

Optimal Sampling Under a Geostatistical Model

Gregory R. Pautsch, Bruce A. Babcock, and F. Jay Breidt

CARD Working Paper 98-WP 200

October 1998

Center for Agricultural and Rural Development
Iowa State University

OPTIMAL SAMPLING UNDER A GEOSTATISTICAL MODEL

Introduction

The marginal product of fertilizer applications varies across crop years and both between and within crop fields. This variability is caused by variations in weather and a nonuniform distribution of soil and production factors such as organic material, nutrient availability, soil moisture, landscape position, pest pressure, soil compaction, drainage, and rooting depth [Sawyer, 1994]. Soil sampling has long been advocated as a means of improving the efficiency of fertility management decisions by better matching fertilizer applications with crop nutrient requirements and nutrient availability. Advances in mapping and sensing technologies have renewed interest in soil sampling as a means of moving to variable rate technologies (VRT) whereby a farmer varies fertilizer applications across space and/or time. Significant research efforts are under way to develop the knowledge and equipment needed to allow farmers to move to variable rate technologies (VRT) [National Research Council, 1997].

Recent studies examining the potential value of switching to a VRT fertilizer program assume producers possess complete information about soil nitrate levels, as well as how to vary fertilizer applications optimally across the field [Babcock and Pautsch, 1997; Lowenberg-DeBoer and Boehlje, 1996; Snyder et al., 1996; Solohub et al., 1996; Hertz, 1994]. In reality, farmers using a VRT strategy will only sample a portion of the field rather than the entire field. The soil samples may then be used, through statistical procedures, to estimate the soil nitrate level in nonsampled sites. These estimates can then be summarized and presented to the producer in the form of a soil nitrate field map where isoclines of equal soil nitrate levels are shown to guide fertilizer rates.

A key factor in such a map is the precision with which the nonsampled points are estimated. Precision can be increased with more soil samples, but at a cost. This paper develops a framework for determining the optimal number of soil samples when applying nitrogen fertilizer under a variable rate program. The optimal sample size is found by

equating the marginal cost of sampling with the marginal benefit of sampling. The marginal benefit of soil sampling is the increased expected returns from an additional soil test. The marginal cost of sampling is the additional cost of obtaining a soil sample.

Two methods are used to estimate soil nitrate levels and subsequent marginal benefits of soil sampling. First, the soil nitrate estimates are treated as “truth” and directly inserted into optimal fertilizing rules. This approach is called *the plug-in method* and is most widely prescribed in precision farming and other agricultural studies. The plug-in method, however, ignores estimation risk and is not consistent with expected utility maximization [Klein et al., 1978]. The second approach accounts for estimation risk by using Bayesian decision rules. The approach is consistent with expected utility maximization [DeGroot, 1970] but with the exception of a few studies [Chalfant et al., 1990; Lence and Hayes 1994, 1995; Babcock et al., 1996] it has not been widely adopted into the agricultural literature.

The Bayesian decision rule is used to determine how increases in soil nitrate variability and the spatial correlation of soil nitrate across a field affect the optimal number of soil samples. Increased variability and decreased correlation would seem to increase optimal sample size because more samples are needed to make reliable estimates of nitrate. We show that increased variability increases the optimal number of soil samples but increased spatial correlation may increase or decrease the optimal sample size depending on the level of marginal sampling costs.

The Model

A geostatistical model is used to simulate the soil nitrate levels occurring naturally in the field. For each of 1,000 draws in a Monte Carlo simulation, soil samples are taken at various nonstrategic points in the field and are used to create an estimated soil nitrate map. Producers then fertilize according to their estimated soil nitrate map and whether or not they account for estimation risk. The accuracy of the map depends upon the sample size. For each sample size, the results of using a VRT fertilizer program are averaged over the same 1,000 draws of possible soil nitrate levels. The marginal benefit of sampling in a VRT fertilizer program is the change in producer returns divided by the change in the number of samples. The producer returns from an SRT fertilizer program

are also averaged over the same 1,000 draws of possible soil nitrate levels and compared with the VRT fertilizer program.

Field Data Simulation

A field is overlaid with a square grid, the nodes of which represent sites for possible soil sampling. All sites possess a common mean soil nitrate level, μ . Soil nitrate levels, however, vary among the sites. A spherical semi-variogram is assumed to portray the spatial dependence of soil nitrate levels within the field. The spherical model is currently the most commonly used semi-variogram in soil science [Han et. al, 1996]. The spherical semi-variogram is given by,

$$\begin{aligned} \gamma(h) &= C_o + C[\frac{3}{2}(h/a) - \frac{1}{2}(h/a)^3] && \text{for } 0 \leq h < a \\ &= C_o + C && \text{for } h \geq a, \end{aligned} \quad (1)$$

where h is the distance between any two sites.

The sill, $C_o + C$, represents the overall variation of soil nitrate levels in the field. This overall variation of soil nitrate levels consists of a local random component, C_o , called the nugget effect and a component, C , that can be explained spatially. The range, a , is the distance at which the nitrate levels in a field become uncorrelated. If the distance between any two sites in a field is less than the range then the nitrate level at one site provides some information about the nitrate level at the other site.

The spatial covariance of nitrate levels within the field is represented by

$$\begin{aligned} \sigma_{ij}(h_{ij}) &= C[1 - \frac{3}{2}(h_{ij}/a) + \frac{1}{2}(h_{ij}/a)^3] && \text{for } 0 \leq h < a \\ &= 0 && \text{for } h \geq a. \end{aligned} \quad (2)$$

The covariance (σ_{ij}) of soil nitrate levels between sites i and j depends on the distance between sites i and j (h_{ij}). The soil nitrate levels between adjacent sites are more related than nitrate levels from sites farther apart. If the distance between sites i and j is greater than or equal to the range, then the corresponding nitrate levels are uncorrelated, $\sigma_{ij} = 0$. Denote the covariance matrix of the soil nitrate levels as $\varphi = [\sigma_{ij}]$.

Cholesky's factorization of the covariance matrix φ is denoted as P , where P is a lower triangular matrix and $PP' = \varphi$. Denote x as the column vector containing the soil nitrate levels on each of the sites. Let x equal $Pz + \mu \mathbf{1}$, where z is a column vector

drawn randomly from a standard normal distribution, $\mathbf{1}$ is the unit column vector, and μ is a constant. In this manner, the soil nitrate levels occurring in the field before fertilizer application are normally distributed with mean μ and covariance structure φ .

Sampling and Soil Nitrate Maps

To simplify the analysis, it is assumed that the true underlying process (semi-variogram) that generates the spatial distribution of soil nitrate levels is known when making estimates. This assumption represents a first step in combining geostatistical procedures and precision farming concepts to derive optimal sample sizes. If the semi-variogram is not known, then one must be estimated from the sampled values. When using the Monte Carlo simulation technique, such an endeavor is difficult and very time consuming when performed for each replication.

Suppose n different sites are sampled and the soil nitrate readings are represented by $w = (w_1, \dots, w_n)'$, where w_j is equal to the soil nitrate level at the j^{th} sampled site. The nitrate levels at sampled sites are then used to estimate the nitrate levels at the nonsampled sites. Since the inherent soil nitrate levels are normally distributed, the joint distribution of $(x_i, w)'$, where x_i is the soil nitrate level at a nonsampled site, is multivariate normal with mean vector $(\mu, \mu \mathbf{1}'_n)$ and covariance matrix,

$$\begin{bmatrix} C_o + C & \text{Cov}(x_i, w_1) & \dots & \text{Cov}(x_i, w_n) \\ \text{Cov}(w_1, x_i) & \text{Cov}(w_1, w_1) & \dots & \text{Cov}(w_1, w_n) \\ \vdots & \vdots & & \vdots \\ \text{Cov}(w_n, x_i) & \text{Cov}(w_n, w_1) & \dots & \text{Cov}(w_n, w_n) \end{bmatrix} = \begin{bmatrix} C_o + C & \varphi'_i \\ \varphi_i & \varphi \end{bmatrix} \quad (3)$$

where φ_i is $n \times 1$ and φ is $n \times n$. The conditional distribution of x_i given w is then normal with mean and variance [Graybill, 1976],

$$\hat{x}_i = E[x_i | w] = \mu + \varphi'_i \varphi^{-1} (w - \mu \mathbf{1}'_n), \quad (4)$$

$$\text{Var}(x_i | w) = (C_o + C) - \varphi'_i \varphi^{-1} \varphi_i. \quad (5)$$

The covariance of the i^{th} nonsampled point with each of the n sampled points is represented by φ_{io} and its transpose is denoted as φ'_{io} . The covariance of the sampled sites with the other sampled sites is represented by φ and its inverse is denoted as φ^{-1} .

If none of the sampled points is within the range of the i^{th} nonsampled site, the covariance between it and all the sampled sites is zero. No additional information on the i^{th} nonsampled site is gained and the conditional mean and variance become the overall unconditional field mean and variance.

A nonstrategic sampling procedure is used in the analysis. Currently, strategic sampling of a field has not been introduced in the precision farming literature. Strategic sampling could take the form of finding a soil nitrate covariate such as topography and sampling according to the field topography. Chin, for example, demonstrates that on Iowa corn fields, intra-field variations in soil nitrate levels are correlated with soil organic material, which, in turn, are correlated with slope, orientation, and the sand and clay content of a soil.

For simplicity, the nonstrategic sampling procedure used for moderate and large sample sizes was to select points at the intersection of every x_1 rows with every x_1 columns. For example, the 25 sample points were selected at the intersection of every 14th row with every 14th column. Table 1 presents the sampling procedure when the sampled points are greater than or equal to 25. For smaller sample sizes, the points were selected to maximize the number of nonsampled sites that could be estimated. The results from four different single sites were averaged and represent the first sample point case. Four sample points were chosen so none of the sites was within 30 grids of one another (the range is 15 grids in any direction); i.e., no points overlapped with another. A fifth sampled site was added, which partially overlapped the previous four sampled sites. A vast majority of the field could be estimated from only five sample points.

Decision Model

The production decision is the amount of nitrogen fertilizer to apply given the relationship between soil nitrate concentrations and yield, the available technology (SRT versus VRT), and the producer's information concerning inherent soil nitrate levels. The soil nitrate concentration, measured in parts per million (ppm), represents the available nitrate in the top 12-inch layer of soil. A producer can alter the soil nitrate concentration by applying an amount of nitrogen fertilizer (F) measured in pounds per acre. The soil nitrate concentration after applying fertilizer (N^{AF}) is assumed to be a linear relationship

of the nitrogen found naturally in the soil (x) and the amount of nitrogen fertilizer applied [Babcock, et al., 1996]. The multiplicative constant k indicates the pounds of fertilizer per acre needed to increase the soil nitrate concentration one ppm,

$$N^{AF} = x + Fk. \quad (6)$$

The existence of a corn yield plateau and an approximately linear response to soil nitrates when nitrates are limiting is supported in the literature [Ackello-Ogutu et al., 1985; Cerrato and Blackmer, 1990; Paris, 1992; Binford et al, 1992]. A review of linear response plateau (LRP) production function research is found in Jomini (1990). The following LRP production relationship is used, assuming that all other input decisions have been made and are at nonbinding levels:

$$Y_i = Y_p - b(N^* - N_i^{AF})I_{\{N_i^{AF} < N^*\}}. \quad (7)$$

For each site i , the indicator variable $I_{\{N_i^{AF} < N^*\}}$ equals 1 when the nitrogen level after fertilizing is less than the critical level of nitrogen (N^*) and equals 0 otherwise. The plateau or maximum corn yield (Y_p) is reached when the soil nitrate concentration after fertilizing is greater than or equal to N^* . When the soil nitrate concentration is less than N^* , the corn yield (Y_i) decreases linearly by a constant level (b) for each ppm less than N^* .

The optimal SRT fertilizer rate is the single rate that, when applied to the entire field, maximizes the producer's expected profit. The spatial correlation and distribution of inherent soil nitrate levels are known, but information on spatial location is not used in SRT. The SRT nitrogen fertilizer optimization procedure is

$$\text{Max}_F E[\pi^{SRT}] = \text{Max}_F E\left[\sum_{i=1}^n P_c(Y_p - b(N^* - (x_i + kF))I_{\{x_i + kF < N^*\}}) - P_F F\right], \quad (8)$$

where n is the number of grid cells in the field, P_c is the price of corn, and P_F is the price of nitrogen fertilizer. Since each x_i is normally distributed with mean μ and variance $C_o + C$, equation (8) is rewritten as,

$$\text{Max}_F \sum_{i=1}^n \left[P_c \{Y_p - b(N^* - (\mu + kF))G\left(\frac{N^* - (\mu + kF)}{\sqrt{C_o + C}}\right) - b\sqrt{C_o + C} g\left(\frac{N^* - (\mu + kF)}{\sqrt{C_o + C}}\right)\} - P_F F \right], \quad (9)$$

where $g(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$ is the standard normal probability density function and

$G(z) = \int_{-\infty}^z g(u)du$ is the corresponding cumulative distribution function. The first-order condition for the optimal SRT fertilizer rate is then

$$P_c G\left(\frac{N^* - (\mu + kF)}{\sqrt{C_o + C}}\right) bk - P_F = 0, \quad (10)$$

from which the optimal SRT fertilizer rate is determined to be

$$F = \frac{N^* - \mu}{k} - \frac{\sqrt{C_o + C}}{k} G^{-1}\left(\frac{P_F}{P_c bk}\right). \quad (11)$$

When using variable rate technology to make fertilizer decisions, the producer possesses a field map of estimated soil nitrate levels. The map is based on the soil samples. Let w represent the vector of sampled nitrate levels at the sampled sites. The producer's posterior beliefs, after the samples are taken, regarding the i^{th} site's inherent soil nitrate level is denoted by $h(x_i|w)$. The optimal expected VRT profit for the entire field is the sum of the optimal expected profit from each site. The optimal VRT fertilizer rate for the i^{th} site is the rate that maximizes the producer's expected profit on that site,

$$\text{Max}_{F_i} E[\pi_i^{\text{VRT}}] = \text{Max}_{F_i} \int_{-\infty}^{+\infty} [P_c(Y_p - b(N^* - (x_i + kF_i)))I_{\{x_i + kF_i < N^*\}} - P_F F_i] h(x_i|w) dx_i. \quad (12)$$

The form of the posterior beliefs about the inherent soil nitrate level depends upon whether the site is a sampled or nonsampled site and whether the producer ignores or accounts for estimation risk. Soil sampling errors are assumed to be zero, so that producers have perfect information about the true nitrogen level at each sampled site. The posterior beliefs about the soil nitrate level at a sampled site become a point density function at the sampled value. Given perfect soil nitrate information, current prices of corn and nitrogen, and the marginal product of nitrogen fertilizer, the economically optimal fertilizer response is to raise the soil nitrate level to the physically optimum level N^* . If the producer were deciding whether or not to fertilize, then the optimal fertilizer prescription would also include application costs. In our analysis, the producer has already decided to fertilize, so application costs are ignored and treated as a fixed cost. At each sampled site, a producer fertilizes in the following fashion:

$$\begin{aligned}
F_i &= (N^* - x_i) / k && \text{if } 0 < x_i < N^* \\
&= 0 && \text{if } x_i \geq N^*
\end{aligned} \tag{13}$$

Producers do not possess perfect information about soil nitrogen levels at nonsampled sites. Instead, producers use estimated soil nitrate levels derived from the sampled sites to make their fertilizer decisions. Nitrogen fertilizer decisions are analyzed under two different assumptions. First, producers ignore estimation risk by directly substituting the estimate for the true unknown level of soil nitrate at each nonsampled site. This method is traditionally referred to as the *plug-in* approach. The posterior density, $h(x_i|w)$, in this case is a point density function at the estimated value $\hat{x}_i = E[x_i|w]$ (equation 4) for each nonsampled site. The optimal fertilizer rate is found by replacing the true soil nitrate level x_i with its estimate \hat{x}_i in equation (13).

The second procedure accounts for estimation risk by using a Bayesian approach. The posterior distribution of the true soil nitrate level for nonsampled sites is found by updating the prior beliefs using Bayes's Theorem. The posterior beliefs are then conditional upon the sampled values at the sampled sites. The posterior density, $h(x_i|w)$, is normal with mean $\hat{x}_i = E[x_i|w]$ and variance $Var(x_i|w)$, given in equations (4) and (5).

The variable rate fertilizer program maximization problem expressed in equation (12) can be rewritten as,

$$Max_{F_i} = \left[P_c \left\{ Y_p - b(N^* - (x_i + kF_i)) G\left(\frac{N^* - (\hat{x}_i + kF_i)}{Var(x_i|w)}\right) - b\sqrt{Var(x_i|w)} g\left(\frac{N^* - (\hat{x}_i + kF_i)}{Var(x_i|w)}\right) \right\} - P_F F \right] \tag{14}$$

where $g(z)$ is the standard normal probability function and $G(z)$ is the corresponding cumulative distribution function. The first-order condition for the optimal VRT fertilizer rate is,

$$P_c G\left(\frac{N^* - (\hat{x}_i + kF_i)}{\sqrt{Var(x_i|w)}}\right) bk - P_F = 0, \tag{15}$$

from which the optimal VRT fertilizer rate is determined to be

$$F_i = \frac{N^* - \hat{x}_i}{k} - \frac{\sqrt{Var(x_i|w)}}{k} G^{-1}\left(\frac{P_F}{P_c bk}\right). \tag{16}$$

Under either procedure, if an estimate other than the mean cannot be made on a site, due to the lack of locally sampled sites, then no additional information has been gained and the optimal VRT response is to use the SRT fertilizer rate.

Optimal Sample Size

Figure 1 shows the expected per acre return over fertilizer costs as a linear, then concave, function of the sample size. The linear portion for very small sample sizes reflects the possibility of drawing samples from sites that are at least twice the range in distance from each other. In this case, each sampled site provides information about the same number of, but different, nonsampled sites. Each additional sampled site on average will affect returns the same as the previously sampled sites. However, if more than one sample provides information about a nonsampled site, the later sample provides less information than the previous samples. As the sample size becomes large, each additional sample provides less and less information about the nonsampled points. Hence, expected per acre returns will eventually become a concave function of the sample size. Expected returns are strictly concave if the range of soil nitrate is high enough so that any two sampled points give information about at least one nonsampled point.

Figure 2 shows the expected marginal benefit to be constant and then decreasing with the number of samples. The marginal cost of sampling is assumed to be constant. The intersection of the marginal benefit with the marginal cost of sampling determines the optimal sample size. If MC_0 represents the marginal cost of sampling, then the marginal cost of sampling exceeds the marginal benefit at all sample sizes. The optimal producer response is to fertilize the field using a single rate fertilizer program. If the marginal cost of sampling is represented by MC_1 , then the optimal producer response is to sample n^* sites and fertilize the field using a variable rate technology program. It is assumed that the cost of investing in the capability of VRT technology has already been made. Otherwise, fixed costs would need to be accounted for in the decision to switch from an SRT fertilizer program to a VRT fertilizer program.

Empirical Analysis

Data

In the analysis, a 2,310 by 2,310 foot field is mapped onto a 70 by 70 unit grid. The field is then divided into 4,900 square units each consisting of 0.025 acres. Each square unit, 33 feet long and 33 feet wide, is assumed to possess a homogenous soil nitrate level. The overall mean and standard deviation for the soil nitrate levels within the field are assumed to be 15 ppm and 5 ppm, respectively. The range of soil nitrate coefficients of variation occurring naturally in Iowa corn fields is estimated to be in the range of {0.08, 0.43} [Chin, 1997]. Our assumed coefficient of variation of 0.33 occurs near the upper end of this interval. Hence, the estimated value of switching to a VRT fertilizer program may be slightly higher than on an average field in Iowa, since greater variability of nitrate levels increases the value of switching to VRT programs [Hennessy and Babcock, 1998].

The nugget of the semi-variogram is assumed to be zero, since samples were assumed to be measured without error. All of the variability in soil nitrate levels can be explained spatially. The range of the semi-variogram is assumed to be 15 grid units (or 495 feet), so that the nitrate level at one point provides some information about the nitrate level at the other points within 15 grid units. Our assumed range is very close to the midpoint of the interval (131 to 900 feet) typically found in precision farming studies of soil nitrate concentrations [Wollenhaupt et al., 1997]. The range of the semi-variogram provides the spatial covariance structure, Φ , of inherent soil nitrate levels within the field. A Monte Carlo simulation is performed by averaging results over 1,000 draws on the same field, drawn from a normal distribution with mean soil nitrate level of 15 ppm, standard deviation of 5 ppm, and covariance structure Φ .

The production process assumes a continuous corn rotation. The corn yield plateau (Y_p) is 148.21 bushels per acre, the slope coefficient (b) is 3.95 bushels per ppm, and the critical level of inherent soil nitrate concentration (N^*) is 24.45 ppm [Babcock and Blackmer, 1992]. To raise the soil nitrate concentration 1 ppm, the producer needs to add 7.63 pounds of nitrogen fertilizer ($k = 1/7.63$) [Babcock and Blackmer, 1992]. The price of corn is \$2.50 per bushel and the price of nitrogen fertilizer is \$0.20 per pound.

Single Rate Fertilizer Program

The SRT fertilizer rate is the field application rate that maximizes the producer's profit over all 1,000 draws. This rate can be thought of as the single rate of fertilizer an experienced producer applies to the field. In Tables 2 and 3, zero sample points represent a single rate fertilizer program. The SRT fertilizer rate is 110.91 pounds of fertilizer per acre and the average per acre returns over fertilizer costs are \$344.34. Under the SRT program, producers over-apply nitrogen fertilizer to insure against possible yield losses [Babcock, 1992; Babcock and Blackmer, 1992]. Producers over-fertilize 85 percent of the grid cells and overapply the field with 5,059 pounds of nitrogen fertilizer. The average per acre yield of 146.61 bushels is 99 percent of the maximum potential yield. Only 15 percent of the grid cells are under-fertilized and only 379 pounds of fertilizer are needed for those grid cells to reach their optimum yield potential.

Variable Rate Program—Plug-In Method

Table 2 presents the per acre yields, fertilizer rates, and returns over fertilizer costs for various sample sizes under the plug-in approach. If producers ignore estimation risk and use a sample size of less than 100 to generate the soil nitrate map, then they are better off using the SRT fertilizer program than a VRT program. Returns decline because the producer uses a suboptimal decision making process by treating the soil nitrate estimates as completely accurate. This process is equivalent to assuming that the producer no longer over-fertilizes to insure against yield losses. The percentage of land over-fertilized and the amount of nitrogen fertilizer overapplied decline. Yields decline by as much as 5.05 bushels per acre as the land underfertilized and the amount of nitrogen fertilizer needed to reach maximum yield potential increase.

Soil nitrate estimates can be generated for every grid cell in the field when the sample size is greater than or equal to 25. In these cases, half of the soil nitrate estimates over-estimate the true soil nitrate level leaving those grid cells undersupplied with nitrogen and half of the soil nitrate estimates underestimate the true soil nitrate level leaving those grid cells over-supplied with nitrogen. The amount of fertilizer over-applied in parts of the field is very close to the amount of fertilizer needed in other parts of the field. Hence, the average fertilizer rate is fairly constant regardless of the amount

of information acquired. The misapplication of fertilizer, however, decreases as the sample size increases, since better estimates are being generated from increased soil nitrate information. Reducing the misapplication of fertilizer increases both yields and returns. However, for yields to equal the SRT level, approximately half of the grid cells (2,450) would need to be sampled. The misapplication of fertilizer is completely eliminated and yields reach their maximum potential when the producer has perfect information by sampling all 4,900 grid cells.

The plug-in approach, despite its suboptimal nature, is often prescribed in the precision farming literature. Producers are typically directed to fertilize so that the average soil nitrate level reaches its critical level. Fertilizer prescriptions are usually equal to the amount of fertilizer needed to raise the average soil nitrate estimate to the critical level of nitrogen.

Variable Rate Program—Bayesian Method

Table 3 presents the per acre yields, fertilizer rates, and returns over fertilizer costs for various sample sizes under the Bayesian approach. The Bayesian approach assumes that producers account for estimation risk. After each sample, producers improve or update their beliefs about the mean and variance of the soil nitrate levels. The additional information reduces the amount of misapplication of nitrogen fertilizer, both the amount of fertilizer needed and the amount of fertilizer over-applied. Regardless of the sample size, a variable rate program using the Bayesian approach always produces higher yields, higher returns, and less over-fertilization than an SRT fertilizer program.

Table 3 shows for many of the sample sizes that the land under-fertilized is approximately 15 percent. With a VRT program the first-order condition for the optimal fertilizer rate given the updated beliefs is given in equation (15), where $G(\cdot)$ represents the probability that the soil nitrate level after fertilizing is less than the critical level of nitrogen (N^*) or equivalently that yield is less the maximum potential yield. Given the values of b , k , P_c , and P_f , $G(\cdot)$ equals 0.1545. Therefore, each nonsampled grid in a field has a probability of 15.45 percent of being under-fertilized and a probability of 84.55 percent of being over-fertilized. Hence, approximately 15 percent of the land that is not properly fertilized will be under-fertilized.

Comparing Tables 2 and 3 reveals that VRT per acre returns over fertilizer costs are always higher with the Bayesian approach than with the plug-in approach. The Bayesian approach deals with estimation risk in a manner that is consistent with expected profit maximization [Lence and Hayes, 1994]. The plug-in approach is easier to implement but it is not consistent with expected profit maximization [Lence and Hayes, 1994]. Producers using a VRT fertilizer program that strictly fertilizes according to an estimated map (plug-in approach) are using a suboptimal decision-making process.

Variability of SRT and VRT Returns

Tables 2 and 3 also present the standard deviation of per acre returns from variations in soil nitrate. This measure reflects the variability of producer returns when using the SRT and VRT fertilizer programs. Under an SRT fertilizer program the variability of producer returns is very low at \$1.56 per acre. The over-fertilization of the SRT program has a stabilizing effect on returns by reducing the risks of yield losses. Under a Bayesian VRT fertilizer program, the variability of producer returns declines even farther whenever the sample size increases. The increased soil nitrate information leads to better mapping accuracy and better decision-making, reducing the variability of returns and over-fertilization. Under a plug-in VRT fertilizer program, the increased information is used suboptimally, leading to suboptimal decision-making and increasing the variability of returns. Eventually, enough information is acquired (and used suboptimally) to reduce the variability of returns below the SRT level.

Marginal Benefit and Cost of Sampling

Table 4 presents the VRT marginal producer and environmental benefits from sampling. Under the plug-in approach, the marginal returns are first negative, then increase to \$14.48, and subsequently decline. Marginal environmental benefit is very large at first, 479.02 pounds of fertilizer for the field, and then declines to 0.11 pounds of fertilizer. The large environmental benefit and large reduction in returns with very small sample sizes is from producers no longer over-fertilizing to insure against yield losses. Instead, producers are treating imperfect soil nitrate maps as truth and, as a result, are suffering from yield losses. If the marginal cost of sampling and other additional VRT

costs exceed \$4.02 per sample, producers are better off with an SRT fertilizer program than a VRT program that fertilizes according to an estimated map.

Under the Bayesian approach, marginal returns over fertilizer costs and marginal environmental benefits decline as the sample size increases. If the marginal cost of sampling and other additional VRT costs exceed \$10.30 per sample, the profit from an SRT fertilizer program exceeds that of a VRT fertilizer program. The marginal environmental benefit is also quite low. The first four sample points each reduce over-fertilization in the field by 38.66 pounds (or 154.64 pounds total).

The marginal cost of obtaining a soil nitrate sample is approximately \$9 per sample [Lowenberg-DeBoer and Swinton, 1997]. Hence, a variable rate fertilizer program using the Bayesian approach appears to be feasible for only very small sample sizes; i.e., five or fewer sample points or sampling approximately 0.1 percent of the possible points in the field. However, other costs of moving to variable rate technology should be included such as new fertilizer spreaders, computer hardware and software, global positioning systems, and any additional labor costs.

Effect of Variability and Correlation on Optimal Sample Size

This section examines the effects of changing the spatial correlation and variability of soil nitrate levels within a field on the marginal benefits from sampling and on the optimal sample size. Marginal costs are assumed to remain constant. The Bayesian method, not the plug-in method, of using estimated soil nitrate mappings is highlighted, since it is consistent with expected profit maximization.

Spatial Correlation

Changing the range in the spherical semi-variogram alters the spatial correlation of soil nitrate levels. The spatial correlation coefficient of soil nitrate levels for a spherical semi-variogram is

$$\rho(h_{ij}) = \frac{C * [1 - \frac{3}{2}(h_{ij} / a) + \frac{1}{2}(h_{ij} / a)^3]}{C_o + C} \quad \begin{array}{l} \text{for } 0 \leq h_{ij} < a, \\ \text{for } h_{ij} \geq a. \end{array} \quad (17)$$

Table 5 shows that increasing the range increases the spatial correlation of soil nitrate readings. If the range is one grid unit, then all the soil nitrate levels in the field are uncorrelated. Sampling at a site provides information only about that site. On the other hand, if the range is 99 grid units, then sampling at one site provides some information about the nitrate levels at all the other sites in the field. The previous analysis assumed the range was 15 grids. For example, the spatial correlation coefficient for sites 5 grids (or 158.75 feet) apart is 0.52 when the range is 15 grids and 0.92 when the range is 99 grids.

To see how an increase in spatial correlation affects the marginal value of soil sampling, note first that the range does not affect either the optimal SRT fertilizer rate or the value of fertilizing according to the SRT rule because a does not appear in either equations (11) or (9). Next note that the value of fertilizing according to VRT under perfect information is not affected by spatial correlation. Under perfect information a does not appear in equation (16) or in equation (14) because $Var(x_i|w) = 0$. Thus the value of moving to VRT under perfect information is unaffected by an increase in spatial correlation. That is, the maximum value that can be obtained from soil sampling in a field is the same regardless of the degree of spatial correlation.

This result does not imply that the marginal benefit curves of VRT are unaffected by spatial correlation. But, because the area under a marginal benefit curve equals the value of perfect information, we know that the area under two marginal benefit curves that differ only with respect to spatial correlation must be equal.

Figure 3 shows the implication of this result. An increase in spatial correlation rotates marginal benefits from MB_1 to MB_2 . An increase in correlation increases the marginal benefit when sample size is low because each sample point reveals more information about adjoining nonsampled points. The two curves must cross at least once, however, because the areas under MB and MB_2 are equal. That is, there is a finite amount of value that can be obtained from soil sampling.

When marginal costs are relatively low, such as MC_L in Figure 3, an increase in spatial correlation reduces the optimal sample size from n to n_2 because marginal benefits at this high optimal sample rate decrease. This decrease in marginal benefit is a result of the increase in prediction capability of all previous sample points. That is, there

is a finite amount of information to be obtained, and with a higher degree of spatial correlation, a greater proportion of this information is revealed by previously sampled points. However, when marginal cost is high, such as MC_H , then an increase in correlation increases the optimal sample size from n_3 to n_4 . Thus, whether an increase in correlation increases or decreases optimal sample size depends on the level of marginal cost.

Spatial Variability

Increased variability in a field increases the potential gain from moving to variable fertilizer applications. To see this note first from equation (11) that increased variability (C) increases the optimal single rate of fertilizer application. The potential amount of fertilizer saved as one moves to variable applications increases with C . This implies that the total area under the marginal benefit curve of soil samples increases with increased spatial variability. If increased variability results in an upward shift in marginal benefits for all sample sizes, then increased variability increases the optimal number of soil samples. However, if increased variability results in a crossing of marginal benefit curves, then the optimal sample size may increase or decrease depending on the level of marginal cost, as was the result under increased spatial correlation.

Figure 4 presents expected marginal benefits for three levels of soil nitrate variability (0.16, 0.33 and 0.5) at nine soil sample levels using the Bayesian decision rule. In this range of variability and for these sample sizes, it is apparent that marginal benefits increase with increased variability. The marginal benefit of each sample increases, since each sample provides more information. The size of the increase in marginal benefits is initially quite large and then decreases rapidly as the sample size increases. Thus, given an interior solution, the optimal sample size increases with increased variability.

Figure 4 also shows that increased variability increases the likelihood of an interior solution, which will also result in increased optimal sample size. Suppose the cost of a soil sample is \$6.00. When the coefficient of variation of nitrate is 0.16, the optimal sample size is 0. That is, marginal benefits are never greater than marginal costs. Increasing variability to 0.33 creates an interior solution and the optimal sample size increases to between 25 and 36 samples. This is simply a reflection that there is a critical

amount of variability that must exist before moving to a variable-rate application method becomes economically feasible.

Conclusions

Studies examining the value of switching to a VRT fertilizer program assume the producer possesses perfect soil nitrate information. In reality, producers estimate soil nitrate levels with soil sampling. The value of switching to a VRT program from an SRT program depends greatly on how the producer uses the estimates and on the quality of the estimates.

Producers failing to account for estimation risk by strictly fertilizing to the estimated soil nitrate map are not following a VRT strategy consistent with expected profit maximization. Despite the inconsistencies, this strategy is most often prescribed to producers in the precision farming literature. To be consistent with expected profit maximization, producers should acknowledge that the soil nitrate mapping is a collection of estimates and does not provide perfect information at nonsampled sites. The soil sample information should be used in a Bayesian fashion to fine-tune or update the producer's beliefs about the soil nitrate levels in nonsampled sites.

The accuracy of the soil nitrate estimates depends on the sample size as well as the degree of spatial correlation and variability among nitrate levels within the field. Larger sample sizes, increased spatial correlation, and decreased variability improve the accuracy of the estimates and increase producer returns.

The marginal benefit of sampling increases for smaller sample sizes when there is a high degree of spatial correlation among nitrate levels. A few sampled sites are able to provide better information to a larger proportion of the field when the degree of correlation is high. Since the marginal cost of soil sampling is substantial, switching to a VRT fertilizer program appears to be more plausible for fields with a high degree of spatial correlation.

The marginal benefit of sampling increases for all sample sizes when there is greater variability in soil nitrate levels. The optimal sample size increases under a VRT fertilizer program. However, expected per acre returns decline under both SRT and VRT fertilizer programs due to the increased uncertainty surrounding soil nitrate levels.

Switching to a VRT fertilizer program from an SRT fertilizer program appears to be more plausible for fields with greater soil nitrate variability.

Table 1. Intersection location of sampled sites for sample sizes of 25 or greater

Intersection	Number of samples											
Location	25	36	49	64	81	100	144	196	324	576	1225	2450
Row	14 th	12 th	10 th	9 th	8 th	7 th	6 th	5 th	4 th	3 rd	2 nd	2 nd
Column	14 th	12 th	10 th	9 th	8 th	7 th	6 th	5 th	4 th	3 rd	2 nd	1 st

Table 2. Production and environmental summary for various sample sizes under the plug-in approach

Sampled Points	Average		Standard		Average		Average		Land		Land		Land	
	Acre Returns (\$)	Acre Returns (\$)	Deviation	Acre Returns (\$)	Acre Yield	Acre	Fertilizer Rate	Under-Fertilized	Fertilizer Needed	Over-Fertilized	Fertilizer Wasted	Properly Fertilized	percent	percent
	U.S. dollars	U.S. dollars		bu.	lb.	percent	lb.	percent	lb.	percent	lb.	percent	percent	percent
0	344.34	1.56		146.61	110.93	15	379	85	5,059	0	0	0		
1	343.35	1.99		145.77	105.40	20	577	80	4,580	0	0	0		
4	340.36	2.91		143.25	88.80	35	1,174	65	3,143	0	0	0		
5	339.53	3.20		142.55	84.21	39	1,340	61	2,747	0	0	0		
25	339.42	2.99		141.56	72.42	50	1,573	50	1,536	1	1	1		
36	340.72	2.44		142.07	72.25	50	1,454	49	1,395	1	1	1		
49	342.01	2.19		142.59	72.29	50	1,330	49	1,278	1	1	1		
64	343.27	1.85		143.10	72.36	49	1,210	49	1,165	2	2	2		
81	344.21	1.79		143.47	72.33	49	1,121	49	1,074	2	2	2		
100	344.83	1.63		143.72	72.39	49	1,061	48	1,021	3	3	3		
144	346.11	1.47		144.24	72.47	48	939	48	908	4	4	4		
196	346.86	1.45		144.54	72.46	48	868	47	836	5	5	5		
324	348.26	1.33		145.11	72.55	46	734	46	713	8	8	8		
576	349.60	1.28		145.65	72.59	44	607	43	591	13	13	13		
1225	351.34	1.26		146.35	72.63	37	441	37	430	26	26	26		
2450	353.06	1.24		147.04	72.67	25	277	24	271	51	51	51		
4900	355.98	1.22		148.21	72.72	0	0	0	0	100	100	100		

Table 4. Marginal production and environmental benefits from sampling

Sample Points	<u>Plug-In Approach</u>		<u>Bayesian Approach</u>	
	<u>Marginal</u> <u>Returns</u>	<u>Marginal</u> <u>Environmental Benefit</u>	<u>Marginal</u> <u>Returns</u>	<u>Marginal</u> <u>Environmental Benefit</u>
0	U.S. dollars	(lb. fertilizer)	(\$)	(lb. fertilizer)
1	-121.74	479.02	10.30	38.66
4	-121.74	479.02	10.30	38.66
5	-102.31	395.87	10.27	32.26
25	-24.11	60.56	8.61	28.93
36	14.48	12.76	8.49	32.21
49	12.18	9.05	7.21	25.29
64	10.27	7.49	6.09	21.35
81	6.79	5.38	4.12	15.10
100	4.02	2.80	2.33	7.98
144	3.56	2.56	2.08	7.37
196	1.76	1.38	1.07	3.89
324	1.34	0.97	0.79	2.79
576	0.65	0.48	0.38	1.36
1225	0.33	0.25	0.20	0.70
2450	0.17	0.13	0.10	0.36
4900	0.15	0.11	0.09	0.31

Table 5. Spatial correlation coefficients for various values of the range

Range (grids)	<u>Distances between sites</u>						
	0	31.75	63.50	158.75	317.50	476.25	feet grids
	0	1	2	5	10	15	
1	1.00	0.00	0.00	0.00	0.00	0.00	
15	1.00	0.90	0.80	0.52	0.15	0.00	
99	1.00	0.98	0.97	0.92	0.85	0.77	

Figure 1. Total Value of Soil Sampling

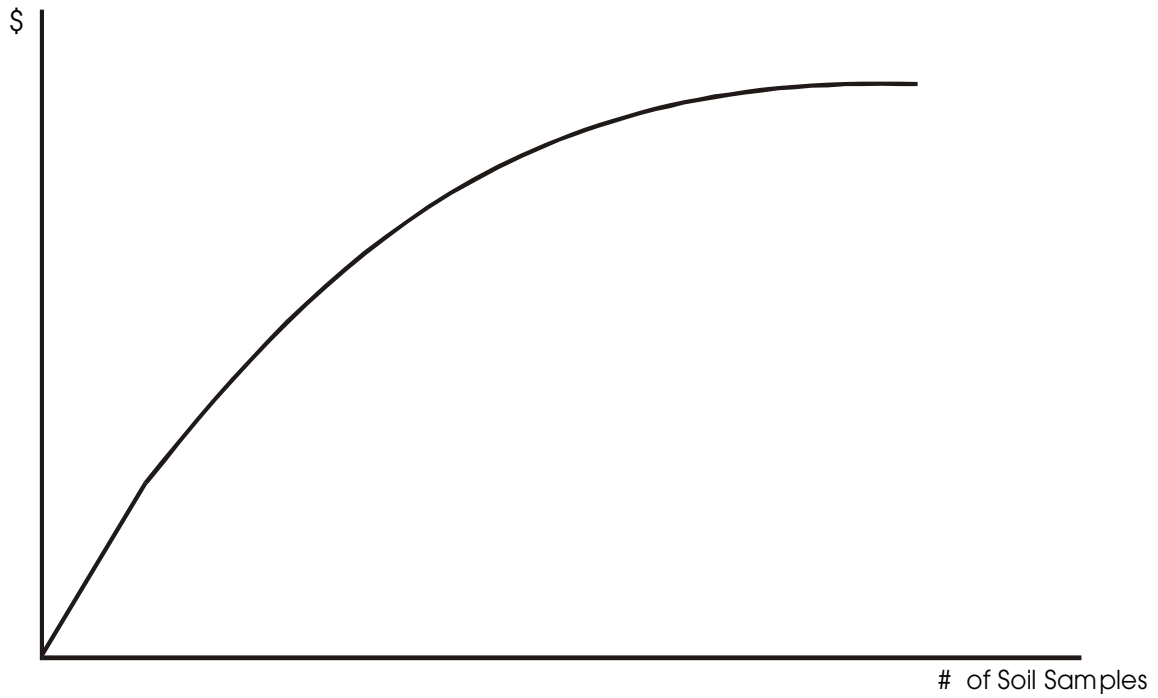


Figure 2. Incremental Value of Soil Sampling and Determination of Optimal Sample Size

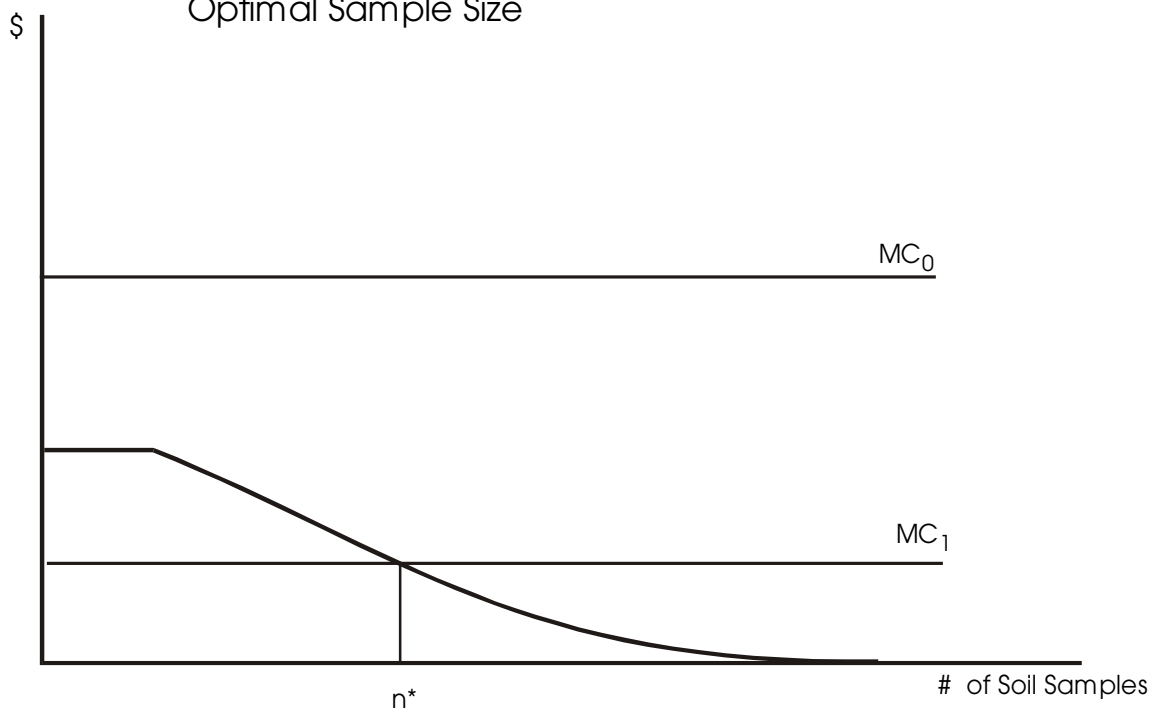


Figure 3. Effect of Increased Correlation on Optimal Sample Size for Different Marginal Cost of Soil Samples

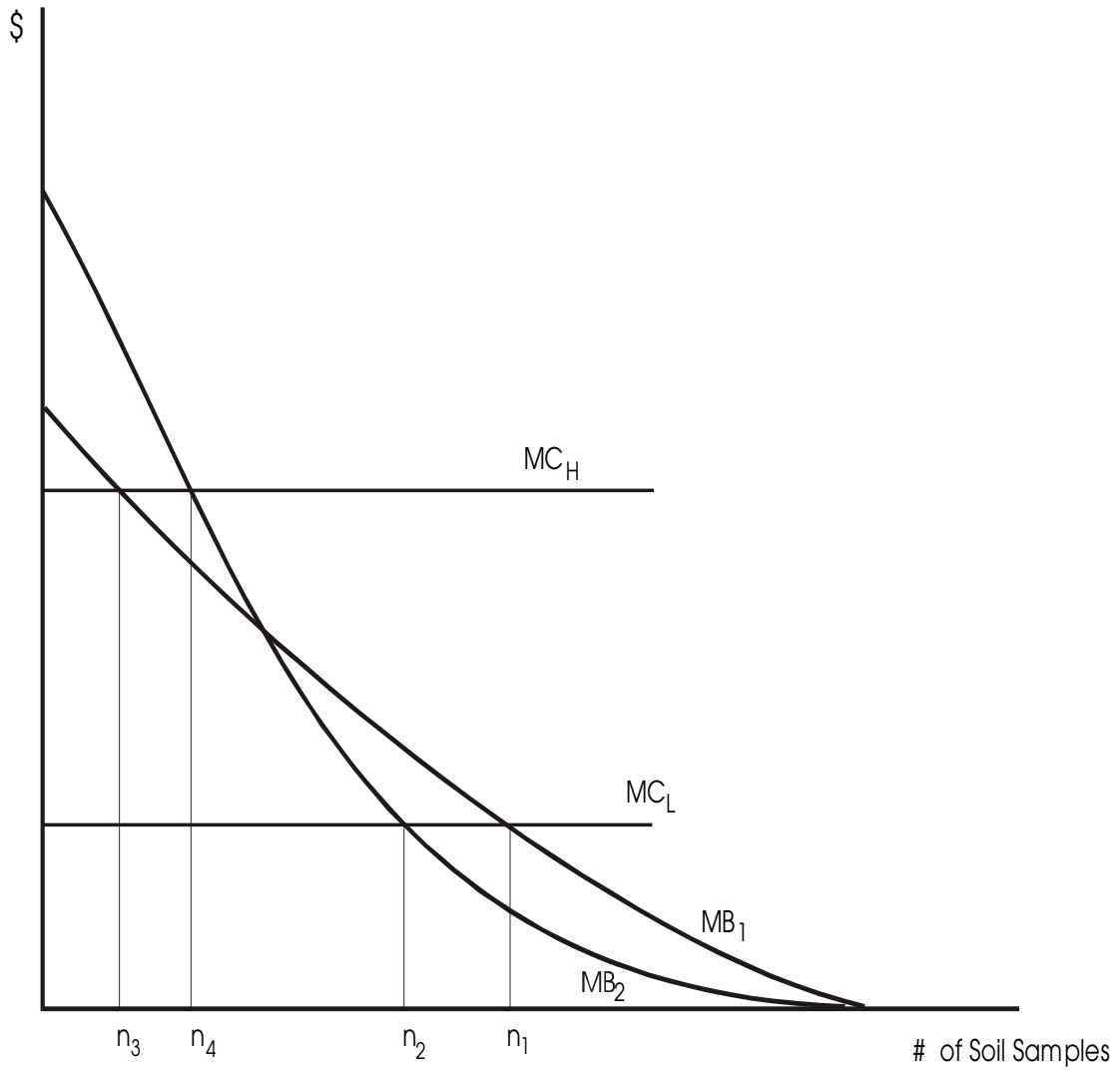
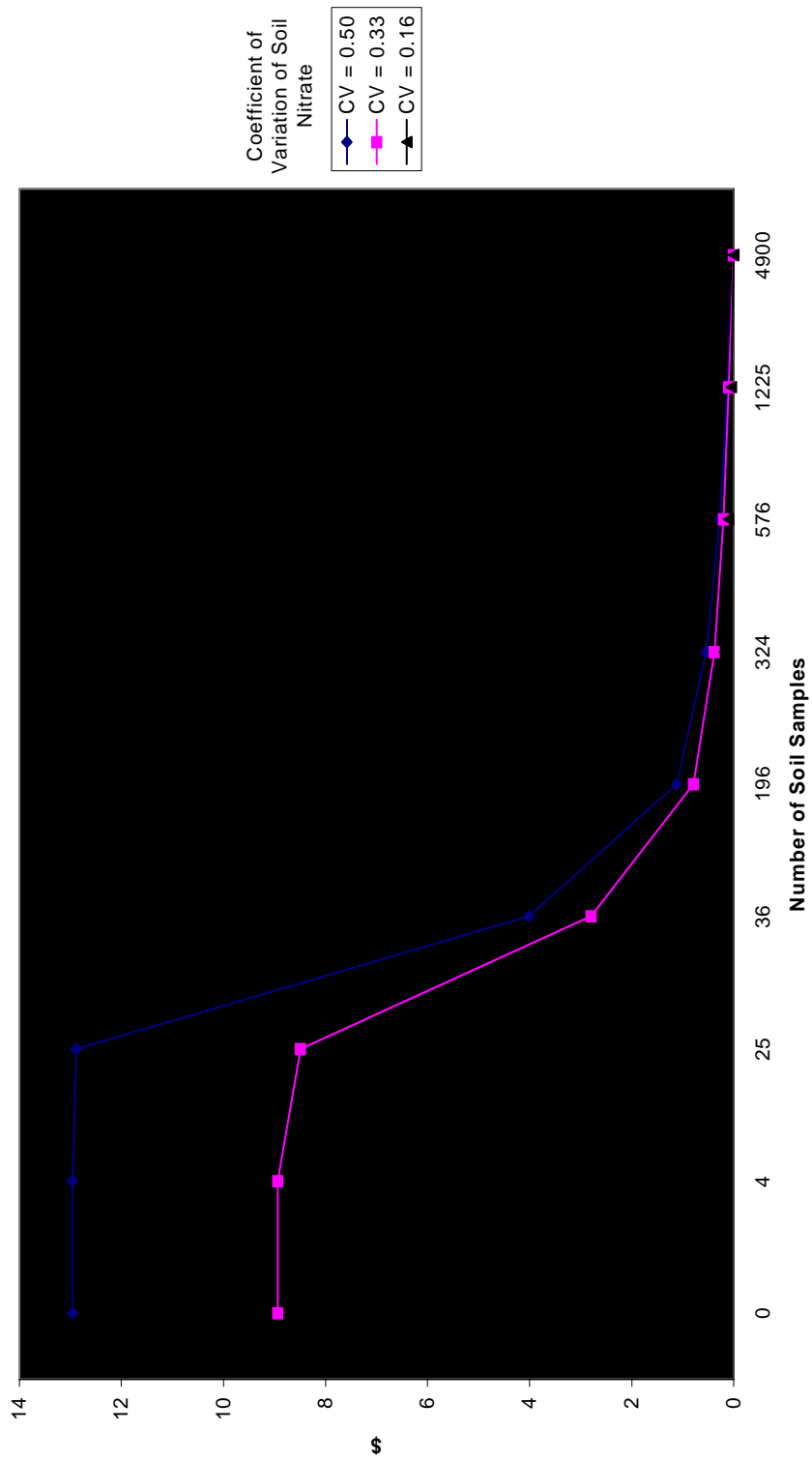


Figure 4. Effect of Increasing Variability on Marginal Benefits of Soil Sampling



REFERENCES

- Ackello-Ogutu, C., Q. Paris, and W. A. Williams. "Testing a von Liebig Crop Response Function Against Polynomial Specifications." *Amer. J. Agr. Econ.* 67(1985): 873-80.
- Babcock, Bruce A. "The Effects of Uncertainty on Optimal Nitrogen Applications." *Rev. Agr. Econ.* 14(1992): 271-80.
- Babcock, Bruce A. and Alfred M. Blackmer. "The Ex Post Relationship Between Growing Conditions and Optimal Fertilizer Rates." *Rev. Agr. Econ.* 16(1994): 353-62.
- Babcock, Bruce A., Alicia L. Carriquiry, and Hal S. Stern. "Evaluation of Soil Test Information in Agricultural Decision-Making." *Appl. Statis.* 45(1996): 447-61.
- Babcock, Bruce A. and Gregory R. Pautsch. *Moving from Uniform to Variable Fertilizer Rates on Iowa Corn: Effects on Rates and Returns*. Card Working Paper 97-WP182, Center for Agricultural and Rural Development, Iowa State University, Ames, IA, October 1997.
- Binford, G. D., A. M. Blackmer, and M. E. Cerrato. "Relationships Between Corn Yields and Soil Nitrate in Late Spring." *Agron. J.* 84(1992): 53-59.
- Carr, P. M., G. R. Carlson, J. S. Jacobson, G. A. Nielsen, and E. O. Skogley. "Farming Soils, not Fields: A Strategy for Increasing Fertilizer Profitability." *J. Prod. Agri.* 4(1991): 57-61.
- Cerrato, M. E. and A. M. Blackmer. "Comparison of Models for Describing Corn Yield Response to Nitrogen Fertilizer." *Agron. J.* 82(1990): 138-43.
- Chalfant, James A., Robert N. Collender, and Shankar Subramanian. "The Mean and Variance of the Mean-Variance Decision Rule." *Amer. J. of Agri. Econ.* 72(1990):966-74.
- Chin, Kian-Chung. *Soil Sampling for Nitrate in Late Spring*, unpublished Ph.D. dissertation, Iowa State University, Department of Agronomy, Ames, IA, 1997.
- DeGroot, M. H. *Optimal Statistical Decisions*, McGraw-Hill, New York, N. Y., 1970.
- Graybill, Franklin A. *Theory and Application of the Linear Model*, Duxbury Press, North Scituate, Mass., 1976.

- Han, Shueng, Robert G. Evans, Sally M. Schneider, and Stephen L. Rawlins. "Spatial Variability of Soil Properties on Two Center-Pivot Irrigated Fields", In *Precision Agriculture*, eds., P. C. Robert, R. H. Rust, and W. E. Larson, pp. 97-106. Proceedings of the Third International Conference on Precision Agriculture, ASA-CSSA-SSSA, Madison, WI, 1996.
- Hennessy, David A. and Bruce A. Babcock. "Information, Flexibility, and Value Added." In *Information Economics and Policy*, forthcoming.
- Hertz, C. *An economic evaluation of variable rate phosphorous and potassium fertilizer application in continuous corn*. M.S. thesis, University of Illinois, Champaign-Urbana, IL, 1994.
- Jomini, P. The Economic Viability of Phosphorous Fertilization in Southwestern Niger: A Dynamic Approach Incorporating Agronomic Principles. Ph.D. dissertation, Department of Agricultural Economics, Purdue University, West Lafayette, IN, 1990.
- Klein, R. W., L. C. Rafsky, D. S. Sibley, and R. W. Willig. "Decisions with Estimation Uncertainty." *Econometrica* 46(1978):1363-87.
- Lence, Sergio H. and Dermot J. Hayes. "The Empirical Minimum-Variance Hedge." *Amer. J. of Agr. Econ.* 76(1994):94-104.
- Lence, Sergio H. and Dermot J. Hayes. "Land Allocation in the Presence of Estimation Risk." *J. of Agr. and Res. Econ.* 20(1995):49-63.
- Lowenberg-DeBoer, J. and M. Boehlje. "Revolution, Evolution, or Dead-End: Economic Perspectives on Precision Agriculture." In *Precision Agriculture*, eds., P. C. Robert, R. H. Rust, and W. E. Larson, pp. 923-44. Proceedings of the Third International Conference on Precision Agriculture, ASA-CSSA-SSSA, Madison, WI, 1996.
- Lowenberg-DeBoer, J. and S. M. Swinton. "Economics of Site Specific Management of Agronomic Crops." In *The State of Site Specific Management for Agriculture*, eds. F. J. Pierce and E. J. Sadler, pp. 369-96, ASA-CSSA-SSSA, Madison, WI, 1997.
- National Research Council (NRC). *Precision Agriculture in the 21st Century: Geospatial and Information Technologies in Crop Management*. Washington, D.C: National Academy Press, 1997.
- Paris, Q. "The Return of von Liebig's Law of the Minimum." *Agron. J.* 84(1992):1040-6.
- Robert, P., S. Smith, W. Thompson, W. Nelson, D. Fuchs, and D. Fairchild. *Soil Specific Management*. University of Minnesota Project Report, Minneapolis, MN, 1990.

- Sawyer, J. E. "Concepts of Variable Rate Technology with Considerations for Fertilizer Application." *J. Prod. Agr.* 7(1994): 195-201.
- Sinclair, T. R. and W. I. Park. "Inadequacy of the von Liebig Limiting-Factor Paradigm for Explaining Varying Crop Yields." *Agron. J.* 85(1993): 91-6.
- Snyder, C., T. Schroeder, J. Havlin, and G. Kluitenberg. "An Economic Analysis of Variable Rate Nitrogen Management." In *Precision Agriculture*, eds., P. C. Robert, R. H. Rust, and W. E. Larson, pp. 989-98. Proceedings of the Third International Conference on Precision Agriculture, ASA-CSSA-SSSA, Madison, WI, 1996.
- Solohub, M. P., C. van Kessel, and D. J. Pennock. "The Feasibility of Variable Rate N Fertilization in Saskatchewan" In *Precision Agriculture*, eds., P. C. Robert, R. H. Rust, and W. E. Larson, pp. 65-73. Proceedings of the Third International Conference on Precision Agriculture, ASA-CSSA-SSSA, Madison, WI, 1996.
- Wollenhaupt, N. C., D. J. Mulla, and C. A. Gotway Crawford. "Soil Sampling and Interpolation Techniques for Mapping Spatial Variability of Soil Properties." In *The State of Site Specific Management for Agriculture*, ed. F. J. Pierce and E. J. Sadler, pp. 19-55, ASA-CSSA-SSSA, Madison, WI, 1997.