. The six replications ran the length of the field. The six randomized treatments within a replication also ran the length of the field in 36 4.6 m wide treatment strips that included nitrogen rates of 0, 67,

101, 134, 168, and 202 kg ha^{-1} applied as anhydrous ammonia. Treatments were applied on November 4, 1994 using a radar controlled variable rate applicator to compensate for variations in applicator speed and to ensure a constant application rate within each treatment strip.

Corn (cv. Pioneer 3531) was planted during the first week of May in 0.76 m rows at approximately 76,500 seeds ha^{-1}. Grain yield was determined by harvesting the center two rows (six row strips) with a Massey Ferguson® plot combine equipped with a ground distance monitor and a computerized Harvestmaster® weigh cell. Experimental strips were divided into 17 harvest segments each 15.4 m in length. Approximately 8 m was discarded from the end of each strip to eliminate border effects. There were 612 yield observations per field. Sub-samples of grain were collected from each area of yield measurement to determine moisture and adjust yields to reflect 15.5 percent moisture. Dikici (2000) reports additional details regarding the experimental design and provides a full summary and descriptive analysis of the data.

Empirical

Estimating equation (5) with this data provides an opportunity to test the precision agriculture (PA) hypothesis. One feature of the data is that it provides observations for each of the six treatments in 102 15.4 m × 27.6 m regions at each location. Therefore, each location can be partitioned into as many as $R = 102$ sites for estimating equation (5) with a full complement of treatments. Another feature is that treatments were randomly assigned across, but not within, strips. Lack of randomization within treatment strips can introduce additional error correlation.

The conceptual framework and experimental design suggest that estimation of equation (5) using ordinary least squares (OLS) will not be efficient. OLS estimates for the standard errors of the α parameter will also be inaccurate. They can be either upward or downward biased (Greene, 1990). The conceptual framework indicates that the covariance of regression

errors will exhibit site and treatment spatial dependencies. The lack of randomization within treatment strips suggests the covariance of regression errors may also exhibit treatment strip correlation. Therefore, estimates of equation (5) should incorporate an error structure that permits treatment strip as well as site and treatment spatial dependencies.

The proposed model is based on the geostatistical framework. First, let $K_x = 2$, so equation (5) becomes

$$y_i = \alpha_{0r_i} + \alpha_{1r_i} x_i + \alpha_{2r_i} x_i^2 + \xi_i \,. \tag{7}$$

The covariance of ξ_i and ξ_j is then defined as

$$\mathrm{E}\left[\left(\xi_i \xi_j\right)\right] = \begin{cases} \sigma_{r_i s_i}^2, & \text{for } i = j \\ \sigma_{r_i s_i} \sigma_{r_j s_j} \left[C_1\left(1 - g_1\left(h_{ij}, s_{ij}\right)\right) + C_s\left(1 - g_s\left(s_{ij}\right)\right) + C_x\left(1 - g_x\left(x_{ij}\right)\right)\right], & \text{for } i \neq j \end{cases} \tag{8}$$

where $\sigma_{r_i s_i}^2 > 0$ and $\sigma_{r_j s_j}^2 > 0$ are the site and strip specific variances for observation i and j; h_{ij} is the segment-wise distance in meters, s_{ij} is the strip-wise distance in meters, and x_{ij} is the difference in variable input between i and j; $1 \geq C_1 \geq 0$, $1 \geq C_s \geq 0$, and $1 \geq C_x \geq 0$ are semi-variogram sill parameters for spatial, treatment strip, and treatment correlation where $1 \geq C_1 + C_s + C_x \geq 0$; and $g_1(h_{ij},s_{ij})$, $g_s(s_{ij})$, and $g_x(x_{ij})$ are semi-variogram distance functions.

Equation (8) exhibits spatially correlated errors when $g_1(h_{ij},s_{ij})$ is specified using a permissible semi-variogram distance function (e.g. spherical, exponential, or Guassian). Equation (8) exhibits errors that are correlated within treatment strips when $g_s\left(s_{ij}\right) = \begin{cases} 0, & \text{for } s_{ij} = 0 \\ 1, & \text{otherwise} \end{cases}$. Treatment correlation is captured when $g_x\left(x_{ij}\right) = \begin{cases} 0, & \text{for } x_{ij} = 0 \\ 1, & \text{otherwise} \end{cases}$. Since the treatment is constant within a strip, conditioning the variance on the site and strip results in an error structure with site, treatment, and treatment strip spatial dependencies. The model can be classified as a heteroscedastic, fixed and random effects, geostatistical model.

Estimation

Equations (7) and (8) can be estimated using a variety of methods after choosing the partition of sites and a distance function for spatial autocorrelation (Schabenberger and Pierce, 2002). The method employed here uses the feasible generalized least squares (FGLS) for the α parameters. Estimates of the covariance parameters are obtained using maximum likelihood (ML) with the α parameters profiled using the FGLS estimator.

The data can be partitioned into 102 different sites with the full complement of treatments, but there are not enough observations in each partition to identify a covariance matrix conditioned on the site, treatment, and treatment strip. Therefore, larger partitions are necessary given the data at hand. To illustrate the benefits of estimating equations (7) and (8) at different levels of resolution, two partitions are explored. The first partitions each location into six contiguous regions by pooling observations from replications 1-3 and 4-6 and segments 1-5, 7-11, and 12-17. The second partitions each location into 48 contiguous regions by pooling observations in two (or three, in one instance) adjacent segments and one replication. With these partitions, site and treatment strip dependent variances can be identified by defining $\sigma_{r_i s_i}{}^2 = \sigma_{r_i}{}^2 \sigma_{s_i}{}^2$ and setting $\sigma_{s_i}{}^2 = 1$ for one strip in replication 1-3 and 4-6 in the first partition or for one strip in each replication in the second partition. Since treatments do not vary within a strip, this also accounts for differences in the variance across treatments.

There are a wide variety of distance functions that can be estimated and compared. However, the model is computationally intensive, which restricts the practicality of comparing a battery of distance functions. Since the primary purpose of the paper is to explore the value of incorporating site, treatment, and treatment strip dependencies into a model with spatially correlated errors, attention is focused on a single distance function. Comparing the fit of a

standard geostatistical model at both locations based on the maximized log-likelihood using the exponential, Gaussian, and spherical distance functions suggests the Gaussian model fit best. Therefore, the full model with site, treatment, and treatment strip spatial dependencies is estimated with the Gaussian function:

$$(9) \qquad g_1\left(h_{ij}, s_{ij}\right) = \begin{cases} 1 - e^{-\left(\frac{\sqrt{h_{ij}^{\,2} + s_{ij}^{\,2}}}{C_R}\right)^2}, & \text{for } \sqrt{h_{ij}^{\,2} + s_{ij}^{\,2}} > 0 \\ 0, & \text{for } \sqrt{h_{ij}^{\,2} + s_{ij}^{\,2}} = 0 \end{cases}$$

where C_R is a shape parameter. The larger the shape parameter the faster the semi-variance approaches the asymptotic sill as the distance between observations increases.

Hypotheses

Eight models are estimated for each location to test a variety of hypotheses. Model 1 uses the partition with six regions and assumes $\sigma_{r_i s_i}^{\,2} = \sigma^2$ for all r_i and s_i, $C_1 = 0$, $C_s = 0$, and $C_x = 0$, which is the ML analogy to OLS. Model 2 uses the partition with six regions and assumes $\sigma_{r_i s_i}^{\,2} = \sigma^2$ for all r_i and s_i, $C_s = 0$, and $C_x = 0$, which is a standard geostatistical model. Model 3 uses the partition with six regions and assumes $\sigma_{r_i s_i}^{\,2} = \sigma_{r_i x_i}^{\,2}$ for all r_i and s_i, and $C_s = 0$. This model incorporates the site and treatment spatial dependencies identified by the conceptual framework as important. Model 4 uses the partition with six regions and is otherwise unrestricted, which adds treatment strip dependencies to model 3. Model 5 is similar to model 4 except it assumes nitrogen does not interact with fixed inputs: $\alpha_{k_x r_i} = \alpha_{k_x r_j}$ for all $k_x > 0$, r_i, and r_j. Models 6, 7, and 8 use the partition with 48 regions, but are otherwise identical to models 1, 4, and 5.

Hypothesis testing can be accomplished using the likelihood ratio statistic (LRS) since the models are nested. The LRS is twice the difference in the maximized likelihood of the unrestricted and restricted model. The statistic is asymptotically distributed χ^2 with the degrees of freedom equal to the number of parameter restrictions.

First, the benefit of incorporating a richer error structure is evaluated by comparing model 1 and 2, 2 and 3, 3 and 4, and 6 and 7. The comparison of models 1 and 2 evaluates the importance of spatial correlation. The comparison of models 2 and 3 evaluates the importance of conditioning the variance and spatial correlation on the site and treatment. The comparison of models 3 and 4 evaluates the importance of also conditioning on treatment strips. These three comparisons are all based on partitioning a field into 6 regions. The comparison of models 6 and 7, evaluates the importance of incorporating site, treatment, and treatment strip spatial autocorrelation when a location is evaluated with 48 rather than 6 regions.

Second, comparing models 4 and 5, and 7 and 8 tests the PA hypothesis. The comparison between models 4 and 5 evaluates whether there are significant differences in crop response to nitrogen between the six regions in the first partition. The comparison between models 7 and 8 evaluates whether there are significant differences in crop response to nitrogen between the 48 regions in the second partition.

Finally, comparing models 4 and 7 evaluates variation in crop response functions within the six regions in the first partition. The test determines if dividing a field into smaller regions for a finer degree of resolution significantly improves the explanatory power of the model.

Results

Hypotheses Tests

Table 1 reports the maximized log-likelihood for each model and the likelihood ratio statistic and degrees of freedom for each model comparison. The regression errors from the crop response function estimates exhibit significant site, treatment, and treatment strip spatial autocorrelation. Model 1 is rejected in favor of model 2 at both locations confirming spatial autocorrelation. Model 2 is rejected in favor of model 3 supporting the implications of the conceptual model—spatial autocorrelation is dependent on the site and treatment. Model 3 is rejected in favor of model 4, which indicates significant correlation within treatment strips. Model 1 is rejected in favor of model 4 and model 6 is rejected in favor of model 7, which shows that dividing fields into smaller regions does not change the importance of employing a richer error structure.

There is significant within field variation in crop response to nitrogen, so there is the potential for variable rate nitrogen applications (VRA) to improve nitrogen returns. Model 5 is rejected in favor of model 4 indicating that crop response to nitrogen varied significantly between the 6 regions in the first partition at both locations. Model 8 is rejected in favor of model 7, indicating that crop response to nitrogen varied significantly between the 48 regions in the second partition at both locations. Model 4 is rejected in favor of model 7, which means crop response functions differ significantly within the 6 regions of the first partition at both locations.

Error Structure

Dividing equation (8) by $\sigma_{r_i s_i}\sigma_{r_j s_j}$ normalizes the semi-variance such that C_1, C_x, and C_s are the proportion of the sill explained by spatial, treatment, and treatment strip effects. The

nugget is $1 - C_1 - C_s - C_x$. Table 2 reports this decomposition along with the shape parameter (C_R) and the average standard deviation for selected models.

The spatial and treatment strip effects are substantial, but the treatment effect is not. Spatial correlation is reduced but not eliminated by estimating crop response functions at a finer degree of resolution (48 vs. 6 regions), implying a finer resolution allows more within field variation in unmanaged inputs to be captured by crop response function estimates. The average standard deviation of error is also reduced when crop response functions are estimated at a finer resolution. Spatial correlation explains between 51 and 63%, while treatment strip correlation explains between 15 and 18% of the semi-variance sill depending on the model and location. The magnitude of the spatial and treatment strip effect is similar for Hanska and Morgan in models 4 and 7. Comparing model 4 and 7 shows, the proportion of the semi-variance sill explained spatially and the shape parameter are lower using 48 rather than 6 regions. Both these factors imply correlation diminishes faster with distance, when 48 regions are used instead of 6.

Variation In Corn Response To Nitrogen

The hypotheses tests indicate that there was potential to increase nitrogen returns by varying applications (VRA) within each of the two fields in 1995. The return to nitrogen above fertilizer costs at each location is calculated as $\pi = \left(\sum_{i=1}^{612} \frac{p_y \left(\alpha_{0r_i} + \alpha_{1r_i} x_i + \alpha_{2r_i} x_i^{\,2} \right) - p_x x_i}{612} \right)$. The optimal VRA is calculated by choosing x_i for $i = 1,..,612$ to maximize π. Alternatively, an optimal uniform rate can be calculated by choosing $x = x_i$ for $i = 1,..,612$ to maximize π. These optimal rates are constrained between 0 and 202 kg ha^{-1} to avoid predicting yields outside the range of the available data.

The potential benefit of VRA can be evaluated by calculating the increased return from switching to the optimal variable rate from the uniform rate recommended by the University of Minnesota's Extension Service (140 kg ha^{-1} for both Hanska and Morgan). To better understand this increased return, it is useful to decompose it into the increase from switching to the optimal uniform rate from the Extension rate and the increase due to switching to the optimal variable rate from the optimal uniform rate. Figure 1 reports this decomposition for Models 1, 4, 6, and 7 where the price of corn and nitrogen equal \$98.21 t^{-1} and \$0.374 kg^{-1}, standard deviation is calculated using a Taylor series expansion (Caselle and Taylor, 1990), and 90 percent confidence intervals are calculated assuming estimated returns are normally distributed.

Three important conclusions can be drawn from Figure 1. First, failing to account for site, treatment, and treatment strip heteroscedasticity and spatial autocorrelation results in a false sense of precision. For both Hanska and Morgan, the confidence intervals for Model 1 are smaller than for Model 4. Similarly, the confidence intervals for Model 6 are smaller than for Model 7. While it is tempting to conclude that the results for Model 1 and 6 are more precise than for Model 4 and 7, the results in Table 1 implies this is not the case. The confidence intervals for Models 1 and 6 are wrong because the assumptions of uncorrelated and homoscedastic errors is violated. Using these inaccurate estimates and confidence intervals can result in erroneous conclusions regarding the profitability of VRA.

There are two distinct ways in which farmers can benefit from PA. One is by determining the optimal uniform rate for a field. The other is by determining how to vary the rate optimally within a field. For Hanska, the return from switching to the optimal uniform rate from the Extension rate is small relative to the return from switching to the optimal variable rate from the optimal uniform rate. The opposite result holds for Morgan. For Hanska, the Extension

rate was close to the optimal uniform rate, so most of the benefit from PA is obtained by varying the rate optimally within the field. For Morgan, the extension rate was too low on average, so most of the benefit from PA could be obtained from identifying a better uniform rate. Farmers do not always have to vary application rates within fields to enjoy the benefits of PA.

There are costs and benefits to estimating crop response functions with a finer degree of resolution. As more crop response functions are estimated for smaller regions of a field, there are fewer observations available to estimate each parameter (a loss in degrees of freedom). This has the effect of increasing the standard error of parameter estimates. Alternatively, by estimating more crop response functions for smaller regions of a field, parameter estimates capture more of spatial variability within the field. This has the effect of reducing the standard error of the parameter estimates. These confounding effects mean that the confidence intervals for Model 4 may be smaller (as seen for Morgan) or larger (as seen for Hanska) than the confidence intervals for model 7. Alternatively, since estimating more crop response functions allows more of the spatial variability to be captured directly with the α parameters, estimates of the potential return to VRA are always larger. At Hanska, estimating more crop response functions more than doubles the estimated potential of VRA, but also decreases confidence in the estimate. At Morgan, estimating more crop response functions again more than double the estimated potential of VRA and also increase the confidence in the estimate.

The conclusion from Table 1 is that VRA can improve the economic returns to nitrogen. This result apparently contradicts the estimate of the increase in returns for Model 4 at Hanska reported in Figure 1. The confidence interval for this estimate includes zero, which suggests VRA doe not significantly increase returns to nitrogen. This contradiction highlights an important distinction. The hypotheses tests in Table 1 evaluate the potential for an increased

return while the hypotheses tests in Figure 1 evaluate the magnitude of the potential increase in the return. While there may be enough information from the experiment to say there is significant potential for VRA, there may not be enough information to precisely estimate the magnitude of this potential.

Figures 2 and 3 report spatially more detailed results for the best fitting model (Model 7) and comparisons to the Extension rate. The figures highlight the degree of within field variability at both locations. For Hanska and Morgan, estimated check strip yields ranged from 2.7.6 to 8.2 t ha^{-1} and 3.6 to 9.8 t ha^{-1} with an average of 6.2 and 6.3 t ha^{-1}. The optimal nitrogen rates ranged from 97 to 202 kg ha^{-1} at Hanska with an average of 154 kg ha^{-1}. These rates correspond to yields ranging from 6.5 to 11.2 t ha^{-1} with an average of 9.4 t ha^{-1}. At Morgan, the optimal application rates ranged from 109 to 202 kg ha^{-1} with an average of 184 kg ha^{-1}. Corresponding yields ranged from 8.3 to 12.9 t ha^{-1} with an average of 10.7 t ha^{-1}. The increase in return when compared to the Extension rate ranged from 0.0 to 176.1 \$ ha^{-1} at Hanska and 0.0 to 274.2 \$ ha^{-1} at Morgan. The standard deviation of this increased return ranged from 0.3 to 76.3 \$ ha^{-1} at Hanska and 0.2 to 78.5 \$ ha^{-1} at Morgan.

Determining the profitability of VRA requires consideration of the cost of implementing a VRA strategy as well as the potential for VRA to increase returns. These costs include for example, the cost of collecting and processing the necessary information, and VRA equipment or application services. For Hanska, if these cost were less than 14.5 \$ ha^{-1} there was greater than a 95 percent chance that VRA could have increased profitability for the whole field. For Morgan, costs less than 48.3 \$ ha^{-1} would have resulted in greater than a 95 percent chance of increased profitability for the whole field. For a cost of 10 \$ ha^{-1}, 8.8 and 37.3 percent of the field at

Hanska and Morgan show a significant increase in profitability. For a cost of 25 \$ ha^{-1}, 6.9 and 32.4 percent of the field at Hanska and Morgan show a significant increase in profitability.

Conclusions

Confirming the precision agriculture hypothesis for variable rate nitrogen applications (VRA) has proven challenging. To confront this challenge, researchers are using increasingly sophisticated statistical models to estimate and compare site-specific crop response functions. While progress has been made, we believe it has been hampered by the lack of a clear conceptual framework to guide and motivate the development of appropriate models and methods. The purpose of this paper was to provide such a framework, while demonstrating its utility. The framework was used to identify a testable hypothesis and develop a statistical model to evaluate that hypothesis. The model was then implemented with data collected in 1995 from two fields in South Central Minnesota.

The conceptual framework and experimental design indicate that regression errors for site-specific crop response functions will likely exhibit site, treatment, and treatment strip spatial autocorrelation. Failing to account for this correlation can result in parameter estimates that are inefficient and standard errors that are misleading. Analysis of the 1995 data confirms the presence of this type of correlation. Failing to account for it produces confidence intervals that are too small and could lead to erroneous conclusions regarding the profitability of VRA.

The conceptual framework also indicates that a necessary condition for precision agriculture is an interaction between managed inputs and other unmanaged factors that vary between or within fields. Testing for this interaction can be accomplished with site-specific crop response functions by testing for a difference in the linear and quadratic response to managed inputs. This type of interaction is confirmed with our data, which indicates crop response to

nitrogen varied significantly within the fields, so there is the potential to increase returns to nitrogen by using VRA. Estimates of this potential suggest that if the cost of implementing VRA was less than 14.5 \$ ha^{-1} at Hanska or 48.3 \$ ha^{-1} at Morgan there was at least a 95 percent chance that VRA would have increased profitability.

Site-specific crop response functions were estimated with fields divided into approximately 0.75 and 0.093 ha regions in order to explore the potential benefits of managing a field with a finer degree of resolution. The benefits of estimating crop response functions with greater resolution are an improvement in model fit and a smaller variance in the regression error that tends to decrease the size of estimated confidence intervals, and a larger estimate for the potential profitability of VRA. The cost is a loss in degrees of freedom that tends to increase the size of estimated confidence intervals. For Hanska, a finer degree of resolution increased the size of the estimated confidence intervals, while at Morgan it decreased it.

The utility of the conceptual framework and statistical model go beyond testing the precision agriculture hypothesis. For example, they can be used to identify factors within a field that influence the productivity of nitrogen applications. They can also be used to define management zones within a field by statistically comparing crop response functions for different regions of the field. The primary limitation of the statistical model is its computational intensity— estimation requires lots of time and a fast computer. However, this limitation will diminish as the speed of computers continues to increase.

References

Bongiovanni, R. and J. Lowenberg-DeBoer (2000). Nitrogen Management In Corn Using Site-Specific Crop Response Estimates From a Spatial Regression Model. *Proceedings of the 5th International Conference on Precision Agriculture*. (eds. P.C. Robert, R.H. Rust, and W.E. Larson). American Society of Agronomy/Crop Science Society of America/Soil Science Society of America, Madison, WI.

Bongiovanni, R. and J. Lowenberg-DeBoer (2001). Precision Agriculture: Economics of Nitrogen Management in Corn Using Site-Specific Crop Response Estimates From a Spatial Regression Model. Selected Paper: *American Agricultural Economics Association Annual Meeting*, Chicago, IL, August 6, 2001.

Davis, J.G., G.L. Malzer, P.J. Copeland, J.A. Lamb, P.C. Robert, and T.W. Bruulsema (1996). Using Yield Variability to Characterize Spatial Crop Response to Applied N. *Precision Agriculture: Proceedings of the 3rd International Conference* (eds. P.C. Robert, R.H. Rust, and W.E. Larson). American Society of Agronomy/Crop Science Society of America/Soil Science Society of America, Madison, WI, pp. 513-519.

Dikici, H. (2000). Seasonal Nitrogen Availability: A Site Specific Approach. Unpublished Ph.D thesis, University of Minnesota, St. Paul, MN.

Hernandez, J.A. and D.J. Mulla (2002). Comparing Statistical Analysis for Landscape Scale Experimental Designs. *6th International Congress on Precision Agriculture and Other Precision Resource Management*, July 15, Minneapolis, MN.

Hurley, T.M., G. Malzer, and B. Killian (2002a). A test of within field variation of corn response to nitrogen in central Minnesota. *Precision Agriculture: Herausforderung an Integrative Forschung, Entwicklung und Anwendung in der Praxis* (eds. A. Werner and A. Jarfe).

Kuratorium für Technik und Bauwesen in der Landwirtschaft e. V. (KTBL), Bartningstraße 49, 64289 Darmstadt. pp. 413-421.

Hurley, T.M., G. Malzer, and B. Killian (2002b). A Test Of Within Field Variation For Corn Response To Nitrogen In Central Minnesota. *6th International Conference on Precision Agriculture and Other Precision Resorce Management*, Minneapolis, MN, July 15.

Hurley, T.M., B. Killian, G. Malzer, and H. Dikici (2001). The Value of Information for Variable Rate Nitrogen Applications: A Comparison of Soil Test, Topographical, and Remote Sensing Information. Selected Paper: *American Agricultural Economics Association Annual Meeting*, Chicago, IL, August 6, 2001.

Lambert, D., R. Bongiovanni, and J. Lowenberg-DeBoer (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.

Malzer, G.L., P.J. Copeland, J.G. Davis, J.A. Lamb, P.C. Robert, and T.W. Bruulsema (1996). Spatial Variability of Profitability in Site-Specific N Management. *Precision Agriculture: Proceedings of the 3rd International Conference* (eds. P.C. Robert, R.H. Rust, and W.E. Larson). American Society of Agronomy/Crop Science Society of America/Soil Science Society of America, Madison, WI, pp. 967-975.

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

Swinton, S.M. and J. Lowenberg-DeBoer (1998). Evaluating the Profitability of Site-Specific Farming. *Journal of Production Agriculture* 11(4):439-446.

Table 1: Maximized log-likelihood and model comparisons.

	Location	
Model	**Hanska**	**Morgan**
	Maximized Log-Likelihood	
1	-2653.71	-2825.80
2	-2493.95	-2672.83
3	-2432.59	-2639.95
4	-2322.91	-2527.79
5	-2336.55	-2548.26
6	-2408.05	-2590.80
7	-2145.04	-2340.72
8	-2240.18	-2478.64
Model Comparisons (Restricted vs. Unrestricted)		
	Likelihood Ratio Statistic	
1 vs. 2	319.5[a]	305.9[a]
d.f.	2	2
2 vs. 3	122.7[a]	65.8[a]
d.f.	11	11
3 vs. 4	219.3[a]	224.3[a]
d.f.	30	30
1 vs. 4	661.6[a]	596.0[a]
d.f.	43	43
6 vs. 7	526.0[a]	500.2[a]
d.f.	81	81
5 vs. 4	27.3[a]	40.9[a]
d.f.	10	10
8 vs. 7	190.3[a]	275.8[a]
d.f.	94	94
4 vs. 7	355.7[a]	374.1[a]
d.f.	164	164

[a] $p < 0.01$

Table 2: Semi-variance parameters and average standard deviation.

	Model					
Parameter	**1**	**2**	**3**	**4**	**6**	**7**
Hanska						
Nugget ($1 - C_1 - C_x - C_s$)	1.00	0.37	0.40	0.26	1.00	0.33
Spatial (C_1)	-	0.63	0.60	0.58	-	0.52
Treatment (C_x)	-	-	0.00	0.00	-	0.00
Treatment Strip (C_s)	-	-	-	0.16	-	0.15
Range (C_R)	-	27.2	25.5	27.1	-	21.3
Average S.D.	1.16	1.25	1.14	1.13	0.78	0.78
Morgan						
Nugget ($1 - C_1 - C_x - C_s$)	1.00	0.45	0.41	0.28	1.00	0.31
Spatial (C_1)	-	0.55	0.59	0.57	-	0.51
Treatment (C_x)	-	-	0.00	0.00	-	0.00
Treatment Strip (C_s)	-	-	-	0.15	-	0.18
Range (C_R)	-	26.3	30.3	30.4	-	16.1
Average S.D.	1.54	1.55	1.59	1.56	1.05	1.06

Note: The Average S.D. is calculated as $\frac{\sum_{i=1}^{N} \sigma_{r_i s_i}}{N}$ t ha^{-1}.

Figure 1: Estimates of the potential value of variable rate nitrogen applications for selected models.

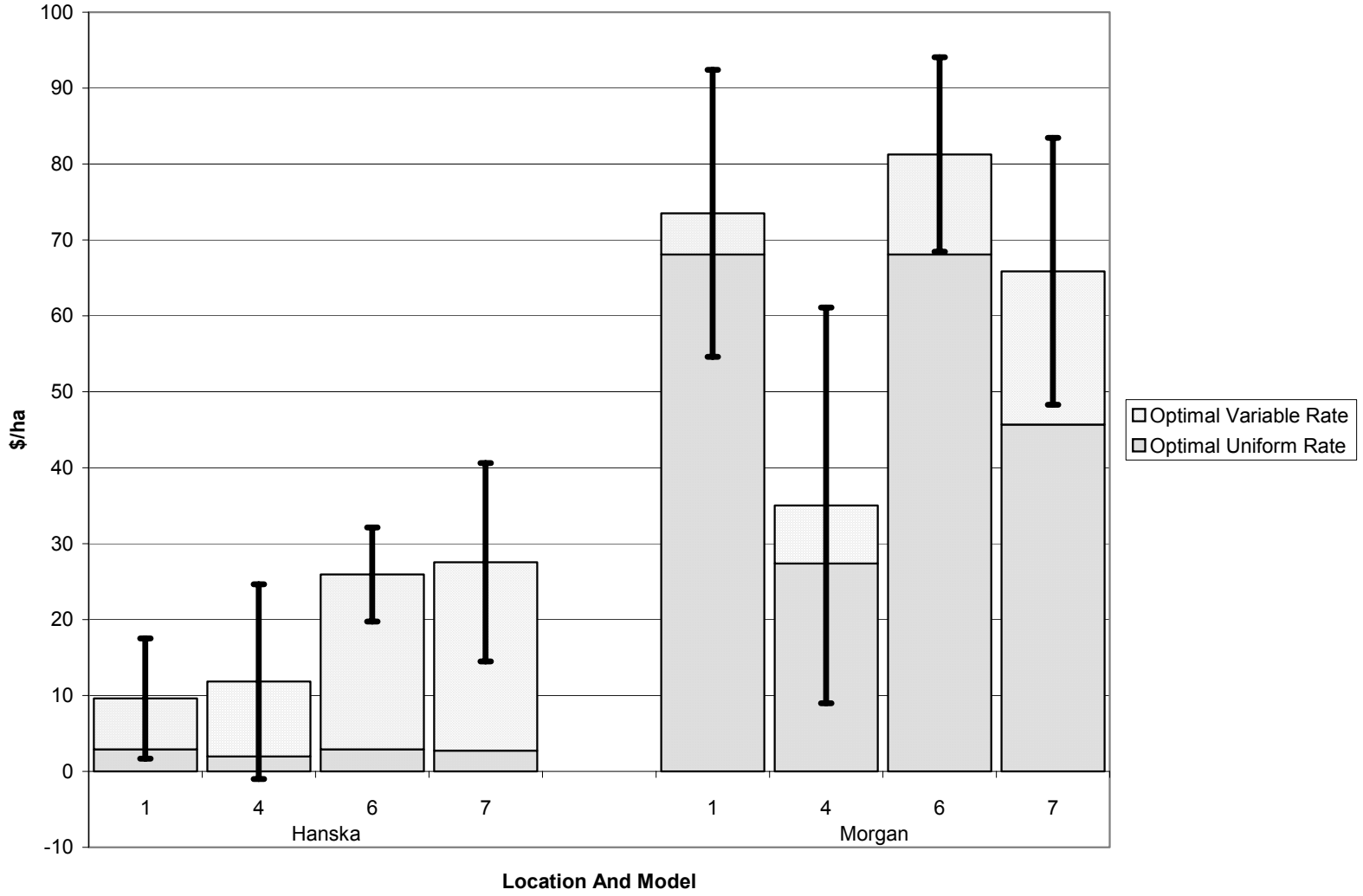
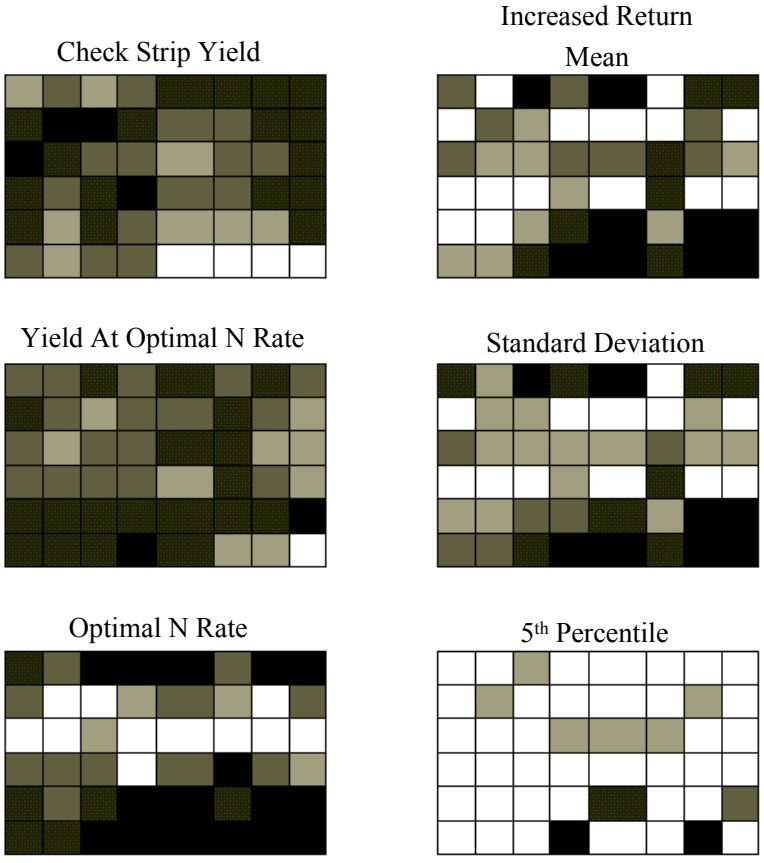
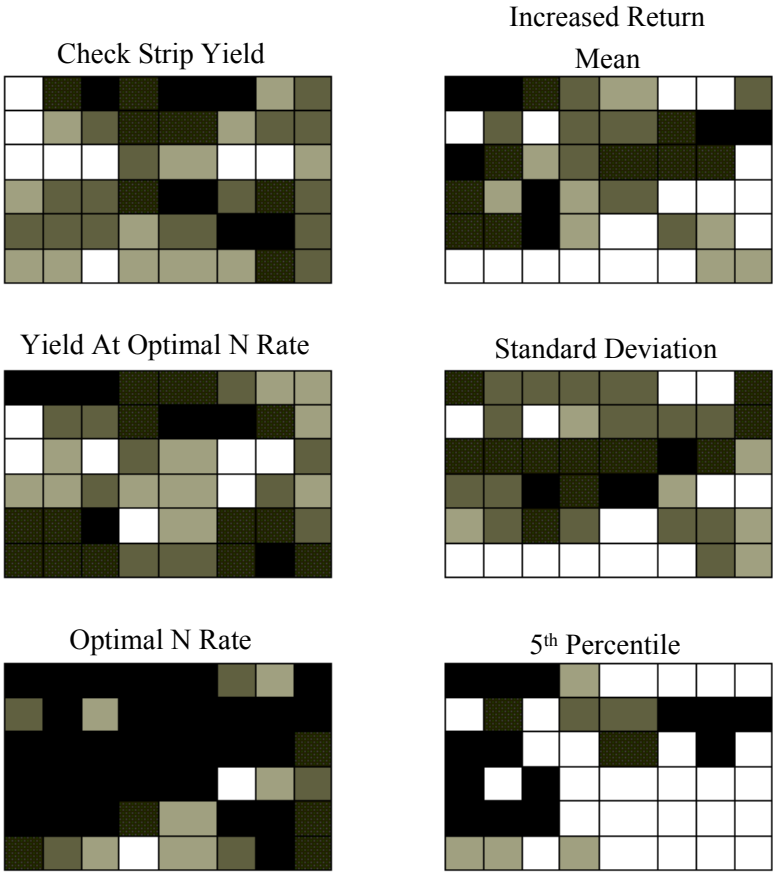


Figure 2: Check strip yield, yield at optimal nitrogen rate, optimal nitrogen rate, and increased return to nitrogen from switching to the optimal variable rate from the University of Minnesota's Extension recommended rated by region for Model 7 at Hanska.



	Check Strip Yield *t ha^{-1}*	Yield At Optimal N Rate *t ha^{-1}*	Optimal N Rate *kg ha^{-1}*	Increased Return: Mean *$ ha^{-1}*	Increased Return: Standard Deviation *$ ha^{-1}*	Increased Return: 5th Percentile *$ ha^{-1}*
	2.7 - 4.3	6.5 – 6.5	97 – 115	0.0 – 2.5	0.3 – 2.8	< 0
	4.3 - 5.7	6.5 – 9.0	115 – 127	2.5 – 6.7	2.8 – 7.9	0 – 10
	5.7 - 6.4	9.0 – 9.5	127 – 144	6.7 – 15.9	7.9 – 3.5	10 – 20
	6.4 - 7.4	9.5 – 10.2	144 – 170	15.9 – 48.3	23.5 – 6.0	20 – 30
	7.4 - 8.2	10.2 – 11.2	170 – 202	48.3 – 176.1	46.0 – 76.3	> 30
Average	6.2	9.4	154			

Figure 3: Check strip yield, yield at optimal nitrogen rate, optimal nitrogen rate, and increased return to nitrogen from switching to the optimal variable rate from the University of Minnesota's Extension recommended rated by region for Model 7 at Morgan.



	Check Strip Yield	Yield At Optimal N Rate	Optimal N Rate	Increased Return Mean	Increased Return Standard Deviation	Increased Return 5th Percentile
	t ha^{-1}	*t ha^{-1}*	*kg ha^{-1}*	*$ ha^{-1}*	*$ ha^{-1}*	*$ ha^{-1}*
	3.6 - 4.7	8.3 – 9.5	109 – 124	0.0 – 16.7	0.2 – 10.7	< 0
	4.7 - 5.7	9.5 – 10.3	124 – 147	16.7 – 40.1	10.7 – 24.9	0 – 10
	5.7 - 6.8	10.3 – 10.9	147 – 161	40.1 – 76.4	24.9 – 44.0	10 – 20
	6.8 - 8.0	10.9 – 11.8	161 – 186	76.4 – 151.7	44.0 – 59.3	20 – 30
	8.0 - 9.8	11.8 – 12.9	186 – 202	151.7 – 274.2	59.3 – 78.5	> 30
Average	6.3	10.7	184			