

ESTIMATION OF SOIL COMPOSITION IN SHRIMP PONDS

G. Ritvo, M. Sherman, A.L. Lawrence, T.M. Samocha*
Texas A&M University

Summary:

We estimate the average amount of phosphorus in the soil of four Texas shrimp ponds. Broad spatial trends in phosphorus levels are observed and the trend is qualitatively common to all four ponds. Under the assumption that the estimated trend is the true mean surface we calculate the bias and variance of the estimated mean. This was done for three different sampling designs, and several sampling rates for each design. Using spatial statistical techniques the diminishing utility of increased sampling is demonstrated, and the design with the lowest bias is identified. The similarities between the four ponds suggest that the results are applicable to similar ponds in Texas, and elsewhere throughout the world.

Keywords: aquaculture, efficiency, shrimp culture, soil sampling, spatial statistics, semivariogram.

* G. Ritvo is Graduate Research Assistant in the Dept. of Soil & Crop Sciences at Texas A&M University, M. Sherman is Assistant Professor in the Dept. of Statistics at Texas A&M University, A.L. Lawrence is Professor and Project leader at Texas Agricultural Experiment Station, and T.M. Samocha is Associate Professor at Texas Agricultural Experiment Station, Texas A&M University-Corpus Christi.

1: INTRODUCTION

The purpose of soil sampling is to obtain information about a particular soil of interest. One main goal is to determine parameters of interest like the total or average amount of various nutrients in the soil over a given region. The purpose of sampling is to estimate such parameters to a certain degree of accuracy at the lowest possible cost (Petersen and Calvin, 1986). If the observed levels are relatively constant across locations then a very small number of samples can accurately determine the parameter of interest. If, however, the samples differ according to spatial trends and are locally very variable then a large number of samples will be needed. A natural goal is to determine a sampling scheme that effectively balances the tradeoff between the accuracy of the point estimator and the cost of obtaining the samples in a specific context. There is, however, no universal tradeoff between accuracy in estimation and cost of sampling. For this reason we settle on the more modest goal of studying the increased precision and accuracy obtained by more intensive sampling.

In soil research the locations for sampling are chosen according to various sample schemes, e.g., simple random sample, stratified random sample or systematic sample methods. In our experiments we used a systematic sample in all cases. The main reason for this was for the ease of sampling. An additional benefit is that systematic sampling is optimal (in a specific sense) by the results of Bellhouse (1977).

A difficulty which occurs in assessing the accuracy of point estimators in the spatial setting is that sample measurements tend to be positively spatially correlated over the field (Warrick, Myers, and Nilsen, 1986). This means that two measurements in a field which are obtained at locations separated by short distances tend to be more similar than two measurements that are separated by greater distances. This correlation needs to be accounted for because it affects the accuracy of our point estimate of the quantity of interest. Semivariogram techniques (see, e.g., Cressie, 1993) help to account for this correlation.

Currently there is relatively little research on the relationship between pond soil chemical composition and the growth of shrimp or other aquatic organisms. Boyd (1976) found that soil phosphorus levels were relatively higher where phosphorus fertilizer was applied, and that the concentrations of organic matter, organic nitrogen, and the soil cation exchange capacity increased with increasing water depth. Boyd (1995) suggested to take samples along several transects from shallow to deep water. He also suggested estimating the average concentration of a pond soil variable from samples obtained via sampling randomly along an “s-shaped” pattern. Another sampling method was employed by Brune and Eversole (1993) who obtained soil samples by dividing each pond into 6 equal areas and sampling at the center point of each.

The above studies suggest sampling procedures, but give little evidence of their efficacy or guidance into how to choose a sampling design. Further, for a particular design they don't address the question of what sampling rates are necessary to obtain a certain degree of accuracy for point

estimators.

The objectives of this study are: 1) to analyze the soil concentration of phosphorus for trends over a pond, and 2) to develop an effective sampling design with a reasonable sampling rate for ponds used in shrimp production. This will be accomplished by determining the accuracy of point estimators of parameters for different sampling designs and sampling rates in a systematic fashion.

2: METHODS OF DATA COLLECTION AND STATISTICAL SETUP

2.1: Methods of Data Collection

Soil was removed from four $300m \times 100m$ ponds in a Texas shrimp farm. This is a common pond size in Texas and occurs elsewhere throughout the world, e.g., China (Wildman et al., 1992). A total of 109 samples were collected from the top $10cm$ of the soil from each pond by a core sampler with a $2.5cm$ diameter. These samples were obtained as follows: a grid shape of 75 samples was removed from each pond where 15 equidistant samples were removed from the long edge ($300m$), and five equidistant samples from the short edge ($100m$). Additionally, 17 samples were taken from each of the shallow and deep edges of each pond. These 17 samples were distributed in a cross shape. Each of the four sides of the cross consists of samples distributed in distances of 1, 5, 10, and $15m$ from the center. These latter observations turn out to be valuable for accurate semivariogram estimation. See Figure 1 for a layout of the sampling design (not to scale).

Each sample was analyzed for its phosphorus concentration in triplicate. We treat the mean of these three measurements as the observed measurement. The sample variance of the three measurements at each location aid in verifying the “nugget effect” in the semivariogram (see Section 3.2). Phosphorus was extracted with extraction solution of $0.025\text{ M H}_4\text{EDTA}$ in 1.4 M ammonium acetate and 1.0 M HCl , adjusted to a pH between 4.2 and 4.3 (Hons, Larson-Vollmer, and Locke, 1990). Phosphorus concentrations were determined via use of an ICP analyzer. The (logged) phosphorus levels (in p.p.m.) for each pond are shown in Figure 2. The measurements exhibit apparent trends but are noisy. Our goals are to estimate this trend for each pond and to accurately estimate the average phosphorus level in each pond.

2.2: Statistical Setup

Let $Z(s)$ be the measured value of (log) phosphorus at location $s \in D$, where D denotes a $300m \times 100m$ pond. We analyze the logarithm of phosphorus because these measurements are approximately normally distributed.

We assume that for each location s

$$Z(s) = \mu(s) + \delta(s),$$

where $\mu(s)$ is the true phosphorus level and $\delta(s)$ is a random error with $E(\delta(s)) = 0$ and $Var(\delta(s)) =$

$\sigma^2(s)$. Our goal is to estimate $\mu(s)$ ($s \in D$), the mean function, and the summary quantity

$$\mu_D := (1/|D|) \int_D \mu(s) ds,$$

the average amount of phosphorus in each pond. We analyze each of the four ponds separately, for this will enable us to generalize conclusions reached to a larger population of similar ponds.

3: MODEL FITTING AND DESIGN EFFICACY

3.1 Fitting the Mean Function

There are several ways to fit a mean function to the observed phosphorus levels. One common method is to assume that the mean at each location has the ANOVA type decomposition of a sum of effects for an overall mean, a column effect, and a row effect. In this example this requires estimating $1+5+15=21$ parameters. An alternative that uses median polishing (see, e.g., Tukey, 1977) also requires estimation of 21 parameters (actually in both cases only 19 parameters are estimable). We believe that the distribution of phosphorus levels is a relatively smooth function of spatial location, and for this reason fit a polynomial trend of degree r in the spatial coordinates, $s = (x, y)$, of the form:

$$\mu(s) = \sum_{0 \leq k+l \leq r} \alpha_{kl} x^k y^l, \quad s = (x, y)$$

(Ripley, 1981, Chapter 4).

For each pond we consider the model with degree $r = 2$ (with additional terms x^3 and y^3 for a total of 8 parameters). To facilitate comparisons between ponds we choose the one model that best fits the common phenomena across ponds. Note that the above model is formally a general linear model, and can be fit using most statistical packages. The models were fit using the S-Plus function $lm(\bullet)$. In all four ponds both cubic terms were insignificant as was the cross-product ($x*y$) term. The resulting parameter estimates in this 5 parameter model and other relevant information is given in Table 1.

We note that the standard errors (and thus the t-values and p-values) may not, in principle, be reliable due to potential correlation between measurements. It turns out, however, that all sampling locations are separated by distances of at least $15m$ because we have used only the 75 equidistant observations (excluding the observations from the crosses) in fitting the mean surface. In the next subsection we see that there is very little, if any, correlation at distances larger than $15m$.

In three of the four ponds we see that the linear and quadratic terms in each direction (x, y) are strong predictors of phosphorus levels. The most significant terms are the y and y^2 terms with x and x^2 terms generally less important predictors. In the four ponds these four predictors explain

between 20 and 35 percent of the variation in phosphorus levels. This is a reasonable amount of explained variation given the noisiness of the observations exhibited in Figure 2. The fitted mean surface, $\hat{\mu}(s)$, for each pond is given in Figure 3. Overall, there is a remarkably good agreement in the results across ponds. The fitted surfaces suggest that elevated levels of phosphorus occur in the center of the pond in the “ y direction”; with high accumulations in both the shallow and deep ends (larger than $250m$ and less than $50m$ in the x direction, respectively) with the highest concentrations in the deep end.

A possible explanation for the observed accumulations is the position of aerators in the ponds. An aerator is a mechanical device that increases the level of dissolved oxygen in the water. In these ponds four aerators are located along each long side (x -axis) of the pond. The paddle wheels are evenly spaced, starting at $10m$ from each edge. The accumulation in the deep ($x = 300m$) and shallow ($x = 0m$) ends of the pond may be a result of the reduced water current where no aerator is present. The high accumulation in the center (along the y -axis) can be explained in the same way. This hypothesis is supported by Smith (1996) who found that in shrimp ponds with size ranging between 2.5 and 2.9 hectares, nutrients such as carbon and phosphorus accumulated in the center of the ponds and near the drain; and attributed it to the currents generated by paddle wheel aerators. Elsewhere, Boyd (1992) suggested that sedimentation occurs in the center of a rectangular intensive shrimp pond as a result of the currents generated by aerators that are positioned in each corner of the pond.

3.2 Estimation of the Average Amount of Phosphorus, μ_D

Let (s_1, \dots, s_k) be a set of k candidate design points. The natural estimator of the average amount of phosphorus (and the only one we consider) is the usual sample mean: $\sum_{i=1}^k Z(s_i)/k$. To facilitate comparisons between the sampling actually done, and alternative sampling rates, we consider only equally spaced designs of the form

$$D_n := \{s : s = (x, y), \text{ where } x = (300 * i)/(3n + 1), y = (100 * j)/(n + 1), i = 1, \dots, 3n, j = 1, \dots, n\},$$

and thus the cardinality of D_n is $N := 3n^2$. Note that this design does not include points on the boundary. We consider the efficacy of an alternative scheme which includes the boundary of each pond in Section 4.2.

Let \bar{Z}_n denote the sample mean of observations $s \in D_n$. If observations are independent and identically distributed then $\mu_{D_n} := E(\bar{Z}_n) = E(Z) = \mu_D$, $Var(Z(s_i)) = \sigma^2$, and thus $Bias(D_n) := E(\bar{Z}_n) - \mu_D = 0$ and $Var(D_n) = \sigma^2/N$. We see the familiar result that the sample mean is unbiased for the population mean and that the variance decreases proportionally to the sample size N . For these data, however, due to the nonconstant mean determined in Section 3.1, the estimator \bar{Z}_n has a non zero bias for any sample size N . Further, due to correlations amongst the errors, $\delta(s)$, the

variance of \bar{Z}_n will not necessarily decrease proportionally in N . Our results show the diminishing utility of increased sampling rates in the spatial setting. See, e.g., Hoel (1961) for related results in the one dimensional setting.

To compute biases and variances of estimators we need to know the true value of μ_D as well as other unknown quantities. For this reason we proceed as follows. Assume that the estimated mean surface, $\hat{\mu}(s)$, is the true mean surface $\mu(s)$. Under this model we have $\mu_D := (1/|D|) \int_D \hat{\mu}(s) ds$ is the target value, and define the bias to be $Bias(D_n) := \mu_{D_n} - \mu_D$. Further, under this model, for any two locations, s and t , we have $Cov(Z(s), Z(t)) = Cov(\hat{\delta}(s), \hat{\delta}(t))$, where $\hat{\delta}(s) := Z(s) - \hat{\mu}(s)$, the residual at location $s \in D$. In the sequel we will simply write $\hat{\delta}(s) = \delta(s)$ because the residuals are the true errors under our model.

Letting $\sigma(s, t) = Cov(\delta(s), \delta(t))$ we have that

$$Var(D_n) = \left(\sum_{s \in D_n} \sigma^2(s) + \sum_{(s,t) \in D_n, s \neq t} \sigma(s, t) \right) / N^2.$$

In analyzing the triplicate measurements at each location there doesn't appear to be a trend in the variability and for this reason we assume that $\sigma^2(s) =: \sigma^2$. The semivariogram, $\gamma(h)$, is defined to be the function $(1/2)Var(\delta(s) - \delta(s+h))$. We assume that $\sigma(s, t) = \sigma(0, (t-s)) =: \sigma(t-s)$, i.e., that the errors are intrinsically stationary and that $\sigma(0, (t-s)) = \sigma(0, d(s, t))$, i.e., the errors are isotropic (these assumptions are reasonable for these data, as seen by viewing directional semivariograms for each pond). The basic parameters in the semivariogram are the value of $\sigma^2 = \lim_{h \rightarrow \infty} \gamma(h)$ and the *nugget* $:= \lim_{h \rightarrow 0} \gamma(h)$. The nugget effect is due to a combination of measurement error and microscale variation. This will be nonzero in our setting. Non-zero measurement error is usually unverifiable in practice, but we have seen through triplicate measurements at each location that we obtain different phosphorus readings at the same location.

The estimated semivariogram is obtained from the natural method of moments estimator

$$\hat{\gamma}(h) := |2N(h)|^{-1} \sum_{N(h)} (\delta(s_i) - \delta(s_j))^2,$$

where $N(h) := \{(s_i, s_j) : d(s_i, s_j) = h\}$ for 21 different (grouped) distances h ranging from 1.2m to 130m. The estimated semivariograms for each pond are given in Figure 4. We see (for ponds 17, 25 and 26) that the semivariograms increase until a distance of about 10 – 20m and then flatten out (the slight decreases for long distances may be due to small biases induced by using $\hat{\delta}(s)$ in place of the unknown $\delta(s)$, as discussed in Cressie (1993), section 3.4.3). We are now including all 109 observations, including the 17 on each cross in Figure 1, for the purpose of estimating the semivariogram. In fact, without these observations, we would obtain an approximately flat semivariogram, which would render the incorrect conclusion that there is no correlation amongst the errors. This would drastically overestimate the utility of increased sampling rates.

We fit the “exponential model” to the estimated semivariogram

$$\hat{\gamma}(h) := \text{nugget} + (\sigma^2 - \text{nugget})(1 - \exp(-ch)), \quad (1)$$

where $c > 0$ is the decay parameter that determines the distances at which correlations of a given strength are present. $c = \infty$ corresponds to independence and c close to zero corresponds to “long range” dependence. For more details on the semivariogram, the estimated semivariogram, and semivariogram models, see, e.g., Isaaks and Srivastava (1989) or Cressie (1993). Equation 1 implies that the estimated covariance function is

$$\hat{\sigma}(h) = I\{h = 0\}(\text{nugget}) + (\sigma^2 - \text{nugget})\exp(-ch).$$

The models were fit to the estimated semivariograms using the S-Plus spatial module. Other software that does this is Geopack, available from the Environmental Protection Agency (1990). Estimated values of the *nugget*, σ^2 , and c parameters were obtained using a nonlinear regression routine in S-Plus. The parameter values obtained in the fitted semivariogram for the four ponds are given in Table 2. In the next section we use these semivariograms to estimate the variability in the estimated phosphorus means.

4: COMPARISON OF SAMPLING RATES AND SCHEMES

4.1 Comparison of Sampling Rates

To compare different sampling rates we compare their biases and variances defined as in Section 3.2 equations 1 and 2, and finally through Mean Square Errors (MSE’s) using the usual decomposition: $MSE(D_n) = Bias^2(D_n) + Var(D_n)$. Table 3 gives the biases, variances, and the MSE, for each sampling rate from $n = 2$ to $n = 10$ ($N = 12$ to $N = 300$). We choose this range of sampling rates because these are considered reasonable in terms of time and resources (actually $N \leq 108$ is close to the limit of feasibility). We see that in all cases the estimates overestimate the true average phosphorus and that the bias in estimating the average phosphorus decreases close to linearly in the sampling rate n (and not in the sample size N). If the observations were independent, then we could expect to see the variances decrease in the sample size N . Due to the correlation detected in Section 3, however, we do not expect this to be the case. In particular, note that for Pond 1 when we increase the sample size from $N = 75$ to $N = 300$ we reduce the variability by a factor of only 1.22, much smaller than the factor of 4 we would expect if the data were independent. Thus, in terms of variability, there is very little reason to sample beyond $N = 75$. Bias is a bigger problem than variance in all four ponds, squared bias accounting for over 50% of MSE in all cases. When the bias and variance are combined into MSE there is more benefit to increased sampling than when we simply consider variability.

4.2 Comparison of Sampling Schemes

Recall that the analysis carried out in Section 4.1 assumed that the design was of the form:

$$D_n := \{s : s = (x, y), \text{ where } x = (300 * i)/(3n + 1), y = (100 * j)/(n + 1), i = 1, \dots, 3n, j = 1, \dots, n\}.$$

We have noted previously that the edges are not included in this design, so we now consider the efficacy of including the edges. Clearly, for dense sampling rates, n , the differences will be relatively small, but we consider only modest, practical sampling rates for soil sampling, $n = 2, \dots, 10$, and in this range the differences could be appreciable. Formally, we now consider the design which includes the edges of the pond, $D_n :=$

$$\{s = (x, y), \text{ where } x = (300 * (i - 1))/(3n - 1), y = (100 * (j - 1))/(n - 1), i = 1, \dots, 3n, j = 1, \dots, n\}.$$

Intuitively, for each fixed sampling rate, n , this design should lead to a smaller $Var(D_n)$ than the design which excludes the edges. The reason is that in this new design, points are generally further separated, leading to smaller covariance terms due to spatial correlation. It is not clear how the biases of the two schemes will compare. Table 4 gives the bias, variance, and MSE for this sampling scheme. We see, as expected, that for each fixed sampling rate the variances are smaller under this new design. We also see, however, that the biases increase in magnitude under this new design, and that the increase in squared bias more than offsets the decrease in variability for $N \leq 147$. These results are uniform across the four ponds, and thus we prefer the design which excludes the edges. In all cases the design with edges underestimates the true amount of phosphorus (in contrast with the previous design). Considering variability, we see that for Pond 1 the variability actually increases when the sampling rate increases from $n = 9$ to $n = 10$ ($N=243$ to $N = 300$). This may seem counter intuitive, but the additional number of covariance terms and the increase in their magnitude more than offset the increased sampling rate. Thus, for data from a stationary model (in particular, a constant mean), increasing the sampling rate leads to a decrease in efficiency. This surprising result corroborates that of Morris and Ebey (1984) in the one dimensional setting. This phenomenon does not occur for the design which excludes the edges.

Recall that the main goal is to estimate μ_D accurately. This will typically be done through a composite sample, i.e., pooling all the samples and obtaining one overall estimate of the mean (without observing the individual values). It seems that a significant, systematic bias is somewhat problematic. The design without edges has a systematic positive bias and the design with edges has a systematic negative bias. This suggests that a compromise design may give closer to unbiased estimates. Consider the equally spaced design of the form

$$\{s = (x, y), \text{ where } x = (300 * (i - .5))/(3n), y = (100 * (j - .5))/n, i = 1, \dots, 3n, j = 1, \dots, n\}.$$

Table 5 gives the biases using this design. We see that the biases are generally smaller than either of the previous two designs considered, especially for small N , which are most likely to be employed in practice. The variances are smaller than the design without edges, and thus we can recommend this compromise design.

5: CONCLUSIONS

We have addressed the spatial distribution of phosphorus in four shrimp ponds in Texas. We have seen that the estimated surface is far from constant, having both significant linear and quadratic trends. This shows that it is nontrivial to estimate summary quantities such as the mean phosphorus level over a pond. We have evaluated the efficiency of the sample mean of the phosphorus measurements under different sampling rates. Our results show the diminishing utility of increased sampling rates due to spatial correlations.

We also compared the efficacy of three different sampling designs. The first design excludes the edges of the pond and the second includes the edges. For reasonable sample sizes, e.g., $N \leq 108$, we've seen that the design excluding edges is superior. Further, the comparison between the two designs suggested a compromise design that has lower bias than either design for small to moderate sample sizes. All results and observations, in particular those on the trends in the estimated surface, the diminishing utility of increased sampling rates, and the superiority of the compromise design which excluded edges were relatively constant across the four ponds. This is relevant because the size of the ponds is approximately the same as many ponds used in Texas and that occur elsewhere throughout the world. Thus, there is reason to believe that the results are generalizable to this larger group of ponds. Further, the techniques and results may well be generalizable to nutrients other than phosphorus.

REFERENCES

- Bellhouse, D.R. (1977). Some optimal designs for sampling in two dimensions, *Biometrika*, 64, 605-611.
- Boyd, C.E. (1976). Chemical and textural properties of muds from different depths in ponds, *Hydrobiologia*, 48, 141-144.
- Boyd, C.E. (1992). Shrimp pond bottom soil and sediment management. In Wyban, G. Ed., *Proceedings of the Special Session on Shrimp Farming*, World Aquaculture Society, Baton Rouge, LA, USA.
- Boyd, C.E. (1995). *Bottom soils, sediments, and pond aquaculture*, Chapman & Hall, New York.
- Brune, D.E., and A.G. Eversole. (1993). Pond nutrient dynamics in crawfish culture ponds. In *Techniques for modern aquaculture*, American Society of Agriculture Engineers, St. Joseph,

MI, USA. pp. 533-542.

Cressie, N.A.C. (1993). *Statistics for spatial data*, Wiley, New York.

Geopack Version 1.0.e-Jan (1990). Robert S. Kerr Environmental Research Laboratory Office of Research and Development, *U.S. Environmental Protection Agency*, Ada, OK.

Hoel, P.G. (1961). Asymptotic efficiency in polynomial estimation, *Annals of Mathematical Statistics*, 32, 1042-1047.

Hons, F.M., Larson-Vollmer, L.A. and Locke, M.A. (1990). NH_4OAc -EDTA-extractable phosphorus as a soil test procedure, *Soil Science*, 149, 249-256.

Isaaks, E.H. and Srivastava, R.M. (1989). *An introduction to applied geostatistics*, Oxford University Press, Oxford.

Morris, M.D. and Ebey, S.F. (1984). An interesting property of the sample mean under a first order autoregressive model, *The American Statistician*, 38, 127-129.

Petersen, R.G. and Calvin, L.D. (1986). Sampling. In Kulte, A. (ed), *Methods of soil analysis, Part 1-Physical and mineralogical methods*, Agronomy, 2nd ed., No. 9, Part 1, 33-51.

Ripley, B.D. (1981). *Spatial statistics*, Wiley, New York.

Smith, P.T. (1996). Physical and chemical characteristics of sediments from prawn farms and mangrove habitat on the Clarence river, Australia, *Aquaculture*, 146, 47-83.

Tukey, J.W. (1977). *Exploratory data analysis*, Addison-Wesley, Reading, MA.

Warrick, A.W., D.E. Myers, and D.R. Nilsen. (1986). Geostatistical methods applied to soil science. In Kulte, A. (ed), *Methods of soil analysis, Part 1-Physical and mineralogical methods*, Agronomy, 2nd ed., No. 9, Part 1, 53-82.

Wildman, M., P. Niemeier, F. Nielsen, J. Beverly, T. Schneider, H. Riha, E. Sanborn, D. Weinder, D. Decker, and B. Rosenberry. (1992). *World shrimp culture*, Vol. 1: Africa, Asia, Europe, Middle East, North America. NOAA Tec. Memo. NMFS-F/SPO-4.

TABLES AND FIGURES

Table 1: Parameter Estimates for Fitted Mean Surfaces

Table 2: Fitted Semivariogram Parameters for Each Pond

Table 3: Bias, Variance and MSE for each sampling rate

Table 4: Bias, Variance and MSE for Edges Design

Table 5: Biases From “Compromise Design”

Figure 1: Sampling Location for Phosphorus Measurements

Figure 2: Observed Phosphorus Levels

Figure 3: Fitted Mean Surface to Phosphorus

Figure 4: Estimated Semivariograms of Residuals from Fitted Mean Surfaces

Table 1: Parameter Estimates for Fitted Mean Surfaces

Pond 1				
	Value	Std. Error	t value	Pr(> t)
y	0.1390	0.0274	5.0704	0.0001
x	-0.0125	0.0047	-2.6289	0.0105
y^2	-0.0013	0.0003	-4.7408	0.0001
x^2	0.00003	0.00001	2.6851	0.0090
Pond 2				
	Value	Std. Error	t value	Pr(> t)
y	0.1096	0.0284	3.8555	0.0003
x	-0.0129	0.0049	-2.6329	0.0104
y^2	-0.0011	0.0003	-4.0621	0.0001
x^2	0.00003	0.00001	2.7090	0.0085
Pond 3				
	Value	Std. Error	t value	Pr(> t)
y	0.0796	0.0266	2.9935	0.0038
x	-0.0019	0.0046	-0.4222	0.6742
y^2	-0.0008	0.0003	-3.0642	0.0031
x^2	0.00001	0.00001	1.1017	0.2744
Pond 4				
	Value	Std. Error	t value	Pr(> t)
y	0.0885	0.0251	3.5308	0.0007
x	-0.0110	0.0043	-2.5414	0.0133
y^2	-0.0009	0.0002	-3.5312	0.0007
x^2	0.00003	0.00001	2.9707	0.0041

Table 2: Fitted Semivariogram Parameters for Each Pond

Pond	Nugget	σ^2	c
1	.617	.891	.017
2	.221	.767	.133
3	.0829	.658	.312
4	.597	1.27	.087

Table 3: Bias, Variance and MSE for each sampling rate

Pond 1, $\mu = 2.9799$			
N	Bias	Variance	MSE
12	.6421	.1179	.5302
27	.4909	.08080	.3218
48	.3984	.06680	.2255
75	.3362	.05984	.1729
108	.2915	.05579	.1408
147	.2579	.05319	.1197
192	.2316	.05141	.1051
243	.2105	.05009	.0944
300	.1933	.04912	.0865
Pond 2, $\mu = 2.9635$			
N	Bias	Variance	MSE
12	.5560	.06551	.3747
27	.4258	.03121	.2125
48	.3460	.01966	.1394
75	.2922	.01453	.0999
108	.2534	.01182	.0760
147	.2243	.01024	.0610
192	.2015	.00922	.0500
243	.1831	.00854	.0421
300	.1682	.00805	.0363
Pond 3, $\mu = 3.0556$			
N	Bias	Variance	MSE
12	.4233	.05484	.2340
27	.3224	.02434	.1283
48	.2615	.01379	.0822
75	.2201	.00894	.0574
108	.1907	.00635	.0427
147	.1686	.00482	.0332
192	.1513	.00386	.0268
243	.1375	.00323	.0221
300	.1262	.00278	.0187

Pond 4, $\mu = 2.8423$

N	Bias	Variance	MSE
12	.4077	.1161	.2823
27	.3133	.06010	.1583
48	.2551	.04098	.1061
75	.2156	.03223	.0787
108	.1872	.02749	.0625
147	.1657	.02461	.0521
192	.1490	.02272	.0449
243	.1355	.02140	.0398
300	.1244	.02044	.0359

Table 4: Bias, Variance and MSE for “Edges” DesignPond 1, $\mu = 2.9799$

N	Bias	Variance	MSE
12	-1.9904	.07642	4.038
27	-.9702	.05149	.9928
48	-.6357	.04455	.4487
75	-.4700	.04120	.2621
108	-.3711	.04087	.1786
147	-.3054	.04040	.1337
192	-.2586	.04020	.1071
243	-.2236	.04014	.0901
300	-.1965	.04015	.0788

Pond 2, $\mu = 2.9635$

N	Bias	Variance	MSE
12	-1.7485	.06392	3.1212
27	-.8497	.02849	.7505
48	-.5560	.01650	.3256
75	-.4106	.01138	.1800
108	-.3241	.00890	.1139
147	-.2667	.00756	.0787
192	-.2257	.00680	.0577
243	-.1952	.00633	.0444
300	-.1714	.00603	.0354

Pond 3, $\mu = 3.0556$

N	Bias	Variance	MSE
12	-1.2757	.05483	1.6822
27	-.6255	.02437	.4156
48	-.4109	.01371	.1826
75	-.3043	.00879	.1014
108	-.2405	.00613	.0640
147	-.1981	.00456	.0438
192	-.1678	.00358	.0317
243	-.1452	.00292	.0240
300	-.1276	.00248	.0188

Pond 4, $\mu = 2.8423$

N	Bias	Variance	MSE
12	-1.3159	.10586	1.8375
27	-.6361	.04818	.4528
48	-.4152	.02990	.2023
75	-.3063	.02242	.1162
108	-.2415	.01904	.0774
147	-.1985	.01720	.0566
192	-.1680	.01615	.0444
243	-.1452	.01552	.0366
300	-.1275	.01512	.0314

Table 5: Biases for “Compromise” Design

N	Pond 1	Pond 2	Pond 3	Pond 4
12	0.2444	0.2059	-0.0537	0.0272
27	0.1810	0.1663	-0.1302	-0.0041
48	0.1799	0.1775	-0.1478	0.0033
75	0.1916	0.1980	-0.1493	0.0176
108	0.2060	0.2200	-0.1449	0.0328
147	0.2204	0.2413	-0.1382	0.0474
192	0.2339	0.2614	-0.1305	0.0610
243	0.2463	0.2804	-0.1225	0.0736
300	0.1518	0.1416	-0.1149	-0.015