

University of Alberta

**A Study on Determining the Sample
Size in Geostatistics**

by

Ying Ming Or

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science
in
Statistics

Department of Mathematical and Statistical Sciences
Edmonton, Alberta, Canada

Fall 2010

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only. Where the thesis is converted to, or otherwise made available in digital form, the University of Alberta will advise potential users of the thesis of these terms.

The author reserves all other publications and other rights in association with the copyright in the thesis, and except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatsoever without the author's prior written permission.

Examining Committee

Dr. Edit Gombay, Statistics, University of Alberta

Dr. Ivan Mizera, Statistics, University of Alberta

Dr. Csaba Szepesvari, Computing Science, University of Alberta

Abstract

After the scientific problem of interest is defined, collecting data is the first stage of any statistical analyses. The question of how large the sample should be is thus of great interest. In this thesis we demonstrate that in a geostatistical experiment determining the minimum sample size to achieve a certain precision of an estimator is often not possible due to inconsistencies of the estimators. This thesis is an empirical work extended from a manuscript (Gombay, 2010), which shows that the laws of large numbers may not hold under the spatial setting. It is demonstrated by a simulation study that the variance of the kriged mean converges to a non-zero constant as the sample size keeps increasing. It then followed by further investigations on the simple and ordinary kriging estimators. The conclusions arrived in this thesis lead for further research on the topic.

Acknowledgements

I owe my deepest gratitude to my supervisor Dr. Edit Gombay for her patience, kindness and generous support of my Master study and research. Without her continuous encouragement and guidance, I believe I would not have been able to complete the writing of this thesis. Her influence on me, not only academically but also personally, is way beyond where words can describe. I could not have imagined having a better advisor and mentor in my life.

Besides my supervisor, I would also like to thank the other members of my examination committee, Dr. Ivan Mizera and Dr. Csaba Szepesvari, who provide constructive feedbacks and encouraging remarks which inspire me to investigate the topic from different angles.

My sincere thanks also goes to the professors in the Department of Mathematical and Statistical Sciences at the University of Alberta, who fed me with numerous statistical knowledge in the graduate courses; and radically altered my ways of thinking.

I am also grateful to my parents who supports my decisions all the time; and especially to my grandmother, for her understandings and unselfishness of letting me to pursue my dream.

Contents

1	Problem	1
2	Review of Literature	3
2.1	Model	3
2.2	Kriging of the Mean	5
2.3	Simple Kriging	9
2.4	Ordinary Kriging	12
2.5	Tools in Spatial Data Analysis	19
3	Performance of the Estimators	28
3.1	Convergence of Sample Mean and Kriging Estimators: The Necessary and Sufficient Condition	28
3.2	Comparing the sample mean and the kriged mean	34
4	Simulation Studies	36
4.1	Generating a Dataset	36
4.2	Kriging of the Mean	37
4.3	Variances of the Kriged Mean	40
4.4	Properties of the Estimators in Kriging Procedures: Simple Kriging and Ordinary Kriging	44
4.4.1	Validating the functions	44
4.4.2	Variances of the Estimators	49
4.5	Random Sampling Designs Vs. Non-Random Sampling Designs: Comparison of the Sample Mean and the Kriged Mean	52
4.6	Example 1 - Dioxin Data Analysis	60
4.6.1	Background and Preliminary Analysis	60

4.6.2	Variogram Modeling and Kriging	63
4.6.3	Model Diagnostics	67
4.7	Example 2 - House Prices Data Analysis	69
4.7.1	Introduction and Preliminary Analysis	69
4.7.2	Variogram Modeling and Spatial Predictions	73
5	Conclusions	79
A	R code	81
A.1	The Drawing of the Theoretical Variogram	81
A.2	The Drawing of the Range Anisotropic Directional Variogram	81
A.3	The Drawing of the Sill Anisotropic Directional Variogram . .	82
A.4	The Drawing of the Nugget Anisotropic Directional Variogram	82
A.5	Calculation of the Theoretical Variance	83
A.6	Dataset Generation	84
A.7	Kriging of the Mean and its Empirical Variance	85
A.8	Validating the function for Simple Kriging in R	88
A.9	Validating the function for Ordinary Kriging in R	91
A.10	Empirical Variances of Simple Kriging Estimator	94
A.11	Empirical Variances of Ordinary Kriging Estimator	96
A.12	Comparison of the Sample Mean and the Kriged Mean Under the Random and Non-Random Design.	98
A.13	Dioxin Data Analysis	102
A.14	Housing Prices Data Analysis	105
B	References	108

List of Tables

1	Empirical Variances of the Kriged Mean	43
2	Ordinary Kriging Prediction Weights	49
3	Empirical Variances of the Simple and Ordinary Kriging Estimator	50
4	Random Design: Comparison of the Sample Mean and the Kriged Mean	57
5	Non-Random Design: Comparison of the Sample Mean and the Kriged Mean	57

List of Figures

1	Theoretical Variogram	23
2	Range Anisotropic Directional Variogram	25
3	Sill Anisotropic Directional Variogram	26
4	Nugget Anisotropic Directional Variogram	27
5	Kriging of the Mean Estimator Variances	43
6	Visualization of Ordinary Kriging Prediction Weights	48
7	Simple and Ordinary Kriging Estimators Variances	51
8	Random and Lattice Sampling Designs	55
9	Random Design: Comparison Plot for Sample Mean and Kriged Mean	58
10	Non-Random Design: Comparison Plot for Sample Mean and Kriged Mean	59
11	Lagged Scatterplot of Dioxin Data	61
12	Lagged Scatterplot of Log-Transformed Dioxin Data	62
13	Lagged Scatterplot of Log-Transformed Dioxin Data	63
14	Variogram Cloud of Dioxin Data	65
15	Experimental and Theoretical Variogram of Dioxin Data	66
16	Ordinary Kriging Predictions of Dioxin Data	66
17	Directional Variogram of Dioxin Data	68
18	Preliminary Plots of Spatial Data	71
19	Variogram Cloud	72
20	Directional Variogram in untransformed location coordinates of Housing Prices Data	73

21	Directional Variogram in transformed location coordinates of Housing Prices Data	74
22	Experimental Variogram in isotropic space of Housing Prices Data	75
23	Contour Plot	77
24	3-D Plot of Predictions	78

1 Problem

The problem of finding the necessary sample size for a randomized experiment and an observational study is very common and practical. Simplifying assumptions are made on the random variable which are often that they are independent and identically distributed (i.i.d.). By specifying a desired bound of error, the number of observations needed for a particular randomized experiment or an observational study can be easily determined.

However, this i.i.d. assumption is not often applicable to the real world. Examples include survey sampling where observations are considered to be correlated when subjects are sampled using a without replacement scheme; or when dealing with temporal data in time series analysis. Similarly, this is the case in the field of geostatistics, where a continuous spatial region is of interest. The task of computing the required sample size then becomes a difficult one. The goal of this thesis is to investigate this simple, but important problem. For instance, the most widely used estimator for estimating the mean in the spatial region, namely the *kriged mean* (definition follows in the next section), will be examined. While looking for a solution of this problem, one may spontaneously ask the following questions:

- Is the kriged mean consistent?
- Is the kriged mean more efficient than the sample mean?

These questions arise naturally when one is looking to find the minimum sample size, and thus they will be considered in detail. It is surprising that there are not many comments in the literature on the properties of these estimators that are often used in geostatistics. Most textbooks and journal articles are only concerned with whether these estimators are unbiased or not, but

somehow they do not investigate their consistency. On the other hand, much research work was done on how to efficiently estimate covariance parameters. This is, indeed, a very important topic in modelling the dependence structure of the spatial region accurately. However, in applications greater attention are placed on the predicted values and thus it is worthwhile to develop better estimators, that are good not only according to the minimum mean-squared error criterion. Modelling the dependence structure is the first step, and this thesis is an intermediate step for future research to be done on finding other estimators better fit for users' purposes, so that more choices are available for different practical situations.

The second phase of this thesis is to extend the empirical study to the two prediction methods often used in geostatistics, *simple kriging* and *ordinary kriging* (definitions follow in the next section). Again, the consistency of the estimators used in the two methods are examined. The last part of this thesis is to see whether the conclusions drawn from the earlier questions are consistent with an application example. A thorough analysis will be included.

2 Review of Literature

2.1 Model

Suppose, that the characteristic in a spatial region of interest is a random variable $S(X)$, associated with each location X . Then, the infinite collection of these random variables is called a random function of the domain. Note that the domain is fixed and it does not change even if the number of measurements increases. We assume that $S(X)$, which will be referred as the underlying signal from now on, is continuous. Further, this underlying signal is assumed to have finite mean and variance. The measurements that are made at locations $X_i, i = 1, 2, \dots, n$, denoted by y_i , are realizations of the value of a random variable Y_i . It is important to mention that Y_i is dependent on the value of $S(X_i)$ at the location X_i ; and measurement error should also be incorporated in the model. With all the above assumptions, a model is,

$$Y_i = S(X_i) + Z_i, \quad i = 1, 2, \dots, n, \quad (1)$$

where $Z_i \sim N(0, \tau^2)$ is the measurement error. Clearly, the random variable Y has conditional mean, $E[Y_i|S(X_i)] = S(X_i)$ and conditional variance, $\text{VAR}[Y_i|S(X_i)] = \tau^2$.

In contrast to classical statistics, where the random variables are assumed to be independent of each other, the underlying signal in the domain has a dependence structure. To model this structure, certain classes of correlation families are chosen. Some families which are commonly used in geostatistics include the Matérn family, spherical family and the powered exponential family (Diggle and Ribeiro, 2007). Depending on the particular application, an

appropriate correlation function should be used. The common feature among these families is that they are all positive semidefinite functions, guaranteeing that the problem of negative variances does not occur. For example, the powered exponential family, has the correlation function of the form,

$$\rho(h) = \exp \left\{ \left(\frac{-h}{\phi} \right)^\kappa \right\},$$

where h is the distance between two measurements. Any distance measures can be used; ϕ is the scale parameter; and κ is a shape parameter determining the analytic smoothness of the underlying signal.

After the model for the underlying signal and the spatial structure has been specified, the next step is to identify the assumptions. The second order intrinsic stationary hypothesis is the most simple one, it is composed of three main parts (Wackernagel, 1998):

(a) The mean is constant in the domain, i.e., for all X in the domain

$$E[S(X)|X = x] = \mu. \quad (2)$$

(b) The covariance depends only on the distance between the two sample locations, i.e., for all X_i, X_j in the domain

$$Cov[S(X_i), S(X_j)|X_i = x_i, X_j = x_j] = \gamma[S(x_i), S(x_j)] = \gamma(\|x_i - x_j\|), i \neq j, \quad (3)$$

where $\| \cdot \|$ denotes the distance between X_i and X_j .

(c) The variance is finite,

$$\text{VAR}[S(X)] < \infty, \quad (4)$$

and it is the same at all locations.

This second order intrinsic stationary hypothesis also implies the *isotropic* (Wackernagel, 1998) conditions, meaning that the covariance functions behave the same in all directions within the spatial region.

In the following three subsections, the properties of the estimator in *kriging of the mean*, *simple kriging* and *ordinary kriging* (Wackernagel, 1998) are examined. Kriging of the mean is a procedure when one is trying to predict the mean of a certain characteristic of interest in the spatial region, using a weighted average estimator. Simple kriging means a prediction of the underlying signal at a certain point in the spatial region is made, assuming the mean of the spatial region is known according to some prior studies or subject matter knowledge. Ordinary kriging has the same purpose as simple kriging, but the prediction is done without assuming that the mean is known, and the mean is often estimated from the available data. Adopting the terminology which is usually used in geostatistics, the *estimation variance* (Wackernagel, 1998) of each estimator is crucial. It helps to compare the performance of different estimators and is actually the same as the mean squared error that are usually used as a criterion to compare estimators in statistics. The name, *kriging variance* (Cressie, 1986), is used for referring to the minimum estimation variance produced by a set of optimal *prediction weights* (Diggle and Ribeiro, 2007).

2.2 Kriging of the Mean

The estimator for kriging the mean is,

$$\hat{\mu}_{\text{KM}} = \sum_{i=1}^n a_i S(X_i), \quad (5)$$

where a_i 's are the prediction weights associated with each sample location. To understand prediction weights more clearly, one can interpret each prediction weight as the contribution of that particular sample point to the predicted value; and we search for these prediction weights through optimization. In order for the estimator in equation (5) to be unbiased, the constraint, $\sum_{i=1}^n a_i = 1$, needs to be imposed. Then,

$$\begin{aligned}
E[\hat{\mu}_{\text{KM}}] &= E\left[\sum_{i=1}^n a_i S(X_i)\right] \\
&= \sum_{i=1}^n E[a_i S(X_i)] \\
&= \sum_{i=1}^n a_i E[S(X_i)] \\
&= \underbrace{\sum_{i=1}^n a_i}_{1} (\mu) \\
&= \mu.
\end{aligned}$$

In geostatistics, people usually look at the variance of the difference between the estimator and the true parameter value, and they refer to this as the estimation variance. In fact, when the estimator is unbiased, one should notice that it is basically the mean squared error of the estimator. However, the term estimation variance will be used throughout this thesis since that is the

terminology used in geostatistics. The estimation variance of $\hat{\mu}_{\text{KM}}$ is,

$$\begin{aligned}
\text{VAR}[\hat{\mu}_{\text{KM}} - \mu] &= E[(\hat{\mu}_{\text{KM}} - \mu)^2] - (E[\hat{\mu}_{\text{KM}} - \mu])^2 \\
&= E[(\hat{\mu}_{\text{KM}}^2 + \mu^2 - 2\hat{\mu}_{\text{KM}}\mu)] \\
&= E[(\sum_{i=1}^n a_i S(X_i))^2] + E[\mu^2] - 2E[\mu(\sum_{i=1}^n a_i S(X_i))] \\
&= \sum_{i=1}^n \sum_{j=1}^n a_i a_j E[S(X_i)S(X_j)] + \mu^2 - 2\mu \sum_{i=1}^n a_i E[S(X_i)] \\
&= \sum_{i=1}^n \sum_{j=1}^n a_i a_j (\gamma[S(X_i), S(X_j)] + E[S(X_i)]E[S(X_j)]) + \mu^2 - 2\mu^2 \\
&= \sum_{i=1}^n \sum_{j=1}^n a_i a_j \gamma[S(X_i), S(X_j)]. \tag{6}
\end{aligned}$$

This estimation variance is used as the criterion to evaluate the goodness of $\hat{\mu}_{\text{KM}}$, and thus should be minimized. As a result, the goal then would be to find a set of weights which will produce the smallest estimation variance. Employing Lagrange's Multiplier's method, equation (6) is minimized subject to the constraint, $\sum_{i=1}^n a_i = 1$. The objective function to be minimized in this case is,

$$s(\mathbf{a}, \lambda) = \sum_{i=1}^n \sum_{j=1}^n a_i a_j \gamma[S(X_i), S(X_j)] - 2\lambda(\sum_{i=1}^n a_i - 1).$$

Taking the partial derivatives with respect to each a_i 's, and together with the constraint, it results in a system of $n + 1$ equations with $n + 1$ unknowns, that

is,

$$\begin{aligned}
\frac{\partial s}{\partial a_1} &= 2 \sum_{j=1}^n a_j \gamma[S(X_1), S(X_j)] - 2\lambda = 0 \\
&\Leftrightarrow \sum_{j=1}^n a_j \gamma[S(X_1), S(X_j)] - \lambda = 0 \\
&\vdots \\
\frac{\partial s}{\partial a_n} &= \sum_{j=1}^n a_j \gamma[S(X_n), S(X_j)] - \lambda = 0 \\
\frac{\partial s}{\partial 2\lambda} &= \sum_{i=1}^n a_i - 1 = 0.
\end{aligned}$$

Expressing the system in matrix form,

$$\begin{bmatrix} \gamma[S(X_1), S(X_1)] & \dots & \gamma[S(X_1), S(X_n)] \\ \vdots & \ddots & \vdots \\ \gamma[S(X_n), S(X_1)] & \dots & \gamma[S(X_n), S(X_n)] \\ 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \lambda \\ \vdots \\ \lambda \\ 1 \end{bmatrix}$$

it becomes obvious that,

$$\sum_{j=1}^n a_j \gamma[S(X_i), S(X_j)] = \lambda, \text{ for } i = 1, \dots, n,$$

and substituting the above expression back into equation (6), the minimum estimation variance, also known as the kriging variance is,

$$\sigma_{KM}^2 = \underbrace{\sum_{i=1}^n a_i}_1 \lambda = \lambda.$$

In other words, the kriging variance is half of the value of the Lagrange multiplier.

2.3 Simple Kriging

The estimator for simple kriging is,

$$\hat{S}_{SK}(x_0) = \mu + \sum_{i=1}^n a_i(S(X_i) - \mu), \quad (7)$$

where a_i 's are again, the prediction weights. However, they are now associated with the residual, $S(X_i) - \mu$, of each sample location rather than the sampled value itself. Similar to the previous section, the unbiasedness of this estimator is shown and then its kriging variance is derived.

Unlike in the case of kriging the mean, no constraints on the prediction weights are needed to make the estimator in equation (7) to be unbiased since,

$$\begin{aligned} E[\hat{S}_{SK}(x_0)] &= E\left[\mu + \sum_{i=1}^n a_i(S(X_i) - \mu)\right] \\ &= \mu + \sum_{i=1}^n a_i(E[S(X_i) - \mu]) \\ &= \mu + \sum_{i=1}^n a_i(\mu - \mu) \\ &= \mu. \end{aligned}$$

Now, in order to obtain an expression of the estimation variance so that it can be minimized by a set of prediction weights, some extra work needs to be done on expanding the terms. For simplicity, the subscript of $\hat{S}_{SK}(x_0)$ will be

dropped in the following calculations. The estimation variance is,

$$\begin{aligned}
\text{VAR}[\hat{S}(x_0) - S(x_0)] &= E[(\hat{S}(x_0) - S(x_0))^2] - (E[\hat{S}(x_0) - S(x_0)])^2 \\
&= E[(\hat{S}(x_0) - S(x_0))^2] \\
&= E[(\hat{S}(x_0))^2] + E[(S(x_0))^2] \\
&\quad - 2E[\hat{S}(x_0)S(x_0)].
\end{aligned} \tag{8}$$

Expanding each term in equation (8),

$$\begin{aligned}
E[(\hat{S}(x_0))^2] &= E[(\mu + \sum_{i=1}^n a_i(S(X_i) - \mu))^2] \\
&= \mu^2 + \sum_{i=1}^n \sum_{j=1}^n a_i a_j E[(S(X_i) - \mu)(S(X_j) - \mu)] + 2\mu \sum_{i=1}^n a_i E[(S(X_i) - \mu)] \\
&= \mu^2 + \sum_{i=1}^n \sum_{j=1}^n a_i a_j \gamma[S(X_i), S(X_j)] + 2\mu \sum_{i=1}^n a_i E[S(X_i) - \mu] \\
&= \mu^2 + \sum_{i=1}^n \sum_{j=1}^n a_i a_j \gamma[S(X_i), S(X_j)].
\end{aligned}$$

$$\begin{aligned}
E[(S(x_0))^2] &= \text{VAR}[S(x_0)] + (E[S(x_0)])^2 \\
&= \gamma[S(x_0), S(x_0)] + \mu^2.
\end{aligned}$$

$$\begin{aligned}
E[\hat{S}(x_0)S(x_0)] &= \gamma[\hat{S}(x_0), S(x_0)] + E[\hat{S}(x_0)]E[S(x_0)] \\
&= \gamma[\hat{S}(x_0), S(x_0)] + \mu^2 \\
&= \gamma[\mu + \sum_{i=1}^n a_i(S(X_i) - \mu), S(x_0)] + \mu^2 \\
&= \sum_{i=1}^n a_i \gamma[S(X_i), S(x_0)] + \mu^2.
\end{aligned}$$

and rewriting equation (8) using the above, it results in,

$$\begin{aligned}
f(\mathbf{a}) = \text{VAR}[\hat{S}(x_0) - S(x_0)] &= \sum_{i=1}^n \sum_{j=1}^n a_i a_j \gamma[S(X_i), S(X_j)] + \gamma[S(x_0), S(x_0)] \\
&\quad - 2 \sum_{i=1}^n a_i \gamma[S(X_i), S(x_0)]. \tag{9}
\end{aligned}$$

Now, this function should be minimized without any constraints. Taking the partial derivatives of equation (9) with respect to each a_i , a system of n equations and n unknowns is obtained. That is,

$$\begin{aligned}
\frac{\partial f(\mathbf{a})}{\partial a_1} &= 2 \sum_{j=1}^n a_j \gamma[S(X_1), S(X_j)] - 2\gamma[S(X_1), S(x_0)] = 0 \\
&\Leftrightarrow \sum_{j=1}^n a_j \gamma[S(X_1), S(X_j)] - \gamma[S(X_1), S(x_0)] = 0 \\
\frac{\partial f(\mathbf{a})}{\partial a_2} &= \sum_{j=1}^n a_j \gamma[S(X_2), S(X_j)] - \gamma[S(X_2), S(x_0)] = 0 \\
&\vdots \\
\frac{\partial f(\mathbf{a})}{\partial a_n} &= \sum_{j=1}^n a_j \gamma[S(X_n), S(X_j)] - \gamma[S(X_n), S(x_0)] = 0
\end{aligned}$$

Solving the above system of equations yields the optimal prediction weights.

With these optimal weights and rewriting the above system of equations,

$$\sum_{j=1}^n a_j \gamma[S(X_i), S(X_j)] = \gamma[S(X_i), S(x_0)], \text{ for } i = 1, \dots, n$$

and substituting it back into the estimation variance, that is equation (9), the kriging variance would be,

$$\sigma_{SK}^2 = \gamma[S(x_0), S(x_0)] - \sum_{i=1}^n a_i \gamma[S(X_i), S(x_0)].$$

2.4 Ordinary Kriging

The estimator for ordinary kriging is,

$$\hat{S}_{OK}(x_0) = \sum_{i=1}^n a_i S(X_i), \quad (10)$$

where a_i 's are the prediction weights associated with each sample location. This estimator has a similar form compared to the estimator for the kriged mean in equation (5). Thus, intuitively there should be a constraint, $\sum_{i=1}^n a_i = 1$ in order for the estimator in equation (10) to be unbiased. Indeed,

$$\begin{aligned} E[\hat{S}_{OK}(x_0)] &= E\left[\sum_{i=1}^n a_i S(X_i)\right] \\ &= \sum_{i=1}^n a_i E[S(X_i)] \\ &= \underbrace{\sum_{i=1}^n a_i}_{1} \mu \\ &= \mu, \end{aligned}$$

yields an unbiased estimator for ordinary kriging.

Again, the subscript of $\hat{S}_{OK}(x_0)$ will be dropped again for the ease of notation

in the following calculations. The estimation variance is,

$$\begin{aligned}
\text{VAR}[\hat{S}(x_0) - S(x_0)] &= E[(\hat{S}(x_0) - S(x_0))^2] \\
&= E[(\hat{S}(x_0))^2 + (S(x_0))^2 - 2(\hat{S}(x_0)S(x_0))] \\
&= E[(S(x_0))^2] - 2E[(\hat{S}(x_0)S(x_0))] \\
&\quad + E[(\hat{S}(x_0))^2].
\end{aligned} \tag{11}$$

Expanding the terms in equation (11),

$$\begin{aligned}
E[(S(x_0))^2] &= \text{VAR}[S(x_0)] + (E[S(x_0)])^2 \\
&= \gamma[S(x_0), S(x_0)] + \mu^2.
\end{aligned}$$

$$\begin{aligned}
E[\hat{S}(x_0)S(x_0)] &= \gamma[\hat{S}(x_0), S(x_0)] + E[\hat{S}(x_0)]E[S(x_0)] \\
&= \gamma[\hat{S}(x_0), S(x_0)] + \mu^2 \\
&= \sum_{i=1}^n a_i \gamma[S(X_i), S(x_0)] + \mu^2.
\end{aligned}$$

$$\begin{aligned}
E[(\hat{S}(x_0))^2] &= \text{VAR}[\hat{S}(x_0)] + (E[\hat{S}(x_0)])^2 \\
&= \text{VAR}\left[\sum_{i=1}^n a_i S(X_i)\right] + \mu^2 \\
&= \sum_{i=1}^n a_i^2 \text{VAR}[S(X_i)] + \sum_{\substack{i \neq j \\ i, j=1}}^n a_i a_j \gamma[S(X_i), S(X_j)] + \mu^2 \\
&= \sum_{i=1}^n \sum_{j=1}^n a_i a_j \gamma[S(X_i), S(X_j)] + \mu^2.
\end{aligned}$$

As a result, rewriting Equation (11),

$$\begin{aligned} \text{VAR}[\hat{S}(x_0) - S(x_0)] &= \gamma[S(x_0), S(x_0)] - 2 \sum_{i=1}^n a_i \gamma[S(x_0), S(X_i)] \\ &\quad + \sum_{i=1}^n \sum_{j=1}^n a_i a_j \gamma[S(X_i), S(X_j)]. \end{aligned} \quad (12)$$

With the constraint, $\sum_{i=1}^n a_i = 1$, Lagrange Multiplier's method is used to minimize the estimation variance by obtaining a set of optimal prediction weights.

In this case, the objective function to be minimized is,

$$\begin{aligned} s(\mathbf{a}, 2\lambda) &= \gamma[S(x_0), S(x_0)] - 2 \sum_{i=1}^n a_i \gamma[S(x_0), S(X_i)] + \sum_{i=1}^n \sum_{j=1}^n a_i a_j \gamma[S(X_i), S(X_j)] \\ &\quad - 2\lambda \left(\sum_{i=1}^n a_i - 1 \right), \end{aligned}$$

where 2λ is the Lagrange multiplier. Taking partial derivatives with respect to each a_i , and together with the constraint, a system of $n + 1$ equations and $n + 1$ unknowns is obtained. That is,

$$\begin{aligned} \frac{\partial s}{\partial a_1} &= -2\gamma[S(x_0), S(X_1)] + 2 \sum_{j=1}^n a_j \gamma[S(X_1), S(X_j)] - 2\lambda = 0 \\ \Leftrightarrow & -\gamma[S(x_0), S(X_1)] + \sum_{j=1}^n a_j \gamma[S(X_1), S(X_j)] - \lambda = 0 \\ \frac{\partial s}{\partial a_2} &= -\gamma[S(x_0), S(X_2)] + \sum_{j=1}^n a_j \gamma[S(X_2), S(X_j)] - \lambda = 0 \\ &\vdots \\ \frac{\partial s}{\partial a_n} &= -\gamma[S(x_0), S(X_n)] + \sum_{j=1}^n a_j \gamma[S(X_n), S(X_j)] - \lambda = 0 \\ \frac{\partial s}{\partial 2\lambda} &= \sum_{i=1}^n a_i - 1 = 0. \end{aligned}$$

Rearranging the terms and put these $n + 1$ equations into matrix notations, it can be expressed as:

$$\begin{bmatrix} \gamma(X_1, X_1) & \gamma(X_1, X_2) & \dots & \gamma(X_1, X_n) \\ \gamma(X_2, X_1) & \gamma(X_2, X_2) & \dots & \gamma(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(X_n, X_1) & \gamma(X_n, X_2) & \dots & \gamma(X_n, X_n) \\ 1 & 1 & \dots & 1 \end{bmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} - \lambda \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \gamma(x_0, X_1) \\ \gamma(x_0, X_2) \\ \vdots \\ \gamma(x_0, X_n) \\ 1 \end{pmatrix}$$

and augmenting the first matrix on the left hand side of the above equation:

$$\begin{bmatrix} \gamma(X_1, X_1) & \gamma(X_1, X_2) & \dots & \gamma(X_1, X_n) & 1 \\ \gamma(X_2, X_1) & \gamma(X_2, X_2) & \dots & \gamma(X_2, X_n) & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma(X_n, X_1) & \gamma(X_n, X_2) & \dots & \gamma(X_n, X_n) & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \\ -\lambda \end{pmatrix} = \begin{pmatrix} \gamma(x_0, X_1) \\ \gamma(x_0, X_2) \\ \vdots \\ \gamma(x_0, X_n) \\ 1 \end{pmatrix}.$$

In order to obtain the prediction weights and the Lagrange Multiplier, one can multiply the inverse of the augmented covariance matrix on both sides from the left,

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \\ -\lambda \end{pmatrix} = \begin{bmatrix} \gamma(X_1, X_1) & \gamma(X_1, X_2) & \dots & \gamma(X_1, X_n) & 1 \\ \gamma(X_2, X_1) & \gamma(X_2, X_2) & \dots & \gamma(X_2, X_n) & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma(X_n, X_1) & \gamma(X_n, X_2) & \dots & \gamma(X_n, X_n) & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix}^{-1} \begin{pmatrix} \gamma(x_0, X_1) \\ \gamma(x_0, X_2) \\ \vdots \\ \gamma(x_0, X_n) \\ 1 \end{pmatrix}.$$

Then, with the optimal prediction weights and the Lagrange Multiplier, and by substituting

$$\sum_{j=1}^n a_j \gamma[S(X_i), S(X_j)] = \gamma[S(x_0), S(X_i)] + \lambda, \text{ for } i = 1, \dots, n \quad (13)$$

into equation (12); the kriging variance in ordinary kriging is,

$$\begin{aligned} \sigma_{OK}^2 &= \gamma[S(x_0), S(x_0)] - 2 \sum_{i=1}^n a_i \gamma[S(x_0), S(X_i)] \\ &\quad + \sum_{i=1}^n a_i \gamma[S(x_0), S(X_i)] + \lambda \\ &= \lambda + \gamma[S(x_0), S(x_0)] - \sum_{i=1}^n a_i \gamma[S(x_0), S(X_i)]. \end{aligned}$$

The only difference between simple kriging and ordinary kriging is that the mean is known in the former. One may wonder what will happen if the kriged mean is used in the simple kriging estimator. Indeed, it is the same as the ordinary kriging system. To show that this is the case, recall that the kriged mean is,

$$\hat{\mu}_{KM} = \sum_{i=1}^n a_i S(X_i),$$

and the simple kriging estimator is,

$$\hat{S}_{SK}(x_0) = \mu + \sum_{i=1}^n a_i (S(X_i) - \mu).$$

To avoid confusions on the notations, the prediction weights, a_i 's, in the simple kriging estimator are replaced by w_i 's. Thus,

$$\hat{S}_{SK}(x_0) = \mu + \sum_{i=1}^n w_i (S(X_i) - \mu).$$

Now, with the kriged mean,

$$\begin{aligned} \hat{S}_{SK}(x_0) &= \hat{\mu}_{KM} + \sum_{i=1}^n w_i (S(x_i) - \hat{\mu}_{KM}) \\ &= \sum_{i=1}^n a_i S(x_i) + \sum_{i=1}^n w_i (S(x_i) - \sum_{i=1}^n a_i S(x_i)) \\ &= \sum_{i=1}^n a_i S(x_i) + \sum_{i=1}^n w_i S(x_i) - \sum_{i=1}^n w_i \sum_{j=1}^n a_j S(x_j) \\ &= \sum_{i=1}^n w_i S(x_i) + \sum_{i=1}^n a_i S(x_i) - \sum_{i=1}^n w_i \sum_{j=1}^n a_j S(x_j) \\ &= \sum_{i=1}^n w_i S(x_i) + \sum_{i=1}^n a_i S(x_i) [1 - \sum_{i=1}^n w_i] \\ &= \sum_{i=1}^n S(x_i) [w_i + a_i (1 - \sum_{i=1}^n w_i)]. \end{aligned}$$

Let $\bar{w} = 1 - \sum_{i=1}^n w_i$, then,

$$\begin{aligned} \hat{S}_{SK}(x_0) &= \sum_{i=1}^n S(x_i) [w_i + a_i \bar{w}] \\ &= \sum_{i=1}^n [w_i + \bar{w} a_i] S(x_i). \end{aligned}$$

If the above estimator is the same as the ordinary kriging estimator (equation (10)), then $\sum_{i=1}^n [w_i + \bar{w} a_i]$ should be equal to 1 as well. Since a_i 's are the

prediction weights of the kriged mean estimator, so $\sum_{i=1}^n a_i = 1$. Then,

$$\begin{aligned}
\sum_{i=1}^n [w_i + \bar{w}a_i] &= \sum_{i=1}^n w_i + \bar{w} \sum_{i=1}^n a_i \\
&= \sum_{i=1}^n w_i + \bar{w} \\
&= \sum_{i=1}^n w_i + 1 - \sum_{i=1}^n w_i \\
&= 1.
\end{aligned}$$

Next, the equivalence of the two kriging systems should be checked. Recall that the kriging system for ordinary kriging (equation (13)) can be written as,

$$\sum_{j=1}^n a_j \gamma[S(X_i), S(X_j)] = \gamma[S(x_0), S(X_i)] + \lambda, \text{ for } i = 1, \dots, n.$$

In simple kriging, the kriging system for finding the minimum estimation variance was,

$$\gamma[S(X_i), S(x_0)] = \sum_{j=1}^n w_j \gamma[S(X_i), S(X_j)],$$

for $i = 1, \dots, n$. For kriging of the mean, the kriging system is,

$$\sum_{j=1}^n a_j \gamma[S(X_i), S(X_j)] = \lambda,$$

for $i = 1, \dots, n$. Therefore, rewriting equation (13) we get,

$$\begin{aligned}
\sum_{j=1}^n [w_j + \bar{w}a_j] \gamma[S(X_i), S(X_j)] &= \underbrace{\sum_{j=1}^n w_j \gamma[S(X_i), S(X_j)]}_{\text{simple kriging}} + \bar{w} \underbrace{\sum_{j=1}^n a_j \gamma[S(X_i), S(X_j)]}_{\text{kriging of the mean}} \\
&= \gamma[S(X_i), S(x_0)] + \bar{w} \lambda \\
&= \gamma[S(X_i), S(x_0)] + \lambda',
\end{aligned}$$

for $i = 1, \dots, n$, subject to $\sum_{i=1}^n [w_i + \bar{w}a_i] = 1$. One can see that the ordinary kriging system can be decomposed into the two kriging systems: simple kriging and kriging of the mean. With the set of optimal prediction weights which minimizes the estimation variance in all three kriging methods, indeed the simple kriging system using the kriged mean and the ordinary kriging system are equivalent. So, unless the known mean assumed in simple kriging is obtained from other sources and is different from the kriged mean, the prediction values and its kriging variance obtained by simple kriging and ordinary kriging would be the same.

2.5 Tools in Spatial Data Analysis

In this section, the techniques of analyzing a spatial dataset are introduced. The focus here is on the general tools that one would need if predictions of a specific characteristic on a spatial region is of interest.

As with other statistical analyses, after the collection of data, a graphical representation is desired to gain a preliminary understanding of the observed data. First, a plot of the observations against the locations can be drawn. If the data is two-dimensional, one can separately plot the observations against the eastings (x) and northings (y) coordinates. The purpose of drawing these

plots is to determine if there is a *spatial trend* (Diggle and Ribeiro Jr., 2007) exist. Spatial trend simply means that the location has some effects on the observations, possibly causing a non-constant mean; which ultimately violates the second-order intrinsic stationary assumption. This problem can be dealt with using an appropriate model or by transformations, but this is not the scope of the discussion. Similarly one can include the dependence of the covariates at the locations in the model.

Due to the existence of correlation in the spatial region, estimates of the parameters in the covariance function, which are used to model the dependence structure are required. The first step is to compute the variogram. For any two realizations in the domain, its *dissimilarity* (Wackernagel, 1998) can be computed by averaging the squared difference between them. Mathematically, it can be expressed as:

$$\gamma^*(h) = \frac{1}{2}[S(X+h) - S(X)]^2.$$

A *variogram cloud* (Wackernagel 1998) is then obtained by plotting $\gamma^*(h)$ versus $|h|$. This helps to determine if there are any inhomogeneities among the sample values. In general, smaller value of h should correspond to a smaller $\gamma^*(h)$ value. From the variogram cloud, the *experimental variogram* (Wackernagel 1998) can be produced by first breaking up $|h|$, into disjoint intervals, and calculate the average dissimilarities within each particular interval. The average dissimilarity is defined as,

$$\bar{\gamma}^*(h) = \frac{1}{2n_\alpha} \sum_{i=1}^{n_\alpha} [S(X_i + h) - S(X_i)]^2,$$

where n_α is the total number of pairs in each interval.

A plot of average dissimilarities versus $|h|$ can then be obtained and this is the experimental variogram. We also have a theoretical variogram. Under the second order intrinsic stationary assumption, the theoretical variogram is defined as,

$$\begin{aligned} V(X, X+h) &= \frac{1}{2}E[(S(X+h) - S(X))^2] \\ &= \frac{1}{2}\text{VAR}[S(X+h) - S(X)]. \end{aligned}$$

By expanding Y_i and Y_j using the model in equation (1), the theoretical variogram can be rewritten as,

$$\begin{aligned} V(Y_i, Y_j) &= \frac{1}{2}E[(Y_i - Y_j)^2] \\ &= \tau^2 + \sigma^2[1 - \rho(h)], \end{aligned} \tag{14}$$

where τ^2 , σ^2 are as defined in Section 2.1, and $\rho(h)$ is the correlation function. Figure 1 shows a typical theoretical variogram and its major components. There are several attributes of the theoretical variogram which are essential, as described in Diggle and Ribeiro Jr. (2007). It includes: (a) *nugget variance*; (b) *range* and (c) *sill*.

(a) Nugget variance (τ^2) - Variation contributed from two sources: (1) measurement errors; and (2) short range spatial variation. If the sampling design does not contain any two points which has a distance less than h_0 apart, and that the correlation function for the process is zero for all $h \geq h_0$, then we cannot tell whether the variation is due to measurement error or short range spatial variation. In order to fit a model to interpolate the data, usually this

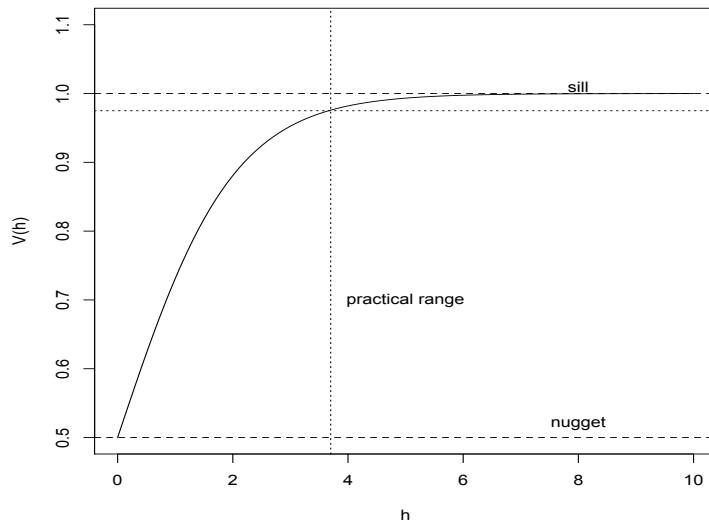
causes a jump at the origin, and this is called the *nugget-effect* (Diggle and Ribeiro Jr., 2007).

(b) Range - A value of h , the distance at which the variogram reaches the sill. When the correlation function only approaches zero asymptotically as h increases, we would take $\rho(h) = 0.05$ and look at its *practical range* (Diggle and Ribeiro Jr., 2007). As a result, the value of the sill in that case becomes $\tau^2 + \sigma^2[1 - 0.05] = \tau^2 + 0.95\sigma^2$.

(c) Sill ($\tau^2 + \sigma^2$) - Variation of the observed process Y . By looking at equation (14), the theoretical variogram can be expressed in terms of the correlation function, $\rho(h)$. When the correlation function equals to zero, equation (14) then equals to σ^2 , which is the signal variance. Most of the time a nugget effect is present and the sample variogram starts at the value, τ^2 . In that scenario, when the correlation decays to zero as the distance increases, the value of the sill is $\tau^2 + \sigma^2$. With the theoretical variogram, initial estimates of the covariance parameters are obtained. They can then be used in the numerical optimization procedures for finding the maximum likelihood estimates for the covariance parameters. With the estimates, spatial prediction can be done using the kriging estimators discussed in the previous sections.

So far variograms and spatial predictions are discussed under second order intrinsic stationary assumption. The first immediate question that a statistician would ask is the validity of the assumptions. In practice, this assumption of second order intrinsic stationary often does not hold and the variation depends on the relative orientation of sample locations, in addition to the distance of the pairs. This is known as *directional effects* or *anisotropy* (Zimmerman, 1993). In Zimmerman (1993), the various types of anisotropies are discussed. However, similar to spatial trend, the strategies of dealing with these viola-

Figure 1: A Typical Theoretical Variogram shown.



tions of the assumption is not the scope of the current discussion. Intuitively, anisotropic behavior can be studied by examining the experimental variograms computed from the data, in different directions of h . For example, the default directions used in R are the E-W direction (0 degrees), NE-SW direction (45 degrees), N-S direction (90 degrees) and NW-SE direction (135 degrees). Furthermore, a tolerance angle needs to be set for each direction since not every sample pair are aligned with one of those directions just mentioned. For illustration, the tolerance angle equal to 22.5 degrees (which is the default setting in R) is set and the NE-SW direction is examined. Then using an arbitrary point as the reference, any other points which fall between 22.5 degrees to 67.5 degrees and 202.5 degrees to 247.5 degrees would be included in the calculation of experimental variogram. Such an experimental variogram taking in consideration the directions between sample pairs is called a *directional variogram*

(Zimmerman, 1993). Usually these directional variograms are plotted together in the same plot for direct comparison.

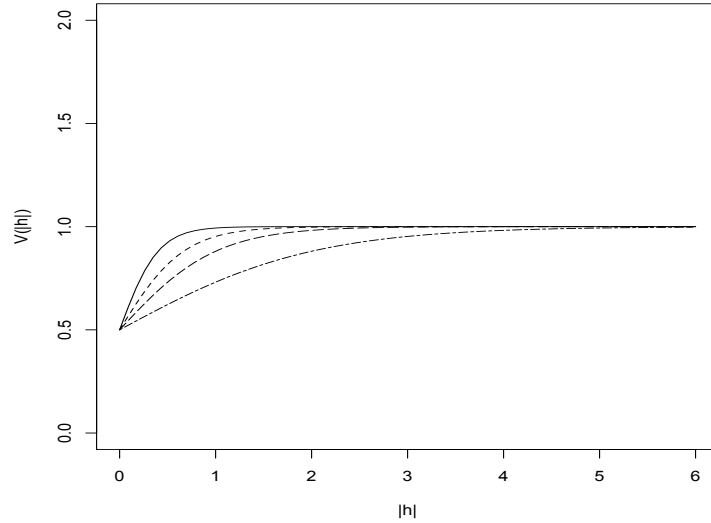
There are various types of anisotropy in geostatistics. The classical term, geometrical anisotropy, is widely accepted in books such as Diggle and Ribeiro Jr. (2007) and Wackernagel (1998). Zimmerman (1993), states that other forms of anisotropic behavior are vaguely defined by the term, *zonal anisotropy*, and there is no general agreement on its definition. As a result, Zimmerman proposed three other types of anisotropy, the *sill anisotropy*, *range anisotropy* and *nugget anisotropy* to avoid the ambiguous meaning of zonal anisotropy. These three types of anisotropy can be used in any combinations to describe a dataset when it is appropriate.

Geometrical anisotropy is the simplest form of anisotropic behavior. To correct it, for two dimensional data, the transformation of coordinates is given by (Diggle and Ribeiro Jr., 2007):

$$(x', y') = (x, y) \begin{bmatrix} \cos(\psi_A) & \sin(\psi_A) \\ -\sin(\psi_A) & \cos(\psi_A) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \psi_R^{-1} \end{bmatrix} \quad (15)$$

where (x, y) are the original coordinates in the anisotropic space, ψ_A the anisotropic angle, $\psi_R > 1$ the anisotropic ratio and (x', y') the transformed coordinates in the isotropic space. The matrix, $\begin{bmatrix} \cos(\psi_A) & \sin(\psi_A) \\ -\sin(\psi_A) & \cos(\psi_A) \end{bmatrix}$, is called the rotation matrix. Its function is to rotate the coordinate axes so that they would be parallel to the principal axes of the ellipse. The matrix, $\begin{bmatrix} 1 & 0 \\ 0 & \psi_R^{-1} \end{bmatrix}$, then either shrinks or dilates the principal axes and the coordinates system would have a spherical shape. Note that ψ_A and ψ_R are two

Figure 2: An example of range anisotropic variogram, with each line representing a different direction.

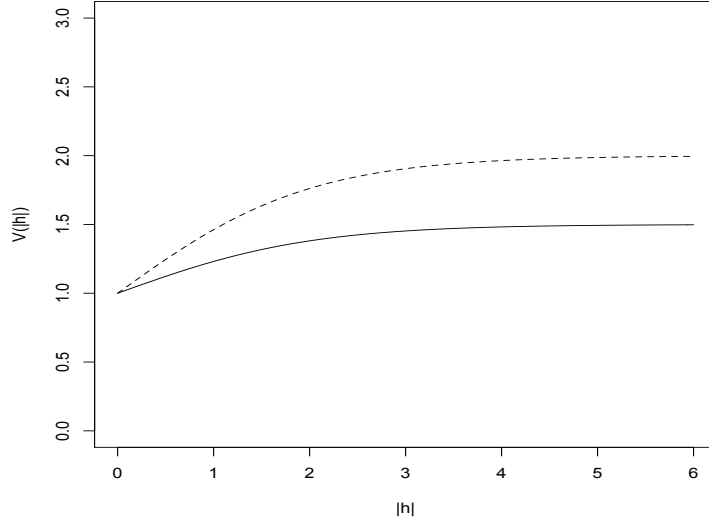


additional parameters which needs to be estimated.

The second type is range anisotropy. An example of range anisotropic behavior is shown in Figure 2. The distance where the sill is reached is different for each line, where each line represents the variogram in a certain direction. Range anisotropy can be either classified as geometric or non-geometric. When the range is the only attribute which varies with direction, a *rose diagram* (Zimmerman, 1993) can help in order to identify whether the anisotropic behaviour is geometric or not. A rose diagram is simply a direction versus range plot, in the polar coordinate system, in which θ is used to denote direction and r to represent the range. If the rose diagram appears to be elliptical, then a geometrical anisotropic model is deemed to be appropriate. Zimmerman (1993) further proposed two improvements for modelling range anisotropy.

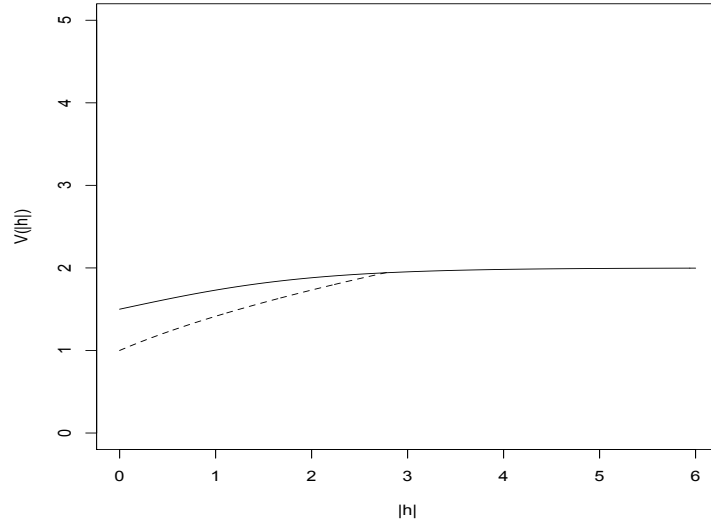
Figure 3 is an illustration of existence of sill anisotropy in directional vari-

Figure 3: An example of sill anisotropic variogram, with each line representing a different direction



ograms. The plateau of each line is achieved at different heights. Sill anisotropy refers to the situation when the value of the sill is different in different directions. Similar to range anisotropy, Zimmerman (1993) in his paper emphasizes the danger of applying the current modelling procedure when evidence of sill anisotropy is present. He believes that there are three possible causes for this type of anisotropy: (1) non-vanishing spatial correlation; (2) spatial trend effect; and (3) measurement errors are not independent or with unequal means. Another possible type of anisotropy mentioned by Zimmerman (1993) is nugget anisotropy, as shown in Figure 4. The two reasons why this might occur are: (1) due to the spatial process or (2) correlated measurement errors. Actually, it can be shown easily by an example of how each factor mentioned above causes the particular type of anisotropic behavior. For readers who are wishing to pursue a more in depth understanding of anisotropies and also the strategies

Figure 4: An example of nugget anisotropic variogram, with each line representing a different direction



of dealing with such situations, they can refer to Zimmerman (1993). The purpose here is only to identify the possible violation of the second order intrinsic stationary assumption.

3 Performance of the Estimators

Having introduced the prediction methods available in geostatistics in the previous section, it is now appropriate to come to the core objective of this thesis. That is, to calculate the minimum sample size required for achieving a desired margin of error in a spatial region. Remembering that two questions mentioned in Section 1 that worth investigating are, the consistency of the kriging estimators and their performance. A manuscript by Gombay (2010) has investigated these two questions in details, and in fact the main objective of this thesis is to verify the theoretical results presented there by simulation studies. The results are shown in Section 4; and in this section we will give a summary of the manuscript, which this thesis is based on.

As we have seen in Section 2, covariance functions have a direct impact on the prediction results obtained from kriging. Under the spatial setting, it was shown in Gombay (2010) that sample mean is not consistent, and the laws of large numbers do not hold. The same also happens in kriging of the mean and ordinary kriging. Gombay (2010) then gives a necessary and sufficient condition, which will be discussed shortly, on the covariance function for the convergence of the estimators. Now let us first look at the consistency problem in spatial statistics.

3.1 Convergence of Sample Mean and Kriging Estimators: The Necessary and Sufficient Condition

The usual setup in geostatistics is that we first fix the spatial region of interest, and then measurements are taken. No matter how many observations are made, the size and location of the spatial region remains unchanged. When

we study the behaviour of the estimators for large samples, in this situation, this would be called *infill asymptotics* (Cressie, 1993). While there is a rich literature on *increasing domain asymptotics* (Cressie, 1993), as in time series analysis, the domain increases as the sample size increases; it is hard to find a discussion on the asymptotic distribution of the estimators in infill situations. Following Gombay's (2010) approach, we will look at the sample mean first, followed by the kriging estimators.

Recall that,

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n S(X_i),$$

is the sample mean. Since X_i are independent and identically distributed random variables, using the conditional variance formula, the variance of the sample mean is,

$$\text{VAR}[\hat{\mu}_n] = \text{VAR}[E[\hat{\mu}_n|X_1 \dots X_n]] + E[\text{VAR}[\hat{\mu}_n|X_1 \dots X_n]]. \quad (16)$$

The first term on the right hand side is zero since it is the variance of a constant, under the stationary mean assumption. Therefore, we only have to look at the second term. Expanding the second term in equation (16), we have,

$$\text{VAR}[\hat{\mu}_n|X_1 \dots X_n] = \frac{1}{n} \text{VAR}[S(X_1)] + \frac{2}{n^2} \sum_{i < j} \gamma(\|X_i - X_j\|). \quad (17)$$

Noticing that in fact, we can express $\frac{2}{n^2} \sum_{i < j} \gamma(\| X_i - X_j \|)$ by a U-statistic with symmetric kernel, that is,

$$U_n = \frac{2}{n(n-1)} \sum_{i < j} \gamma(\| X_i - X_j \|),$$

which gives us the minimum variance unbiased estimator for the quantity, $E\{\gamma(\| X_i - X_j \|)\}$. Therefore, substituting back into equation (17), we obtain,

$$\text{VAR}[\hat{\mu}_n] = E\{\gamma(\| X_i - X_j \|)\} + \frac{1}{n} [\text{VAR}[S(X_1)] - E\{\gamma(\| X_i - X_j \|)\}].$$

So from Gombay (2010) we see that the following Proposition is true:

Proposition 1 (Inconsistency of the Sample Mean) *For independent and identically distributed random locations, X_1, \dots, X_n , and under the second order intrinsic stationary assumptions (i.e. conditions (2), (3) and (4)),*

$$\lim_{n \rightarrow \infty} \text{VAR}[\hat{\mu}_n] = E\{\gamma(\| X_1 - X_2 \|)\} + \frac{1}{n} [\text{VAR}[S(X_1)] - E\{\gamma(\| X_1 - X_2 \|)\}],$$

therefore unless $E\{\gamma(\| X_1 - X_2 \|)\} = 0$, otherwise the sample mean is not a consistent estimator of μ .

We can easily see that $E\{\gamma(\| X_i - X_j \|)\} \geq 0$ as otherwise we would have negative $\text{VAR}[\hat{\mu}_n]$ for large n . Next, recall that the kriged mean is,

$$\hat{\mu}_{KM} = \sum_{i=1}^n a_i S(X_i),$$

and we can express its variance as,

$$\text{VAR}[\hat{\mu}_{KM}] = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n a_j \gamma[S(X_i), S(X_j)]. \quad (18)$$

Using the conditional variance formula again, similar to the case for sample mean, we only have to look at the expectation of equation (18). That is,

$$\begin{aligned} E[\text{VAR}[\hat{\mu}_{KM}|X_1 \dots X_n]] &= \frac{1}{n} \text{VAR}[S(X_1)] + \frac{1}{n} E\{E[\sum_{i=1}^n \sum_{j \neq i}^n a_j \gamma(\|X_i - X_j\|) | a_1 \dots a_n]\} \\ &= \frac{1}{n} \text{VAR}[S(X_1)] + E[\gamma(\|X_1 - X_2\|)] - \frac{1}{n} E[\gamma(\|X_1 - X_2\|)] \\ &= E\{\gamma(\|X_i - X_j\|)\} \\ &\quad + \frac{1}{n} [\text{VAR}[S(X_1)] - E\{\gamma(\|X_i - X_j\|)\}], \end{aligned} \quad (19)$$

which is the same as the conditional variance of sample mean. Obviously, equation (17) and (19) are exactly the same, showing that the variance does not approach zero even if the sample size is asymptotically large enough, when $E\{\gamma(\|X_i - X_j\|)\} \neq 0$. In other words, both estimators are not consistent and the laws of large numbers do not hold.

Proposition 2 (Inconsistency of the Kriged Mean) *For independent and identically distributed random locations, X_1, \dots, X_n , and under the second order intrinsic stationary assumption (i.e., conditions (2), (3) and (4)),*

$$\lim_{n \rightarrow \infty} \text{VAR}[\hat{\mu}_{KM}] = E\{\gamma(\|X_1 - X_2\|)\} + \frac{1}{n} [\text{VAR}[S(X_1)] - E\{\gamma(\|X_1 - X_2\|)\}],$$

therefore unless $E\{\gamma(\|X_1 - X_2\|)\} = 0$, otherwise the kriged mean is not a consistent estimator of μ .

Similar problem occurs in ordinary kriging. Recall that the estimator is,

$$\hat{S}_{OK}(x_0) = \sum_{i=1}^n a_i S(X_i),$$

and the kriging variance is,

$$\sigma_{OK}^2 = \lambda + \gamma[S(x_0), S(x_0)] - \sum_{i=1}^n a_i [S(x_0), S(X_i)],$$

where,

$$\lambda = \sum_{j=1}^n a_j \gamma[S(X_i), S(X_j)] - \gamma[S(x_0), S(X_i)], \text{ for } i = 1, \dots, n.$$

Combining the two equations, we get,

$$\begin{aligned} \sigma_{OK}^2 &= \text{VAR}[\hat{S}(x_0) - S(x_0) | X_1 \dots X_n] \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n a_j \gamma[S(X_i), S(X_j)] - \frac{1}{n} \sum_{i=1}^n \gamma[S(x_0), S(X_i)] \\ &\quad + \gamma[S(x_0), S(x_0)] - \sum_{i=1}^n a_i \gamma[S(x_0), S(X_i)] \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n a_j \gamma[S(X_i), S(X_j)] + \gamma[S(x_0), S(x_0)] \\ &\quad - \sum_{i=1}^n (a_i + \frac{1}{n}) \gamma[S(x_0), S(X_i)]. \end{aligned} \tag{20}$$

Rearranging the terms in equation (11), we have,

$$\begin{aligned} E[(\hat{S}(x_0))^2] &= \text{VAR}[\hat{S}(x_0) - S(x_0)] - E[(S(x_0))^2] - 2E[\hat{S}(x_0)S(x_0)] \\ &= \text{VAR}[\hat{S}(x_0) - S(x_0)] - \gamma[S(x_0), S(x_0)] - \mu^2 + 2 \sum_{i=1}^n a_i \gamma[S(X_i), S(x_0)] + 2\mu^2. \end{aligned}$$

Then,

$$E[(\hat{S}(x_0))^2] - \mu^2 = \text{VAR}[\hat{S}(x_0) - S(x_0)] - \gamma[S(x_0), S(x_0)] + 2 \sum_{i=1}^n a_i \gamma[S(X_i), S(x_0)],$$

and thus,

$$\text{VAR}[\hat{S}(x_0)] = \text{VAR}[\hat{S}(x_0) - S(x_0)] - \gamma[S(x_0), S(x_0)] + 2 \sum_{i=1}^n a_i \gamma[S(X_i), S(x_0)].$$

Substituting equation (20) into the above expression,

$$\begin{aligned} \text{VAR}[\hat{S}(x_0)] &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n a_j \gamma[S(X_i), S(X_j)] + 2 \sum_{i=1}^n a_i \gamma[S(X_i), S(x_0)] \\ &\quad - \sum_{i=1}^n (a_i + \frac{1}{n}) \gamma[S(x_0), S(X_i)] \\ &= \sum_{i=1}^n \left\{ \frac{1}{n} \sum_{j=1}^n a_j \gamma[S(X_i), S(X_j)] + a_i \gamma[S(X_i), S(x_0)] \right\} \\ &\quad - \frac{1}{n} \sum_{i=1}^n \gamma[S(x_0), S(X_i)]. \end{aligned}$$

If we proceed similarly with kriging of the mean, we obtain,

$$\text{VAR}[\hat{S}(x_0)] = E[\gamma(\|X_1 - X_2\|)] + \frac{1}{n} \{E[\text{VAR}[X_1]] - E[\gamma(\|X_1 - X_2\|)]\}.$$

Therefore, the consistency problem remains in ordinary kriging, unless $E[\gamma(\|X_1 - X_2\|)] = 0$. In fact, $E[\gamma(\|X_1 - X_2\|)] = 0$ is a necessary and sufficient condition for the three estimators to be consistent under the spatial setting. However, many covariance models, for example the ones that we mentioned earlier, all have $E[\gamma(\|X_1 - X_2\|)] > 0$.

Proposition 3 (Inconsistency of the Ordinary Kriging Estimator) *For*

independent and identically distributed random locations, X_1, \dots, X_n , and under the second order intrinsic stationary assumption (i.e., conditions (2), (3) and (4)),

$$\lim_{n \rightarrow \infty} \text{VAR}[\hat{S}(x_0)] = E\{\gamma(\|X_1 - X_2\|)\} + \frac{1}{n}[E[\text{VAR}[S(X_1)]] - E\{\gamma(\|X_1 - X_2\|)\}],$$

therefore unless $E\{\gamma(\|X_1 - X_2\|)\} = 0$, otherwise the ordinary kriging estimator is not a consistent estimator of $S(x_0)$.

3.2 Comparing the sample mean and the kriged mean

For random locations, X_1, \dots, X_n , we see that the conditional variances of $\hat{\mu}_n$ and $\hat{\mu}_{KM}$ are virtually the same. Thus, we conclude that the kriged mean does not outperform $\hat{\mu}_n$, for large sample sizes. However, the situation is a little different for non-random grid sampling design. According to Gombay (2010), an approximation for the variance is,

$$\text{VAR}[\hat{\mu}_{KM}] = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n a_j \gamma(\|x_i - x_j\|) + O\left(\frac{1}{n}\right).$$

Notice that,

$$\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n a_j \gamma(\|x_i - x_j\|) + O\left(\frac{1}{n}\right) \geq \min_{1 \leq j \leq n} \frac{1}{n} \sum_{i=1}^n \gamma(\|x_i - x_j\|) + O\left(\frac{1}{n}\right),$$

and $\sum_{i=1}^n \sum_{j=1}^n a_j \gamma(\|x_i - x_j\|)$ and $\sum_{i=1}^n \gamma(\|x_i - x_j\|)$ are approximating sums for $\int_D \int_D \gamma(\|u - v\|) du dv$ and $\int_D \gamma(\|x - u\|) du$, respectively. As a result, given that the integrals exist, the reduction of variance of using the kriged mean will not be larger than the difference between $\int_D \int_D \gamma(\|u - v\|) du dv$

and $\inf_{x \in D} \int_D \gamma(\|x - u\|) du$. In Section 4, we are going to calculate these two integrals under our model and compare them with the empirical variance obtained by simulation studies.

4 Simulation Studies

4.1 Generating a Dataset

In this section, we will give a brief description of the steps we used to generate the spatial data for the simulation studies that follows. We actually follow the unconditional simulation steps presented in section 3.10 of Diggle and Ribeiro (2007). The detailed R code can be found in the appendix.

Step 1 - *To obtain a simulated two-dimensional data, we generate two different vectors from the uniform distribution, $U(0, 1)$, with lengths equal to the sample size (n) desired, to represent the data locations. We then combine the two vectors together and form a $n \times 2$ location matrix.*

Step 2 - *Under the second-order stationary assumption, the covariance function depends only on the distances between two spatial locations; in this case we use Euclidean distance, although other kinds of distance measures can be used. After specifying the parameters for the correlation function, we can now compute the variance-covariance matrix.*

Step 3 - *Next, we apply diagonalization to the variance-covariance matrix; and obtain a diagonal matrix which contains the singular values in its diagonal. Then, we multiply this diagonal matrix by the corresponding eigenvectors. This results in a matrix, A , which we can treat as a linear transformation matrix.*

Step 4 - *We generate another vector, z with length equal to the sample size*

we used in step 1, its components are independent standard normal, $N(0, 1)$ random variables. By multiplying the matrix, A , obtained in step 3 with z , we simulated a vector of realization values for the underlying spatial signal, S . That is, $S = Az$.

Step 5 - Due to the second order stationary assumption which states that the mean is constant over the spatial region, we can get the realizations of the random variable y by adding a constant to the vector S in step 4.

By combining the $n \times 2$ location matrix in step 1 and the vector y , we now have a simulated two dimensional spatial dataset.

The dataset generated by the above five steps are used throughout the empirical studies done in Section 4. These include: (a) kriging the mean; (b) validating the built-in functions in R for performing simple kriging and ordinary kriging; and (c) computing the empirical variance of the estimators for various kriging procedures by simulation. Note that in order for the readers be able to reproduce the results presented in this thesis for (a) and (b), the seed 571 was used to generate the dataset.

4.2 Kriging of the Mean

We first generate a dataset using the steps outlined in the previous section. As mentioned before, for the sake of simplicity, we will use the powered exponential correlation functions to model the dependence structure in the imaginary spatial region, which is a unit square. In particular, the values of the parameters are: $\sigma^2 = 1$, $\phi = 0.25$, $\kappa = 1$ and $\tau = 0$. Furthermore, we will set the true

mean as 10.

To be able to write a function in R for implementing the method of kriging the mean, we will need to manipulate the matrices derived in Section 2. The manipulations are shown, followed by the results of the kriged mean, kriging variance and the weights for a sample of size $n = 100$.

Rewriting the matrices in page 13, with 1's omitted in the last row of the variance-covariance matrix, we have the system of equations as,

$$\begin{bmatrix} \gamma[S(x_1), S(x_1)] & \dots & \gamma[S(x_1), S(x_n)] \\ \vdots & \ddots & \vdots \\ \gamma[S(x_n), S(x_1)] & \dots & \gamma[S(x_n), S(x_n)] \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \lambda \\ \vdots \\ \lambda \end{bmatrix}.$$

Now, we can rearrange the equation to,

$$\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \gamma[S(x_1), S(x_1)] & \dots & \gamma[S(x_1), S(x_n)] \\ \vdots & \ddots & \vdots \\ \gamma[S(x_n), S(x_1)] & \dots & \gamma[S(x_n), S(x_n)] \end{bmatrix}^{-1} \begin{bmatrix} \lambda \\ \vdots \\ \lambda \end{bmatrix}$$

$$\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \lambda \underbrace{\begin{bmatrix} \gamma[S(x_1), S(x_1)] & \dots & \gamma[S(x_1), S(x_n)] \\ \vdots & \ddots & \vdots \\ \gamma[S(x_n), S(x_1)] & \dots & \gamma[S(x_n), S(x_n)] \end{bmatrix}^{-1}}_W \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}.$$

With $w_{ij} = \gamma[S(x_i), S(x_j)]$, we have,

$$\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \lambda \begin{bmatrix} \sum_{j=1}^n w_{1j} \\ \vdots \\ \sum_{j=1}^n w_{nj} \end{bmatrix},$$

and thus,

$$\lambda = \frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} = \frac{1}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}}.$$

With this new expression for Lagrange Multiplier, the kriged mean, kriging variance and the prediction weights can now be easily calculated. The results appear in the R console window are pasted below. One way to check whether the program is actually computing the kriged mean is by summing up the optimal prediction weights. Remember, that we have imposed the constraint, $\sum_{i=1}^n a_i = 1$, to come up with an unbiased estimator in kriging the mean. Therefore, this should hold as well in our example. The prediction weights are explicitly printed out below, and indeed, the sum of these weights is equal to 1.

```
>mean
      [,1]
[1,] 9.41205
>krige.var
[1] 0.1554925
```



```
>weights
```

```
[1] -5.106823e-04  1.114605e-02  4.336087e-02  2.643137e-03  1.153814e-03
[6]  1.727726e-03  3.723752e-04  1.675179e-02  2.171049e-02  1.632545e-02
[11]  2.883099e-02  3.351455e-02 -3.299657e-03 -1.996395e-03 -1.027529e-03
[16]  3.958129e-03  6.641148e-02 -8.503879e-04 -1.164438e-04 -3.749313e-03
[21] -8.065376e-03  1.106822e-03  7.873935e-04  3.186733e-02  2.769606e-02
[26]  3.255292e-03  2.693462e-03  6.341622e-03  1.644526e-02 -3.581716e-03
[31]  4.079638e-02  2.936754e-03  1.781537e-03  4.727978e-04  1.331523e-02
[36]  1.195609e-03  6.892881e-04  4.097841e-03  2.569735e-03  5.281232e-03
[41] -7.196993e-03  1.281802e-02 -6.073575e-03  3.981107e-03  2.529499e-02
[46]  3.332096e-03  5.986693e-03 -1.114503e-03  3.094430e-03  6.297633e-02
[51]  2.828627e-02  6.453335e-03  3.653306e-03 -6.375875e-04  2.427305e-04
[56]  2.362941e-03  1.346472e-03 -1.865804e-03  4.005762e-02  8.559120e-02
[61]  5.515757e-02  9.905833e-03  6.388065e-03  4.368503e-04  3.925253e-03
[66]  2.593894e-03 -2.776409e-04  4.393475e-02 -7.740037e-03 -2.552849e-05
[71]  3.765216e-03  3.581013e-03  3.138372e-03  3.318967e-03  2.002241e-03
[76]  1.381272e-03 -4.921192e-03  5.144090e-02 -2.804065e-03  4.750240e-03
[81]  3.978076e-02  2.314530e-03 -3.014454e-03  2.826213e-03  5.012584e-05
[86]  2.527591e-02  7.251327e-03  3.679866e-02  8.207566e-04  2.813179e-03
[91]  9.789059e-03  1.130029e-02  2.123862e-03  3.345970e-03 -3.018091e-04
[96]  3.449868e-04  3.806795e-03  9.680764e-03  5.906725e-03  2.506288e-03
```

4.3 Variances of the Kriged Mean

In this section, we are going to present the empirical variance of the kriged mean calculated under different sample sizes using Monte Carlo method. The

same set of parameter values as in the Section 4.2 were used, and datasets with sample sizes: 10, 50, 100, 200, 500, 600, 900 and 1000 were investigated. For example, we first generate a sample with sample size $n = 10$ and calculate its kriged mean. We repeat this step for $N = 10$ times, in other words, we generate 10 such samples and calculate the kriged mean of each sample. Then, we compute the variance of these 10 values and this is the empirical variance of the kriged mean. We do the the same for the other sample sizes just mentioned.

Table 1 shows the variances for different numbers of N , the number of simulations. We do not see the empirical variances decreasing when we increase the sample size. Rather, we see that it stays close to a value of 0.14 even if we increase $n = 500$ to $n = 1000$. To calculate the theoretical variance of the kriged mean, recall from equation (19),

$$\text{VAR}[\hat{\mu}_{\text{KM}}] = E\{\gamma(\|X_i - X_j\|)\} + \frac{1}{n}[\text{VAR}[S(X_1)] - E\{\gamma(\|X_i - X_j\|)\}].$$

Under our powered exponential model with $\phi = 0.25$ and $\kappa = 1$,

$$\text{VAR}[S(X_1)] = \gamma[S(X_1), S(X_1)] = \exp\left(\frac{-h}{0.25}\right) = \exp(0) = 1.$$

As,

$$E\{\gamma(\|X_i - X_j\|)\} = \int_0^1 \int_0^1 \exp\left\{\left(\frac{-\|x - y\|}{0.25}\right)\right\} dx dy,$$

we can use the “adapt” library in R to approximate this two-dimensional integral. Detailed R code can be found in the appendix. R gives the value of

0.09 for this integral and thus,

$$\text{VAR}[\hat{\mu}_{\text{KM}}] = 0.09 + \frac{1}{n}[1 - 0.09].$$

So depending on the sample size, the theoretical variance of the kriged mean is going to be different. As $n \rightarrow \infty$, the second term becomes negligible and the theoretical variance of the kriged mean approaches 0.09. That is,

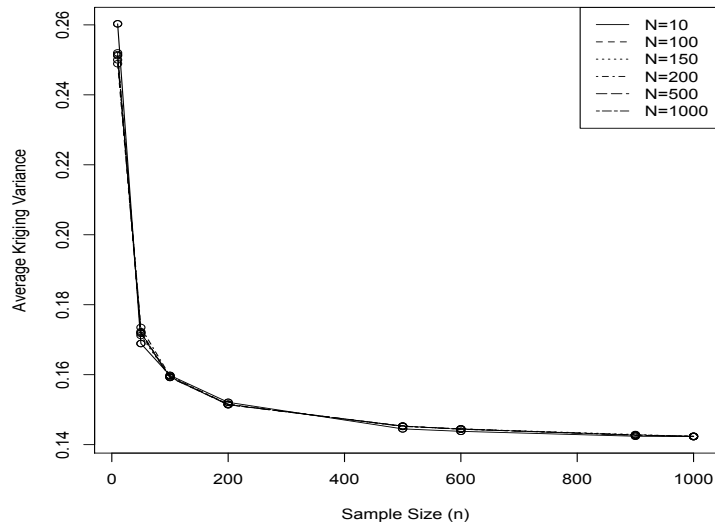
$$\text{VAR}[\hat{\mu}_{\text{KM}}] \rightarrow 0.09, \text{ as } n \rightarrow \infty.$$

The corresponding theoretical variance is shown in the brackets under each sample size in Table 1. Although the theoretical variance and the empirical variance are not exactly the same, they are very close to each other. More importantly, for both situations the variances do not decrease much when the sample size is large. Empirically, we see a difference of 0.10 when n is increased from 10 to 50, but only decrease by 0.0006 (the largest difference) when n is increased from 900 to 1000. Similarly, for theoretical variance it decreases by 0.08 when n is changed from 10 to 50, while only a difference of 0.0001 can be seen when n is changed from 900 to 1000. This indicates for a sufficiently large sample size, the variances, theoretical or empirical, approach a constant and do not decrease any further. We can then conclude that the kriged mean is not consistent, meaning that it is not possible to achieve a specific bound of error with any sample size if it is made smaller than the theoretical variance of the kriged mean.

Table 1: Theoretical and Empirical Variances for the Kriged Mean with Various Sample Sizes and Number of Simulations.

Number of Simulations (N)	Sample Sizes (n)							
	10	50	100	200	500	600	900	1000
	(0.1810)	(0.1082)	(0.0991)	(0.0946)	(0.0918)	(0.0915)	(0.0910)	(0.0909)
10	0.2603	0.1689	0.1598	0.1521	0.1445	0.1438	0.1424	0.1423
100	0.2489	0.1723	0.1592	0.1515	0.1453	0.1445	0.1427	0.1423
150	0.2520	0.1712	0.1593	0.1516	0.1452	0.1445	0.1429	0.1423
200	0.2500	0.1735	0.1596	0.1515	0.1452	0.1445	0.1427	0.1424
500	0.2515	0.1718	0.1592	0.1514	0.1453	0.1444	0.1427	0.1424
1000	0.2512	0.1720	0.1594	0.1514	0.1453	0.1444	0.1428	0.1424

Figure 5: Empirical Variances of the Kriged Mean versus Sample Sizes for Different Number of Simulations.



4.4 Properties of the Estimators in Kriging Procedures:

Simple Kriging and Ordinary Kriging

The plan of this section is as follows: First, the built-in function in R for performing simple kriging and ordinary kriging are investigated. The consistency of these functions with the theoretical results presented in Sections 2.3 and 2.4 are of vital importance for obvious reasons. Before utilizing such functions in simulations, the results of the kriging estimate made on the same location using a single dataset, along with its kriging variance and the prediction weights associated with each sample location produced by two methods are shown. After validating the functions used for both types of kriging, investigation on these two estimators can be done by simulations. Again, Monte-Carlo methods are used to calculate the empirical variances of each estimator.

4.4.1 Validating the functions

Unlike in the case of kriging the mean, Diggle and Ribeiro Jr. (2007) has developed the geoR package in R for predictions using simple kriging and ordinary kriging. Simple kriging and ordinary kriging can then be easily carried out by calling a function in R, after defining a proper covariance model and the prediction locations. However, the mechanisms behind these programming language is not made visible to the users in R. Also, according to Diggle and Ribeiro Jr. (2007), a different approach compared to the setup of this thesis was used in computing the kriging estimates for both types of kriging. Therefore, it is necessary to ensure that the results obtained by calling the R functions are consistent with the computations of the estimates in Section 2. The same set of covariance parameter values as in Section 4.2 are used to

generate a sample of size 10 for comparison purposes. Readers can find the detailed R code in the appendix and should be able to reproduce the results using the seed 571. For illustration purposes, the point (0.5,0.5) is chosen as the prediction location and the mean 9 is assumed to be the known mean. In what follows, the values of the kriging estimate, kriging variance and the prediction weights of each sample location are extracted from R. The output of simple kriging using the methods of Section 3 are presented first, followed by the output from the built-in function.

The output generated by following the way presented in Section 2.3 are:

```
> pred
      [,1]
[1,] 9.2956
> kriging.variance
      [,1]
[1,] 0.3712162
> weights
      [,1]
[1,] 1.043811e-03
[2,] 3.191851e-05
[3,] 1.118982e-03
[4,] 9.572313e-04
[5,] 5.567122e-02
[6,] 4.026260e-02
[7,] 5.617616e-04
[8,] 2.185532e-04
```

```
[9,] 6.301987e-01
[10,] 1.446616e-04
```

And the output generated by the built-in function are:

```
> sk$predict
      data
9.2956

> sk$krige.var
[1] 0.3712162

> w
[1] 1.043811e-03  3.191851e-05  1.118982e-03
[4] 9.572313e-04  5.567122e-02  4.026260e-02
[7] 5.617616e-04  2.185532e-04  6.301987e-01
[10] 1.446616e-04
```

One can immediately conclude that the two results are exactly the same. Therefore, it is justified to use the built-in function for performing simple kriging, and it can be used in simulations to examine the properties of the estimator.

In a similar fashion, the built-in function for performing ordinary kriging is checked. The same dataset is used and the results of the kriging estimate at the point (0.5,0.5), kriging variance and the prediction weights for each sample location is shown below.

The output generated by following the way presented in Section 2.4 are:

```
> pred
      [,1]
```

```

[1,] 9.333303
> kriging.variance
      [,1]
[1,] 0.323121
> weights
[1] 0.02598289 0.02951080 0.02604279
[4] 0.02875157 0.08721244 0.06017567
[7] 0.03588410 0.02333283 0.65610963
[10] 0.02699727

```

And the output generated by the built-in R function are:

```

> ok$predict
      data
9.333303
> ok$krige.var
[1] 0.323121
> w
[1] 0.02598289 0.02951080 0.02604279
[4] 0.02875157 0.08721244 0.06017567
[7] 0.03588410 0.02333283 0.65610963
[10] 0.02699727

```

Again, the two results agree with each other and thus it is safe to use the built-in function to perform ordinary kriging in simulations. The properties of the estimators, specifically, the variance of the two estimators used in these two kriging procedures can now be explored.

The prediction weights with their corresponding locations are shown in Table 2,

Figure 6: Circle Plot of Ordinary Kriging Prediction Weights. The target point at $(0.5, 0.5)$ is represented by a black circle. Circles are plotted with centres at the measurement locations with diameter representing the magnitude of prediction weights.

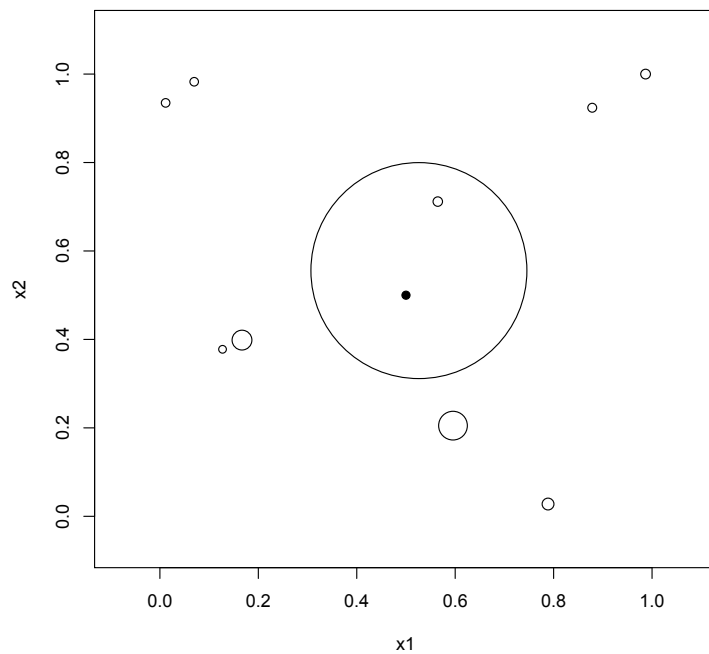


Table 2: Ordinary Kriging Prediction Weights with their corresponding locations.

Location Coordinates		Prediction Weights
x_1	x_2	
0.06966785	0.98261857	0.02598289
0.98683762	0.99997644	0.02951080
0.01162989	0.93485441	0.02604279
0.56463547	0.71166349	0.02875157
0.59553526	0.20503937	0.08721244
0.16680966	0.39845670	0.06017567
0.78866925	0.02777584	0.03588410
0.12726980	0.37765095	0.02333283
0.52627892	0.55562496	0.65610963
0.87845702	0.92386763	0.02699727

and a circle plot is drawn in Figure 6. Generally speaking, observations which are farther away from the target point, (0.5,0.5) shares a smaller weight. We can see that the observation taken at (0.5263,0.5556), which is closest to the target point, possesses the heaviest prediction weight. Due to the presence of this observation, the measurement taken at (0.5646,0.7117) does not give us much more information about the characteristic value at the target point and thus it is only given with a small weight.

4.4.2 Variances of the Estimators

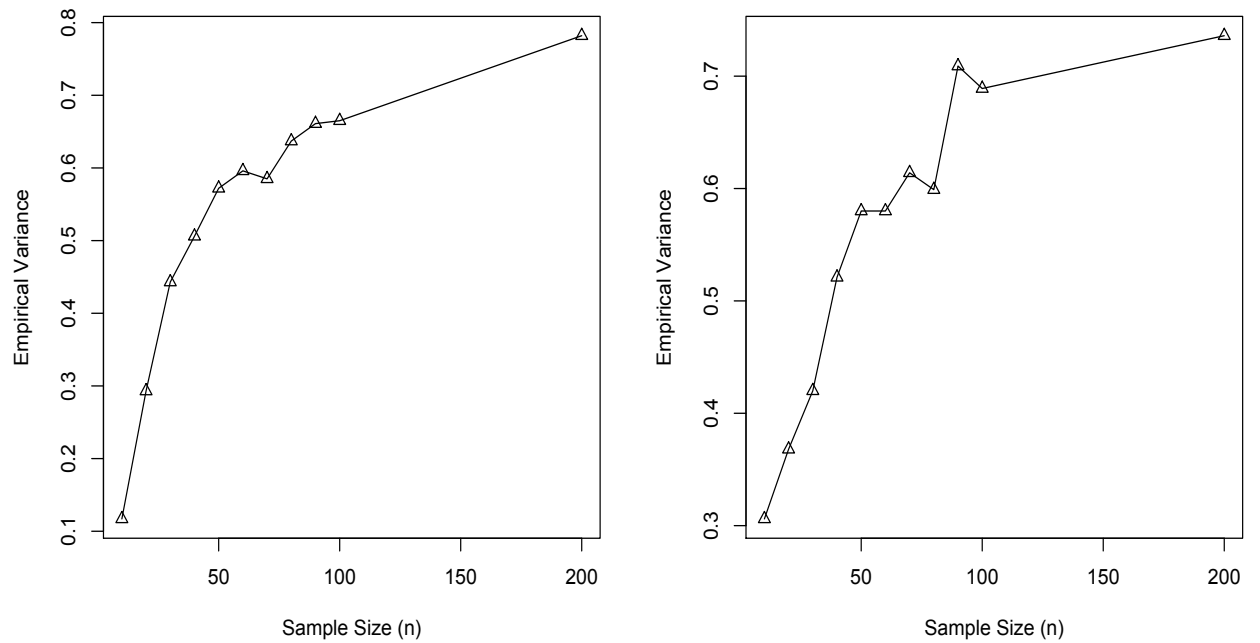
The main aim of this section is to find the empirical variances of the simple kriging and ordinary kriging estimators by simulation studies. Similar to what was done in Section 4.3, where the empirical variances of the kriged mean is computed, a sample with certain number of observations are first generated. Due to computational feasibility, smaller sample sizes and fewer number of simulations are used compared to the ones in Section 4.3. Nevertheless, these

Table 3: Empirical Variances of Simple and Ordinary Kriging Estimator, with $N = 1000$.

Sample Size	Empirical Variance	
	Simple Kriging	Ordinary Kriging
10	0.117	0.306
20	0.293	0.368
30	0.443	0.420
40	0.506	0.521
50	0.572	0.580
60	0.596	0.580
70	0.585	0.614
80	0.637	0.599
90	0.661	0.709
100	0.665	0.689
200	0.782	0.736

are sufficient to gain an idea of the trend in the variances of the estimator. For each sample, the kriging estimate at the point $(0.5, 0.5)$ is computed for convenience purpose. Then, $N=1000$ samples with the same sample size are generated and the kriging estimate is calculated for each sample. The variance of these estimates are found and reported in Table 3, and this is the empirical variance of the kriging estimator. The above procedure is repeated for other sample sizes, up to the size of 200.

Figure 7: The left panel plots the empirical variances of the simple kriging estimator and the right panel plots the empirical variances of the ordinary kriging estimator, both uses the information obtained from Table 3 of the simulation study.



According to the above results, the variance of the estimator neither converges to zero nor decreasing to a constant. Instead, it keeps increasing when the sample size is increased. Whether it will approach to a constant for a large enough sample size is of question, and due to computational power, it is not feasible to examine the empirical variance of the estimator beyond the sample size of 200 in both cases. To conclude this section, one can at least say the kriging estimators, either simple or ordinary, are not consistent.

4.5 Random Sampling Designs Vs. Non-Random Sampling Designs: Comparison of the Sample Mean and the Kriged Mean

Besides determining the sample size, choosing an appropriate sampling design is another important problem to consider before data collection. It is important to know how the design affects the predictions for a given region. In geostatistics, there does not exist an optimal design because it would depend on the particular application. Rather, knowing the advantages and drawbacks of the different designs can help us to decide which better fits the situation. Generally speaking, we can classify geostatistical designs into two categories: *completely random* designs and *lattice* designs (Diggle and Ribeiro Jr., 2007). Figure 8 is a schematic representation showing the difference of the two designs. The sampling designs have influences on both parameter estimations and spatial predictions, the two common objectives in geostatistics; and now, we will discuss their advantages and disadvantages under each objective.

To estimate the covariance parameters more accurately, we need to have a compromise between the two designs. For example if we wish to estimate the

nugget variance, sufficiently small enough distance between two sample points is required in order to detect this variation. The lattice design may not be able to detect this effect. On the other hand, a random design may not be able to capture the whole range of spatial variations due to the close proximity of the sample points in some cases. Sometimes there would be a problem for the random design to detect directional effects as well.

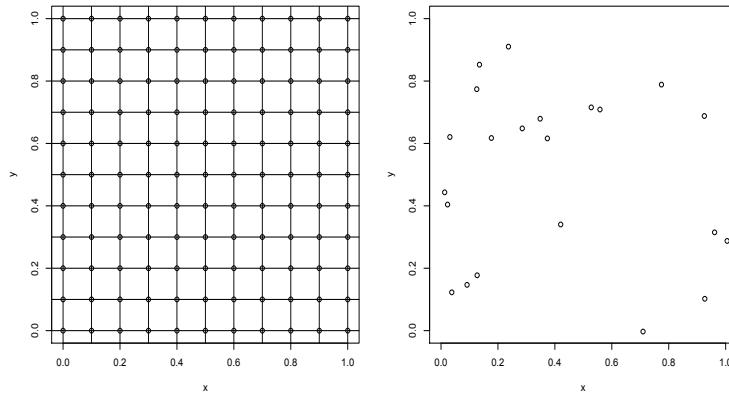
Usually, we judge the performance of an estimator by its variance; and actually, when the objective is to make spatial predictions, we can also use this criterion to investigate the two sampling designs. Diggle and Ribeiro Jr. (2007) did a simulated experiment and found that the lattice design is superior than the random designs. The results can be found on page 205 in the book. This is not surprising because theoretical calculations in Gombay (2010) also showed that the asymptotic variance of the kriged mean is lower in the case of non-random designs. In the remaining of this subsection, simulation results on the asymptotic variance are presented and also the variance of the sample mean under each design is computed for the comparison of the predictive performance of the two estimators.

To illustrate, we will restrict the spatial region to be a unit square. For completely random designs, realizations are obtained from diagonalization of the covariance matrix calculated by using the information of the random locations, generated by the uniform distributions, using the steps presented in Section 4.1. For lattice designs, instead of using the uniform distribution to generate the random locations, we create a regular grid by dividing the unit square into equal portions according to the respective sample size. Realizations are then obtained by the same way as in the random designs. Again, the same set of covariance parameters values as in Section 4.2 are used here. 2000 samples are

then simulated for each sample size. Attention should be paid to two aspects here: (a) the variance of the sample mean and the kriged mean under the same design; and (b) the variance of the kriged mean in the completely random and lattice designs. Table 4 summarizes the results of the simulated experiment for the random design.

We can see that the empirical variance of the kriged mean is lower than that of the sample mean, meaning that the predictive performance is higher in kriging. The same is true in the lattice design settings, and the results are shown in Table 5. It should be noted here that for kriging of the mean under the random setting, the variance does not decrease further with increasing sample size and in fact, stays at the value of around 0.14 for our model. However, under the regular grid setting, even for $n=49$ the empirical variance of the kriged mean is 0.1346, which is already smaller than the minimum variance it can be achieved in the random design. And for $n=1600$, its variance decreases to be around 0.13; meaning that the asymptotic variance is lower for the non-random design. In fact, by looking at the simulation results, the sample mean has a smaller asymptotic variance under the non-random design as well. Furthermore, for small sample sizes such as $n=16$, the variance of the kriged mean is 0.1588 while it is 0.206 under the random design. Remember from Section 3 that the reduction of variance cannot be larger than the difference between $\int_D \int_D \gamma(\|u - v\|) du dv$ and $\inf_{x \in D} \int_D \gamma(\|x - u\|) du$. From Section 4.3, we have already calculated $\int_D \int_D \gamma(\|u - v\|) du dv$ to be 0.09 for our model. By using the “integrate()” function in R, the one-dimensional integral, $\inf_{x \in D} \int_D \gamma(\|x - u\|) du$ can be easily computed. Since we are looking at the unit square, therefore the integral needs to be computed becomes, $\inf_{x \in [0,1]} \int_D \gamma(\|x - u\|) du$. Because we are using the exponential model and the exponential covariance function is a

Figure 8: The left panel shows the lattice sampling design on the $[0,1]$ grid and the right panel shows the completely random sampling design on the $[0,1]$ grid.



monotone decreasing function, so the infimum of the integral should exist at $x = 1$. A sequence of values between 0 and 1 were generated in R and the integrals were calculated. Indeed, the minimum value occurs at $x = 1$ and it is 0.0114. The reduction of variance thus cannot be larger than the difference between 0.09 and 0.0114, which is 0.0786. This is verified in the simulation study, as shown in Table 5. From practical point of view, it is common to have a small dataset for analysis. Therefore, under such circumstance, the prediction of the mean would be more precise if a regular grid is used as the sampling design.

Lattice design also has another advantage over the other design. It is easy for implementation. For example, suppose that in an agricultural study, an experimenter is interested in knowing the number of insects present in a farm. It would be difficult to take measurements at the location that we have randomly chosen prior to the data collection step; as identifying the location in

the field is not a simple job. Instead, if we are planning to use the regular grid sampling design; then an experimenter could easily obtain the measurements at the pre-specified locations by knowing the distance that he has walked for, after taking the measurement at the first random location. In conclusion, lattice designs are more useful from practical perspective because of the feasibility of implementation and it is more statistically efficient when we are trying to estimate the mean of a spatial region.

Table 4: Empirical Variances for the Sample Mean and the Kriged Mean under the Random Design.

Sample Size	Empirical Variance	
	Sample Mean	Kriged Mean
10	0.274	0.249
20	0.236	0.206
30	0.218	0.185
40	0.220	0.182
50	0.208	0.171
60	0.208	0.168
70	0.204	0.165
80	0.211	0.168
90	0.203	0.163
100	0.199	0.157
500	0.293	0.143
1000	0.195	0.142

Table 5: Empirical Variances for the Sample Mean and the Kriged Mean under the Non-Random Design.

Sample Size	Empirical Variance	
	Sample Mean	Kriged Mean
16	0.1689	0.1588
49	0.1570	0.1346
100	0.1745	0.1446
225	0.1889	0.1439
400	0.1735	0.1356
625	0.1846	0.1377
900	0.1853	0.1441
1600	0.1828	0.1306

Figure 9: Random Design: Empirical Variances of the Sample Mean and the Kriged Mean for N=10000 Simulations.

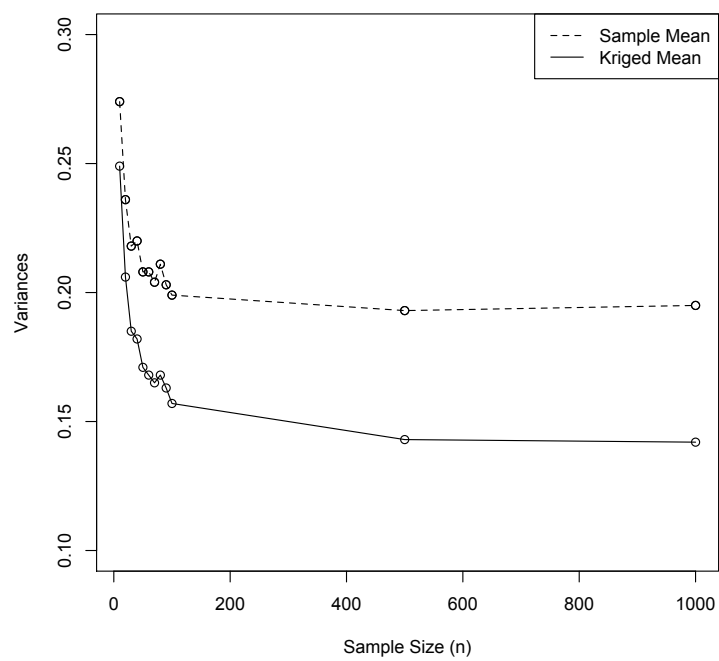
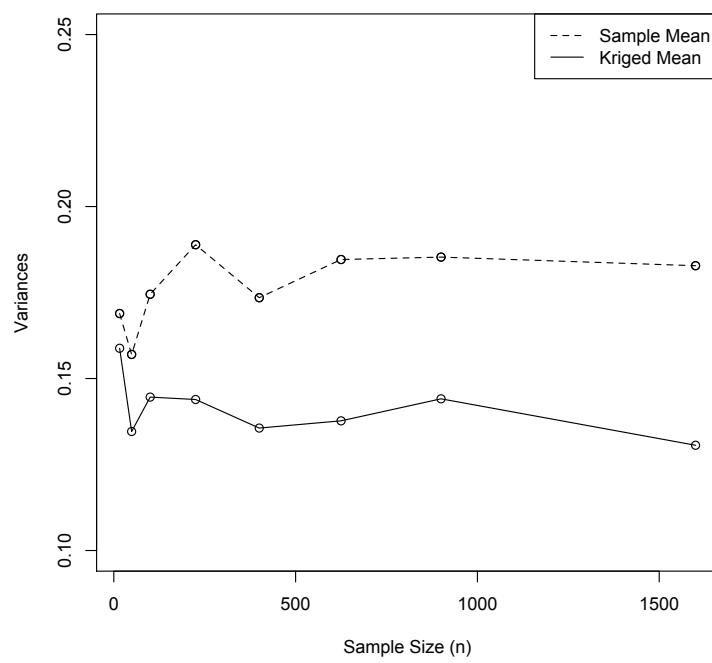


Figure 10: Non-Random Design: Empirical Variances of the Sample Mean and the Kriged Mean for N=2000 Simulations.



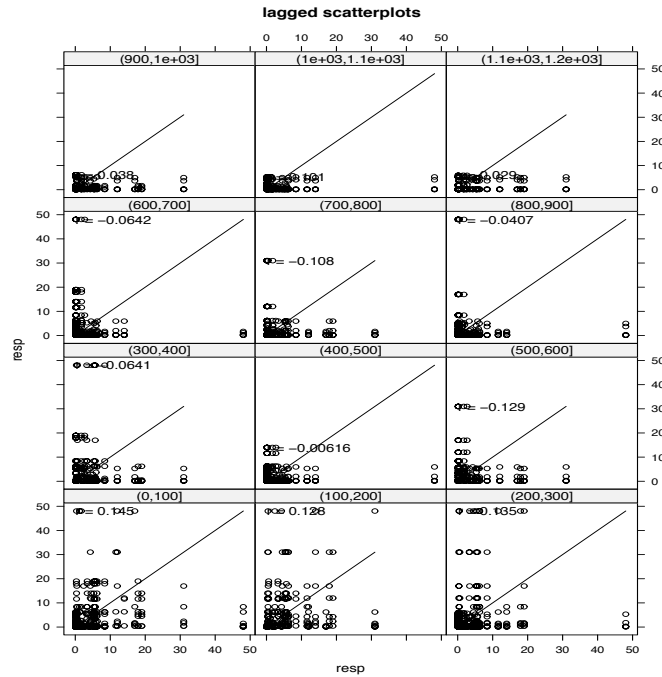
4.6 Example 1 - Dioxin Data Analysis

4.6.1 Background and Preliminary Analysis

An environmental dataset is going to be analyzed in this section. The dataset is obtained from the article, “Geostatistical Analysis of Hazardous Waste Site Data”, from the Journal of Environmental Engineering. In 1971, a truck driver dumped an unknown quantity of dioxin (2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD)) contaminated waste in a rural area of Missouri. Because there is a state highway nearby the original dumped site, so the U.S. Environmental Protection Agency believe that vehicles driven by might have transported the contaminated soil to other areas. An investigation was called in 1983 to determine the areas of the highway shoulder requiring cleanup. Those areas with TCDD concentration higher than $1 \mu\text{g}/\text{kg}$ will be identified as contaminated sites, and remedial actions will be taken. Measurements of TCDD concentrations were made on both sides of the highway, up to a distance of 30 feet away from the highway shoulder and samples were taken at every 10-feet. A certain starting point in the highway was chosen by the investigator and measurements were made for every 200-feet (or sometimes 50-feet or 250-feet) until it is 3555-feet away from the starting point. This results in 127 observations and the exact locations, along with the TCDD concentrations are recorded. In this manner, the sampling design is close to a regular grid; and we will treat each observation as a point sample on the grid.

The second order intrinsic stationary assumption seems to be reasonable here since the polluted soil were transported away by the vehicles using the highway, somewhat similar to the diffusion process. Therefore, the constant mean assumption is justified. Using the `summary()` function of a geostatistical dataset,

Figure 11: Lagged Scatterplot for the Dioxin Data.



we found out that the largest separation distance between a pair of points is 3555-feet. With the lagged scatterplots, which breaks up the separation distances into different groups and plot all the pairs of points within each particular lag distance, it enables us to examine whether sample correlation exists in the data. We will look at the lags up to 1200-feet, one-third of the maximum separation distance. Figure 11 shows the lagged scatterplot of the dioxin data. We can see that the points are clustered together on the bottom-left corner of each scatterplot, so we take the log of the TCDD concentrations.

Figure 12 clearly shows correlations exist between the measurements, as the sample correlations are 0.362 and 0.407 for groups $(0, 100]$ and $(100, 200]$, respectively. By further reducing the lag distance in each group to 50, we can

Figure 12: Lagged Scatterplot of Log-Transformed Dioxin Data, with lags=100 in each group .

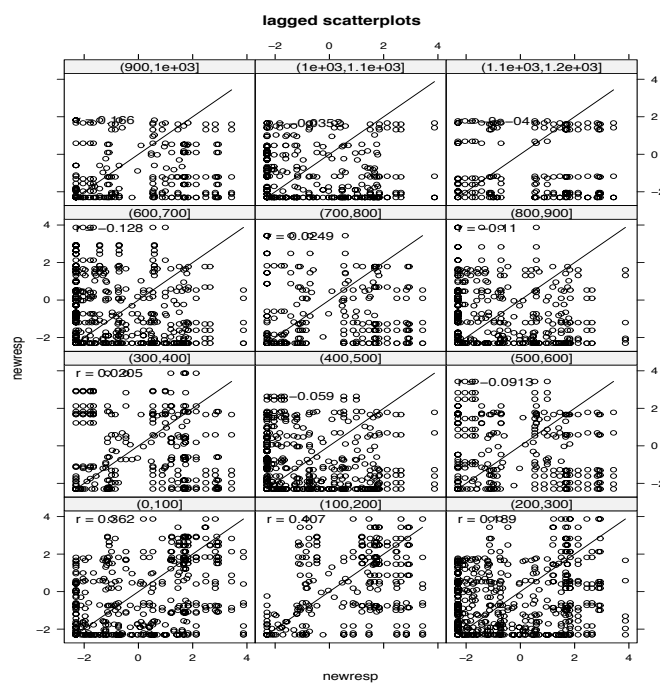
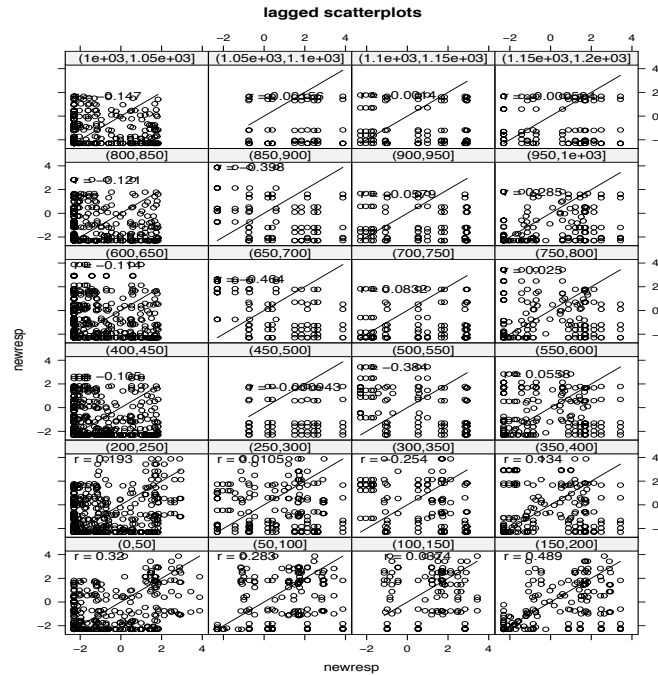


Figure 13: Lagged Scatterplot of Log-Transformed Dioxin Data, with lags=50 in each group.



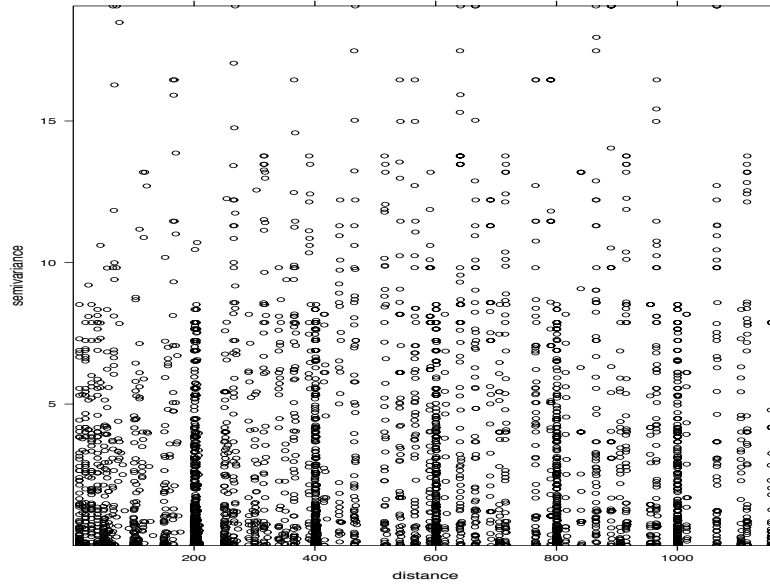
see from Figure 13 that some moderate sample correlation values still exist in several groups. As a result, we are confident that correlations exist in the data and kriging can be applied.

4.6.2 Variogram Modeling and Kriging

The variogram cloud is shown in Figure 14. Note that this variogram cloud is made up of $\frac{n(n-1)}{2}$ differences. Since there are 127 observations, so 8001 points will be found. The author of the paper mentioned for some locations, the TCDD concentrations were actually below the detection limit of the instrument, and the experimenter used $\frac{1}{2}$ of the detection limit value to be the

observed value for those particular locations. This might exaggerate the difference of the realization values for a pair of locations and thus causing a nugget effect due to measurement error. Indeed, this is evident in the variogram cloud. For small lag distances, we can see the inhomogeneity of points, confirming the reason of the existence of nugget effect; as otherwise the lagged scatterplot will not show moderate sample correlations for small lags. Not surprisingly, the experimental variogram shows the presence of the nugget effect. To model the dependence structure, we choose from a class of available covariance models. The most commonly used covariance families in geostatistics are: exponential, Matérn and spherical. To visually look at the shape of the covariance functions, with different values of parameters, one can call out the `show.vgms()` function in the `gstat` package. Here, we will use the exponential model. Actually, it is hard to find the estimates of the parameters as we are fitting a non-linear function to the experimental data. In classical non-linear regression, choosing appropriate starting values for numerical optimization is a critical step. Whether sensible estimates can be obtained and the number of iteration steps needed depends on these values. This same problem occurs in fitting the variogram as well. As there is no general rule for choosing such values, one should be cautious in using the results from R. Some estimates will be given even when one uses some insensible starting values, so whether the resulting estimates make practical sense to the experimental data is of question. Several initial values have been tried in this case: 3.5 for the partial sill, 300 for the range and 1 for the nugget seem to work reasonably well. The output from R is given below:

Figure 14: Variogram Cloud of Dioxin Data.



model	psill	range
Nug	2.106748	0.000
Exp	1.314107	819.613

The left panel of Figure 15 shows the experimental variogram and the right panel plotted the fitted line together with the experimental variogram. We believe that this is a reasonable fitting as we can see, for large distances the semivariances begin to decrease. This is usual in practice as the correlation between two locations becomes less evident when the separation distance is large, and it causes some strange behaviour of the variogram.

With the estimated variogram, we can now do kriging. Recall that at the beginning we assumed that the mean is constant, so kriging of the mean can be done. We have also calculated the sample mean to compare the two estimates. From R, the sample mean is -0.57 with variance 2.8775; while the kriged mean

Figure 15: Experimental and Theoretical Variogram of Dioxin Data.

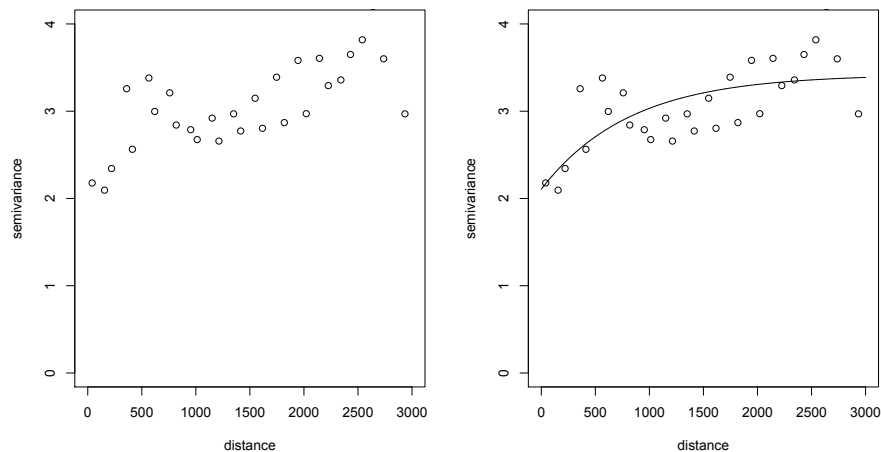


Figure 16: Ordinary Kriging Predictions of Dioxin Data.



is -0.9503 with kriging variance 0.4397. Clearly, the kriged mean is superior than the sample mean in the non-random design setting, with a smaller kriging variance; which is consistent with both theoretical and simulation results.

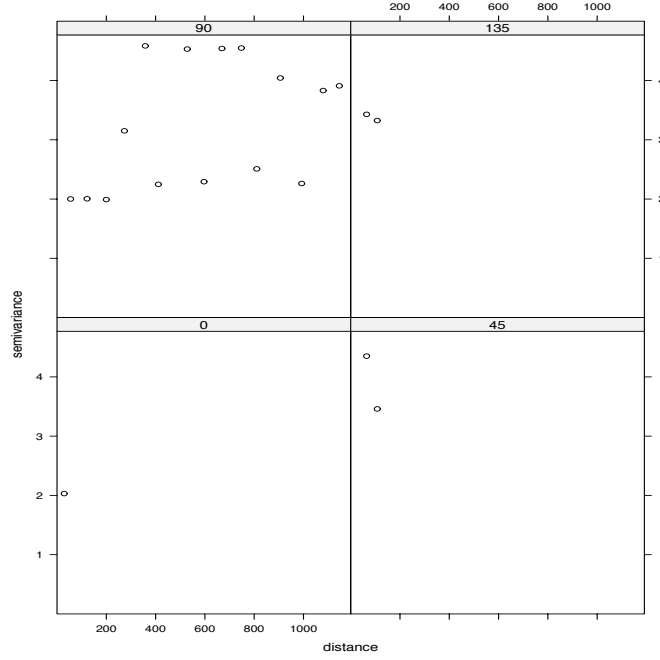
The main purpose of this investigation is to find the areas which are contaminated by dioxin residues and require cleanup. As a result, ordinary kriging should be performed for the locations where observations have not been made. Recall that, from Section 2.4, the kriging variances are the same when we do simple kriging with kriged mean; or ordinary kriging directly. Therefore, we will proceed with ordinary kriging in R using the `gstat` package. Figure 16 shows the spatial predictions of the region, using the same boundaries of the dataset. To identify the areas which needs to be cleaned up, we need to see the part of the region which has the TCDD concentration higher than $0 \mu\text{g/kg}$, under the transformed scale. By looking at Figure 16, we can see that the areas which are not far away from the starting point of the measurements are the most severely polluted areas, and require the remedial actions.

4.6.3 Model Diagnostics

The validity of the second order intrinsic stationary assumption is very important in this analysis. This assumption is composed of two parts: (a) the mean is constant in the spatial region; and (b) the covariance only depends on the magnitude of distance between the observations. We believe that the nature of the region under investigation has already justified the first part of the assumption, that is, the diffusion process carried out by vehicles using the highway averages out the TCDD concentrations of the whole region.

We can check the second part of the assumption by directional variograms. Defining the four directions to be: N-S direction (0 degrees), NE-SW direction

Figure 17: Directional Variogram of Dioxin Data.



(45 degrees), E-W direction (90 degrees) and NW-SE direction (135 degrees) with a tolerance angle of 22.5 degrees; the directional variograms are shown in Figure 17. We can see that except for the E-W direction, there were not many other sample points which are available for estimation of each particular variogram. The lack of sample points is mainly due to the sampling design of the study. With this constraint on the dataset itself, the experimental and theoretical variogram computed earlier is sufficient for predictions. We claim that the assumption is satisfied.

Besides checking the validity of the assumptions, based on the given data, we can also evaluate how good our variogram model is. Cross validation is a common method for comparing models. Usually, the leave-one-out cross

validation procedure will delete one observation from the whole dataset, and then estimate the parameters by the remaining observations. The residual between the value predicted by the model developed using the covariates for the deleted observation and the observed value for the same observation is calculated, with this procedure applied to every observation in the dataset. And finally the average of the residuals is calculated and it is compared to other choices of models. However, in our case, only one variogram model is computed from the whole dataset; so a slight modification of the method is used. The residual value between each observed value and the predicted value for the same observation is calculated. Each residual is then being standardized by its kriging standard error. The resulting value is called the cross validation z-score in R. Using the `krige.cv()` function in the `gstat` package, the z-score for each observation will be computed automatically. If the mean and variance for this z-score is 0 and 1, respectively, then we can conclude that the variogram model is good for the given data. The cross-validation z-score was calculated for the variogram model which we have used for kriging. The mean of the z-score is -0.0024 and the variance is 0.9613, therefore we can conclude that the theoretical variogram we have used is a reasonable fit.

4.7 Example 2 - House Prices Data Analysis

4.7.1 Introduction and Preliminary Analysis

At the beginning of this thesis, we have mentioned the kriging techniques developed in geostatistics can also be applied to other fields of studies. An environmental dataset, which is a common application of the kriging techniques, was analyzed in the Section 4.6. We now shift the attention to analyze an

economical dataset. Housing prices in Baltimore, United States are recorded and we wish to make predictions on the housing prices in a particular neighbourhood. The usual two-dimensional coordinate system is used to identify the locations of the houses, and the price of each house was recorded. In this case, it would be reasonable to assume that the prices of houses are correlated with each other: with nearby houses selling for a similar price, and houses which are farther apart have a larger discrepancy of prices.

To visually examine the data, the `plot()` function for a geostatistical dataset is used. From Figure 18, we see that the observations which are below the 1st quartile, depicted by the circles, are concentrated in the middle region and the bottom right part of the plot; while the observations which are above the 3rd quartile, depicted by the crosses, are concentrated on the top left corner of the plot. Figure 18 also shows the data values plotted against the two location coordinates separately, and a histogram. The histogram, however, does not give us too much information since we are dealing with correlated data here. Remember that the second order stationary intrinsic assumption is associated with the kriging analysis. However, it seems that the constant mean assumption would not be valid for this kind of data. One way to relax this assumption is to include the spatial trend in the model, and usually the two coordinates will be used as the covariates. It should be noted here that there is no physical meaning of the two coordinates in the model, they are used to describe the trend, but they are not able to provide an explanation. Nevertheless, the first order spatial trend will be used in the analysis.

Figure 18: A Glance of the Data. The top left panel shows the locations of the observations according to the quartiles they belong to. The top right and bottom left panel plots the observations against the y and x coordinates separately, with the lowess curve added. The bottom right panel is a histogram of the data values.

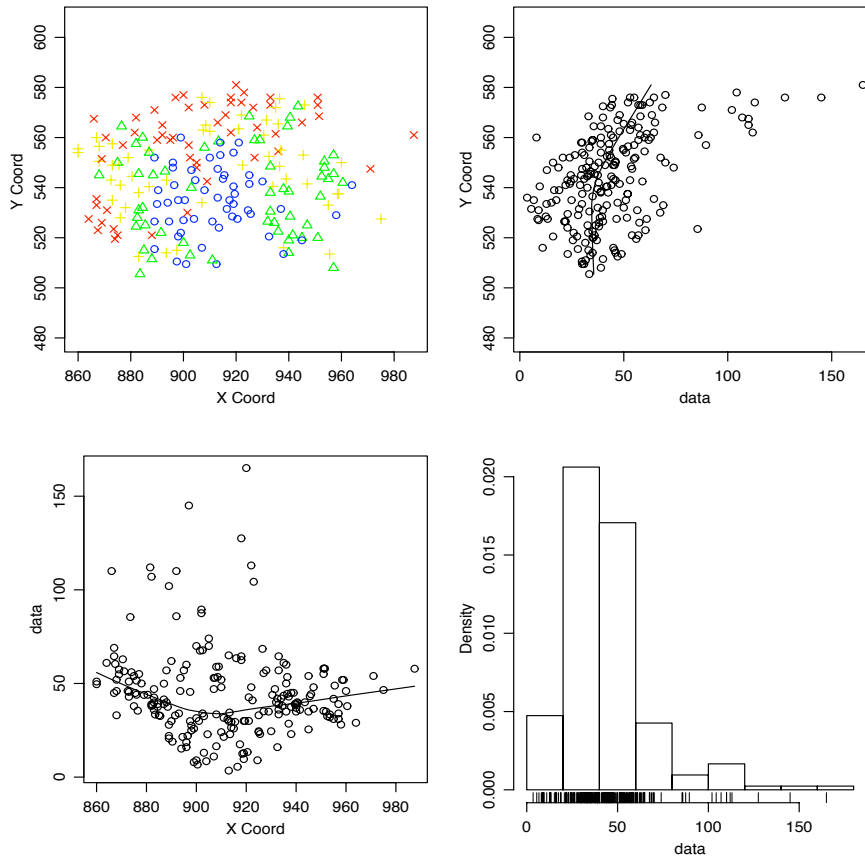


Figure 19: Variogram Cloud of the Housing Prices Data.

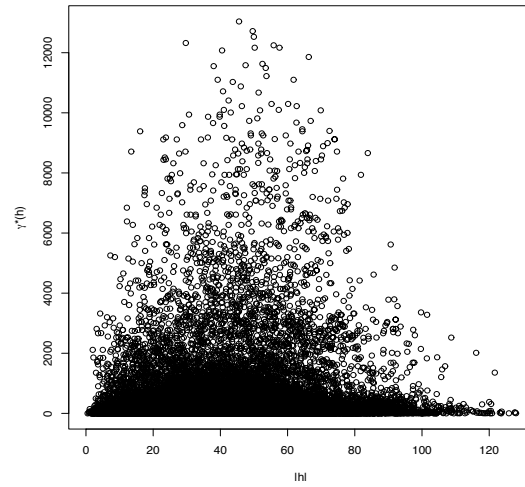
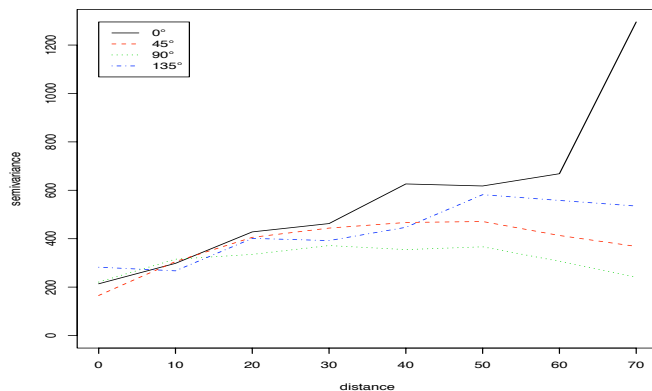


Figure 20: Directional Variogram in untransformed location coordinates of Housing Prices Data.

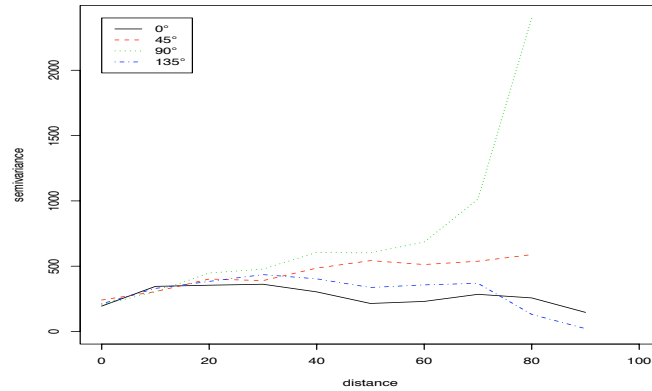


4.7.2 Variogram Modeling and Spatial Predictions

A variogram cloud is shown in Figure 19. The dissimilarities actually diminish when $|h|$ is greater than the value of 80. This might be due to the distance between the two points, as if too large it does not have much contribution in estimating the dependence structure. This is of particular importance when looking at the experimental variogram, as a sharp decrease on the average dissimilarities will be found when the distance is beyond 80. Therefore, we can restrict our attention for semivariances below this value.

To examine whether directional effects exist in the data, directional variograms are drawn in Figure 20. Obviously, we can see that nugget and sill anisotropy exist in the data. Therefore, we first tried to transform the original coordinates from the anisotropic space into an isotropic space, by specifying the anisotropy

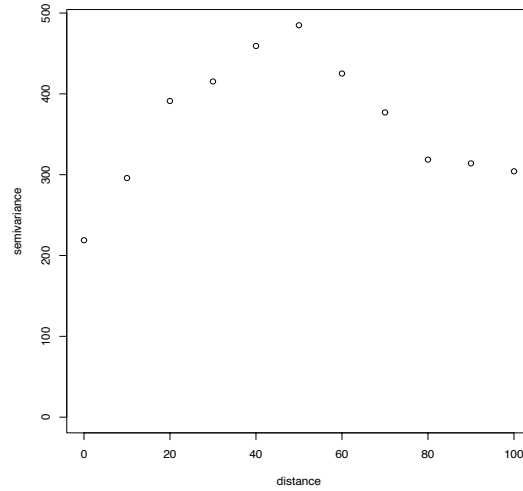
Figure 21: Directional Variogram in transformed location coordinates of Housing Prices Data.



angle and ratio mentioned in Section 2.5. Usually, these two values are unknown and being treated as additional parameters to be estimated. Instead of this approach, we use the trial and error approach to find out the “best” estimates of these parameters. Directional variograms are then used to justify whether the estimates are good or not. Different values have been tried and it turns out that the anisotropy angle equal to $\frac{3\pi}{2}$ and the anisotropy ratio equal to 1.5 work well. The directional variograms are shown in Figure 21.

In this figure, we can see that the problem of nugget anisotropy are corrected, as all variograms start at the same point. We should be cautious about the unusual behaviour of the two directional variograms for NW-SE direction and E-W direction. Starting at $h=30$, the semivariances begin to decrease and we therefore take 30 as the approximate range for the four directions, as the sill was almost reached at that point. With this range value, the value of the sill

Figure 22: Experimental Variogram in isotropic space of Housing Prices Data of Housing Prices Data.



becomes approximately 500 for the four directions.

With these newly transformed location coordinates, we proceed our analysis as in the isotropic space. The experimental variogram is shown in Figure 22. We use this variogram computed from the data to obtain initial estimates for the pararmeters of the exponential covariance function. The initial estimates used are as follows: (a) $\tau^2 = 200$, (b) $\phi=50$ and (c) $\sigma^2=500$. Maximum likelihood estimation is used by the `likfit()` function in the `geoR` package, with the estimates shown below:

tausq	sigmasq	phi
159.5461	359.8294	16.7403.

Now, we are ready to make spatial predictions by kriging. We chose a regular grid running from 500 to 590 for the horizontal direction and -660 to -570 for the vertical direction, under the transformed coordinates. The predictions are shown in the contour plot. The left panel of Figure 23 shows the prediction values in the grid. Higher housing prices are located on the right side of the plot, where lower housing are concentrated middle of the plot. The right panel of Figure 23 added the original data on the same contour plot, with larger circles represent a higher price at that location. We can see that the pattern of the observed data is the same compared to the prediction values. To immediately see the pattern of the whole spatial region, a plot using the `persp()` function in R is drawn in Figure 24. With this picture, one can find out the location of lowest and highest housing prices in Baltimore, by transforming the coordinates back to the original scale.

Figure 23: Contour Plot of predictions. The left panel shows the prediction values. The right panel shows the prediction values along with the original data values.

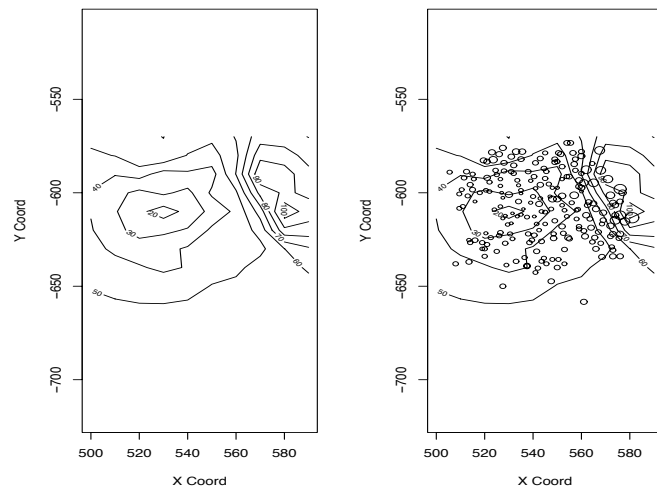
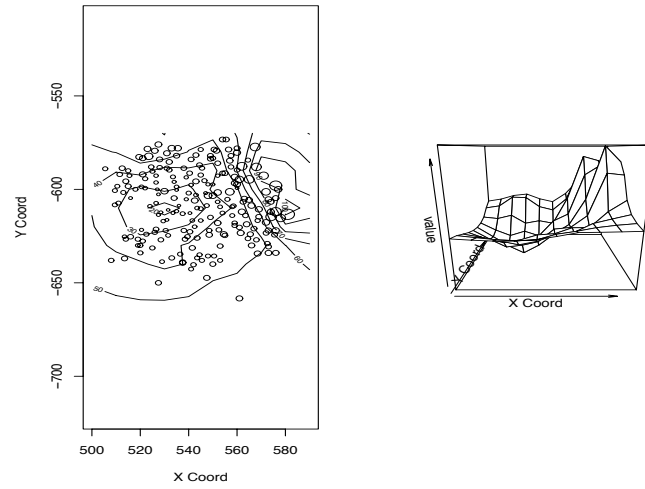


Figure 24: Different Plots for showing the predicted values.



5 Conclusions

This thesis has basically arrived to the conclusion that we are not able to calculate a minimum sample size, if kriging of the mean is used to estimate the mean of the spatial region. As shown theoretically in a manuscript by Gombay (2010) and by the simulation studies done in this thesis, the kriged mean is not consistent. In other words, if the desired bound of error is made to be smaller than the asymptotic variance of the estimator, there is no sample size which is going to achieve that particular bound. Then, the thesis extends its scope and investigate whether the same situation occurs for the estimators used in simple kriging and ordinary kriging. It turns out that, by simulation studies, the empirical variance of these two estimators keep increasing when the sample size increases. Although the variances do not approach a constant in either of these cases, it verifies the fact that determining the minimum sample size required to achieve a specific bound of error is not an easy question in geostatistics. Even though we do not have a definite answer for the question of minimum sample size needed for a particular application, we are able to draw the conclusion that the predictive performance of the kriged mean is higher under a non-random sampling design. It was shown by the empirical studies that for a similar sample size, its variance is lower under the non-random design. In fact, we see that for some small sample sizes, the variance is below the constant which the kriged mean is approaching in the random design. Therefore, we conclude that the non-random design is superior than the random design, from both statistical and practical perspective. Further, the properties of the sample mean was also investigated under the spatial setting, and comparisons have been made to the kriged mean under the two

sampling designs. We end the thesis by two data analyses, where one of the datasets satisfy the second order intrinsic stationary assumption and the other one violates this assumption.

Throughout the development of geostatistical techniques in mining industry, much of the researchers' attention were towards variogram modeling. There is no doubt that estimation of the parameters is important, but we should also invest some time to investigate the properties of the estimators; as estimations and predictions are the two main branches in Statistics. Going in this direction, our last word in this thesis is that it is worth doing kriging under the spatial setting.

A R code

A.1 The Drawing of the Theoretical Variogram

```
x = seq(0,5,length=600)

curve(exp(x)/(1+exp(x)),from=0,to=10,xlab="h",ylab="V(h)",ylim=c(0.5,1.1))

abline(h=1,lty=2)

abline(h=0.5,lty=2)

abline(h=0.975,v=3.7,lty=3)

text(x=8,y=0.52,labels="nugget")

text(x=5,y=0.7,labels="practical range")

text(x=8,y=1.01,labels="sill")
```

A.2 The Drawing of the Range Anisotropic Directional Variogram

```
x = seq(0,6,length=70)

curve(exp(x)/(1+exp(x)),from=0,to=6,xlim=c(0,6),ylim=c(0,2),lty=6,xlab="|h|",ylab="V(|h|)")

y1 = exp(2^x)/(1+exp(2^x))

lines(x,y1,lty=5)

y2 = exp(3^x)/(1+exp(3^x))

lines(x,y2,lty=2)

y3 = exp(5^x)/(1+exp(5^x))

lines(x,y3,lty=1)
```

A.3 The Drawing of the Sill Anisotropic Directional Variogram

```
x = seq(0,6,length=70)
curve((exp(x)/(1+exp(x)))+0.5,from=0,to=6,xlim=c(0,6),ylim=c(0,3),xlab="|h|",ylab="V(|h|)")
y4 = 2^exp(x)/(1+exp(x))
lines(x,y4,lty=2)
```

A.4 The Drawing of the Nugget Anisotropic Directional Variogram

```
x = seq(0,6,lenght=70)
curve(exp(x)/(1+exp(x))+1,from=0,to=6,xlim=c(0,6),ylim=c(0,5),xlab="|h|",ylab="V(|h|)",lty=1)
x = seq(0,2.8,length=500)
y5 = sqrt(x+1)
lines(x,y5,lty=2)
```

A.5 Calculation of the Theoretical Variance

```
### For calculating  $\inf_{x \in D} \int_D \exp \{(\|x - u\|)/0.25\} du$ 
```

```
z = seq(0,1,0.0001)
```

```
output = c()
```

```
for(i in 1:length(z))
```

```
{
```

```
f = function(x)
```

```
{
```

```
exp(-sqrt((z[i])^2+x^2)/0.25)
```

```
}
```

```
inf = integrate(f,0,1)
```

```
output[i] = inf$value
```

```
list("output"=output)
```

```
}
```

```
output
```

```
### For calculating  $\int_D \int_D \exp \{(\|u - v\|)/0.25\} dudv$ 
```

```
library(adapt)
```

```
f = function(x)
```

```
{
```

```
x1 = x[1]
```

```
y1 = x[2]
```

```
exp(-sqrt(x1^2+y1^2)/0.25)
```

```
}
```

```
adapt(2,c(0,0),c(1,1),functn=f)
```

A.6 Dataset Generation

```
library(geoR)

set.seed(571)

one.sample = function(n,mu,sigma,phi,kappa,nugget)

{

x1 = runif(n,0,1)

x2 = runif(n,0,1)

loc = cbind(x1,x2)

cov = varcov.spatial(coords=loc,cov.model="powered.exponential",kappa=kappa,nugget=nugget,cov.pars=c(sigma,phi))

decom = eigen(cov$varcov)

D = sqrt(diag(decom$values))

A = decom$vectors%*%D

z = rnorm(n,0,1)

S = A%*%z

Y = mu+S

dat = cbind(loc,Y)

dat = as.geodata(dat)

list("geodata"=dat)

}
```

A.7 Kriging of the Mean and its Empirical Variance

```
### To generate a dataset for kriging the mean

dat = one.sample(100,10,1,0.25,1,0)

### Kriging of the mean

invcov = varcov.spatial(coords=dat$geodata$coords,cov.model="powered.exponential",kappa=1,nugget=0,cov.pars=c(1,0.25),inv=T)

invcov = invcov$inverse

rowsum = apply(invcov,1,sum)

invcovsum = sum(rowsum)

krige.var = 1/invcovsum

weights = krige.var*rowsum

mean = t(weights)%*%dat$geodata$data

### To print the kriged mean

mean

### To print the kriging variance

krige.var

### To print the kriging weights

weights

sum(weights)
```

```

### Function to generate a random sample and calculate the kriged mean,kriging variance and kriging
weights.

kriging.mean = function(n,mu,sigma,phi,kappa,nugget)

{

x1 = runif(n,0,1)

x2 = runif(n,0,1)

loc = cbind(x1,x2)

cov = varcov.spatial(coords=loc,cov.model="powered.exponential",kappa=kappa,nugget=nugget,cov.pars=c(sigma,phi))

decom = eigen(cov$varcov)

D = sqrt(diag(decom$values))

A = decom$vectors%*%D

z = rnorm(n,0,1)

S = A%*%z

Y = mu+S

dat = cbind(loc,Y)

dat = as.geodata(dat)

invcov = varcov.spatial(coords=loc,cov.model="powered.exponential",kappa=kappa,nugget=nugget,cov.pars=c(sigma,phi),inv=T)

invcov = invcov$inverse

rowsum = apply(invcov,1,sum)

invcovsum = sum(rowsum)

krige.var = 1/invcovsum

weights = krige.var*rowsum

mu = t(weights)%*%Y

list("kriged.mean"=mu,"kriging.variance"=krige.var,"kriging.weights"=weights)

}

### Function for doing the Monte Carlo experiment

```

```

simu1 = function(N,n,mu,sigma,phi,kappa,nugget)

{

krige.var = c()

for(i in 1:N)

{

output = kriging.mean(n,mu,sigma,phi,kappa,nugget)

krige.var[i] = output$kriged.mean

}

avg.krige.var = sum((krige.var-mean(krige.var))^2)/N

avg.krige.var

}

### Calculating the empirical variance of the kriged mean for sample size=10, with number of simulations=1000.

simu1(1000,10,10,1,0.25,1,0)

```


A.8 Validating the function for Simple Kriging in R

```
### Function to generate a random sample with seed 571.

simple.krigen=function(n,mu,sigma,phi,kappa,nugget)

{

  set.seed(571)

  x1=runif(n,0,1)

  x2=runif(n,0,1)

  loc=cbind(x1,x2)

  cov=varcov.spatial(coords=loc,cov.model="powered.exponential",kappa=kappa,nugget=nugget,cov.pars=c(sigma,phi))

  decom=svd(cov$varcov)

  D=diag(decom$d)

  A=decom$v%*%D

  z=rnorm(n,0,1)

  S=A%*%z

  Y=mu+S

  dat=cbind(loc,Y)

  data=as.geodata(dat)

  list("data"=data)

}

### Creating the random sample with sample size=10.

example2=simple.krigen(10,10,1,0.25,1,0)

dat=example2$data

### Using section 2.3 method.

ml=likfit(dat,ini=c(1,0.01),cov.model="powered.exponential")

phi=ml$phi
```

```

sigmasq=m1$sigmasq

mu=9

x1=dat$coords[,1]

x2=dat$coords[,2]

rho=matrix(numeric(length(x1)*length(x2)),ncol=length(x2))

for(i in 1:(length(x1)))

{

for(j in 1:(length(x1)))

{

if(i>j) rho[i,j]=0 else if(j>i) rho[i,j]=exp((-sqrt((x1[j]-x1[i])^2 + (x2[j]-x2[i])^2))/phi) else rho[i,j]=1

}

rho

}

cov.matrix=sigmasq*(rho+t(rho)-diag(1,length(x1),length(x2)))

inv.cov.matrix=solve(cov.matrix)

t1=rep(0.5,10)

t2=rep(0.5,10)

h=sqrt((t1-x1)^2 + (t2-x2)^2)^.5

tar.cov=sigmasq*(exp(-h/phi))

weights=inv.cov.matrix%*%tar.cov

kriging.variance=sigmasq-(t(weights)%*%tar.cov)

kriging.variance

pred=mu+t(weights)%*%(dat$data-mu)

pred

### Using the R built-in function.

locs=c(0.5,0.5)

```

```
kc=krige.control(type="sk",beta=mu,cov.pars=c(sigmasq,phi))
```

```
sk=krige.conv(dat,krige=kc,loc=locs)
```

```
w = krweights(cbind(x1,x2),loc=locs,krige=kc)
```

```
sk$predict
```

```
sk$krige.var
```

```
w
```

A.9 Validating the function for Ordinary Kriging in R

```
### Function to generate a random sample with seed 571.

ordinary.kriging = function(n,mu,sigma,phi,kappa,nugget)

{

  set.seed(571)

  x1 = runif(n,0,1)

  x2 = runif(n,0,1)

  loc = cbind(x1,x2)

  cov = varcov.spatial(coords=loc,cov.model="powered.exponential",kappa=kappa,nugget=nugget,cov.pars=c(sigma,phi))

  decom = eigen(cov$varcov)

  D = sqrt(diag(decom$values))

  A = decom$vectors%*%D

  z = rnorm(n,0,1)

  S = A%*%z

  Y = mu+S

  dat = cbind(loc,Y)

  data = as.geodata(dat)

  list("data"=data)

}

### Creating the random sample with sample size=10.

example = ordinary.kriging(10,10,1,0.25,1,0)

### Using Section 2.4 method.

dat = example$data

ml = likfit(dat,ini=c(1,0.01),cov.model="powered.exponential")

phi = ml$phi
```

```

sigmasq = m1$sigmasq

x1 = dat$coords[,1]

x2 = dat$coords[,2]

rho=matrix(numeric(length(x1)*length(x2)),ncol=length(x2))

for(i in 1:(length(x1)))

{

for(j in 1:(length(x1)))

{

if(i>j) rho[i,j]=0 else if(j>i) rho[i,j]=exp((-sqrt((x1[j]-x1[i])^2 + (x2[j]-x2[i])^2))/phi) else rho[i,j]=1

}

rho

}

cov.matrix = sigmasq*(rho+t(rho)-diag(1,length(x1),length(x2)))

T1 = c(rep(1,length(x1)))

T2 = c(rep(1,length(x2)),0)

cov.matrix = cbind(cov.matrix,T1)

cov.matrix = rbind(cov.matrix,T2)

inv.cov.matrix = solve(cov.matrix)

t1 = rep(0.5,10)

t2 = rep(0.5,10)

h = sqrt((t1-x1)^2+(t2-x2)^2)

tar.cov = sigmasq*(exp(-h/phi))

tar.cov = c(tar.cov,1)

weights = inv.cov.matrix%%tar.cov

lambda = weights[length(x1)+1]

weights = weights[-(length(x1)+1)]

```

```

weights

kriging.variance = sigmasq - t(weights)%*%tar.cov[-(length(x1)+1)] - lambda

kriging.variance

pred = t(weights)%*%dat$data

pred

### Using the R built-in function.

locs = matrix(c(0.5,0.5),1,2)

kc = krige.control(type="ok",cov.pars=c(sigmasq,phi))

ok = krige.conv(dat,krige=kc,loc=locs)

w = krweights(cbind(x1,x2),loc=locs,krige=kc)

ok$predict

ok$krige.var

w

```

A.10 Empirical Variances of Simple Kriging Estimator

```
### Function to generate random sample for simple kriging.

simple.kriging = function(n,mu,sigma,phi,kappa,nugget)

{

x1 = runif(n,0,1)

x2 = runif(n,0,1)

loc = cbind(x1,x2)

cov = varcov.spatial(coords=loc,cov.model="powered.exponential",kappa=kappa,nugget=nugget,cov.pars=c(sigma,phi))

decom = eigen(cov$varcov)

D = sqrt(diag(decom$values))

A = decom$vectors%%D

z = rnorm(n,0,1)

S = A%%z

Y = mu+S

dat = cbind(loc,Y)

data = as.geodata(dat)

list("data"=data)

}

### Inputting the sample size.

n = 10

### Calculating the simple kriging estimate, kriging variance for each of the 1000 random samples.

kriging.mean = c()

for(i in 1:1000)

{

example = simple.kriging(n,10,1,0.25,1,0)
```

```

dat = example$data

ml = likfit(dat,ini=c(1,0.25),cov.model="powered.exponential")

locs = c(0.5,0.5)

kc = krige.control(type="sk",beta=10,obj.mod=ml)

sk = krige.conv(dat,krige=kc,loc=locs)

kriging.mean[i] = sk$predict

}

### Calculating the empirical variance of simple kriging estimator.

output = sum((kriging.mean-mean(kriging.mean))^2)/1000

output

```


A.11 Empirical Variances of Ordinary Kriging Estimator

```
### Function to generate random samples for ordinary kriging.

ordinary.kriging = function(n,mu,sigma,phi,kappa,nugget)

{

x1 = runif(n,0,1)

x2 = runif(n,0,1)

loc = cbind(x1,x2)

cov = varcov.spatial(coords=loc,cov.model="powered.exponential",kappa=kappa,nugget=nugget,cov.pars=c(sigma,phi))

decom = eigen(cov$varcov)

D = sqrt(diag(decom$values))

A = decom$vectors%*%D

z = rnorm(n,0,1) S = A%*%z

Y = mu+S

dat = cbind(loc,Y)

data = as.geodata(dat)

list("data"=data)

}

### Inputting the sample size.

n = 10

### Calculating the ordinary kriging estimate, kriging variance for each of the 1000 random samples.

kriging.mean = c()

for(i in 1:1000)

{

example = ordinary.kriging(n,10,1,0.25,1,0)
```

```

dat = example$data

ml = likfit(dat, ini=c(1,0.25), cov.model="powered.exponential")

loci = c(0.5,0.5)

kc = krige.control(type="ok", obj.mod=ml)

ok = krige.conv(dat, krige=kc, loc=loci)

kriging.mean[i] = ok$predict

}

### Calculating the empirical variance of ordinary kriging estimator.

output = sum((kriging.mean-mean(kriging.mean))^2)/1000

output

```

A.12 Comparison of the Sample Mean and the Kriged Mean Under the Random and Non-Random Design.

```
### Random sampling design ### Function to create random samples, and also calculate its sample
mean, sample variance, kriging mean and kriging variance

ks.sample = function(n,mu,sigma,phi,kappa,nugget)

{

x1 = runif(n,0,1)

x2 = runif(n,0,1)

loc = cbind(x1,x2)

cov = varcov.spatial(coords=loc,cov.model="powered.exponential",kappa=kappa,nugget=nugget,

cov.pars=c(sigma,phi))

decom = eigen(cov$varcov)

D = sqrt(diag(decom$values))

A = decom$vectors%*%D

z = rnorm(n,0,1)

S = A%*%z

Y = mu+S

dat = cbind(loc,Y) dat = as.geodata(dat)

invcov = varcov.spatial(coords=loc,cov.model="powered.exponential",kappa=kappa,nugget=nugget,

cov.pars=c(sigma,phi),inv=T)

invcov = invcov$inverse

rowsum = apply(invcov,1,sum)

invcovsum = sum(rowsum)

krige.var = 1/invcovsum
```

```

weights = krige.var*rowsum

krige.mu = t(weights)%*%Y

mu = mean(Y)

var = sum((Y-mu)^2)/n

list("kriged.mean"=krige.mu,"kriging.variance"=krige.var,"kriging.weights"=weights,"mean"=mu,"variance"=var)

}

### Function for carrying out the Monte Carlo experiment.

simu2 = function(N,n,mu,sigma,phi,kappa,nugget)

{

var.mean = c()

var.krige.mean = c()

for(i in 1:N)

{

output = nr.sample(n,mu,sigma,phi,kappa,nugget)

var.mean[i] = output$mean

var.krige.mean[i] = output$kriged.mean

}

avg.var = sum((var.mean-mean(var.mean))^2)/N

avg.krige.var = sum((var.krige.mean-mean(var.krige.mean))^2)/N

list("avg.var"=avg.var,"avg.krige.var"=avg.krige.var,"vari"=var.mean,"vari2"=var.krige.mean)

}

### Calculating the empirical variance of the sample mean, kriged mean under the random sampling

design for sample size=10 and number of simulations=1000.

simu2(10000,10,1,0.25,1,0)

```

```

### Non-random sampling design ### Function to generate a non-random grid of measurements

nr.sample = function(n,mu,sigma,phi,kappa,nugget)

{

x1 = seq(0,1,1/n)

x2 = seq(0,1,1/n)

loc = matrix(ncol=2)

for(i in 1:(n+1))

{

T1 = cbind(rep(x1[i],n+1),x2)

loc = rbind(loc,T1)

}

loc = loc[-1,]

cov = varcov.spatial(coords=loc,cov.model="powered.exponential",kappa=kappa,nugget=nugget,cov.pars=c(sigma,phi))

decom = eigen(cov$varcov)

D = sqrt(diag(decom$values))

A = decom$vectors%*%D

z = rnorm((n+1)^2,0,1)

S = A%*%z

Y = mu+S

dat = cbind(loc,Y)

dat = as.geodata(dat)

invcov = varcov.spatial(coords=loc,cov.model="powered.exponential",kappa=kappa,nugget=nugget,cov.pars=c(sigma,phi),inv=T)

invcov = invcov$inverse

rowsum = apply(invcov,1,sum)

invcovsum = sum(rowsum)

krige.var = 1/invcovsum

```

```

weights = krige.var*rowsum

krige.mu = t(weights)%*%Y

mu = mean(Y)

var = sum((Y-mu)^2)/n

list("kriged.mean"=krige.mu,"kriging.variance"=krige.var,"kriging.weights"=weights,"mean"=mu,"variance"=var)

}

### Function for doing the simulation experiment

simu3 = function(N,n,mu,sigma,phi,kappa,nugget

{

var.mean = c()

var.krige.mean = c()

for(i in 1:N)

{

output = one.sample(n,mu,sigma,phi,kappa,nugget)

var.mean[i] = output$mean

var.krige.mean[i] = output$kriged.mean

}

avg.var = sum((var.mean-mean(var.mean))^2)/N

avg.krige.var = sum((var.krige.mean-mean(var.krige.mean))^2)/N

list("avg.var"=avg.var,"avg.krige.var"=avg.krige.var,"vari"=var.mean,"vari2"=var.krige.mean)

}

```

A.13 Dioxin Data Analysis

```
library(sp)

library(gstat)

library(geoR)

dat = read.table("dioxin.txt",sep=" ",header=T)

dat2 = read.geodata("dioxin.txt",sep=" ",header=T)

### Exploratory Data Analysis (EDA)

loca = cbind(dat[,1],dat[,2])

resp = dat[,3]

dat3 = SpatialPointsDataFrame(loca,dat)

### Investigate sample correlations

postscript(file="Correlation1.ps",height=10,width=8,horizontal=F)

hscat(resp ~ 1,dat3,c(seq(0,1200,by=100)))

dev.off()

newresp = log(resp)

dat4 = as.data.frame(cbind(loca,newresp))

dat5 = as.geodata(dat4)

dat4 = SpatialPointsDataFrame(loca,dat4)

postscript(file="Correlation2.ps",height=10,width=8,horizontal=F)

hscat(newresp ~ 1,dat4,c(seq(0,1200,by=100)))

dev.off()

postscript(file="Correlation3.ps",height=10,width=8,horizontal=F)

hscat(newresp ~ 1,dat4,c(seq(0,1200,by=50)))

dev.off()

hscat(newresp ~ 1,dat4,c(seq(0,1200,by=10)))
```

```

#### Variogram modelling

postscript(file="Cloud2.ps",height=10,width=8,horizontal=F)

plot(variogram(newresp 1,dat4,cloud=TRUE))

dev.off()

par(mfrow=c(1,2))

v1 = variogram(newresp 1,dat4,cutoff=3000,width=100)

plot(v1)

fit = fit.variogram(v1,vgm(3.5,"Exp",300,1))

fit

x = seq(0,3000,by=1)

y = c()

for(i in 1:length(x))

{

y[i] = fit$psill[1]+(fit$psill[2])*(1-exp(x[i]/-fit$range[2]))

}

#### Theoretical Variogram with Sample Variogram

drawx = v1[,2]

drawy = v1[,3]

plot(drawx,drawy,xlim=c(0,3000),ylim=c(0,4),xlab="distance",

ylab="semivariance")

lines(x,y)

#### Ordinary Kriging

predx = rep(c(1:3500),7)

predy = c(rep(5,3500),rep(15,3500),rep(25,3500),rep(35,3500),rep(45,3500), rep(55,3500),rep(65,3500))

pred.xy = as.data.frame(cbind(predx,predy))

#### Create the Spatial Pixels Object for Ordinary Kriging

```



```

pred = SpatialPoints(cbind(pred.xy$predx,pred.xy$predy))

pred = SpatialPixels(pred)

di.ok = krige(newresp 1,dat4,pred,fit)

postscript(file="Prediction1.ps",height=10,width=8,horizontal=F)

spplot(di.ok["var1.pred"], main = "Ordinary Kriging Predictions")

dev.off()

postscript(file="Variance1.ps",height=10,width=8,horizontal=F)

spplot(di.ok["var1.var"], main = "Ordinary Kriging Variance")

dev.off()

}

### Directional Variogram

v.dir = variogram(newresp 1,dat4,alpha=(0:3)*45)

plot(v.dir)

### Cross-validation

x2 = dat[,1]

y2 = dat[,2]

hscat(newresp x2+y2,dat4,c(seq(0,1200,by=100)))

v2 = variogram(newresp x2+y2,dat4,cutoff=3000,width=100)

plot(v2)

fit2 = fit.variogram(v2,vgm(1.5,"Exp",300,2))

fit2

di.ok2 = krige(newresp 1,dat4,pred,fit2)

spplot(di.ok2["var1.pred"], main = "Ordinary Kriging Predictions")

spplot(di.ok2["var1.var"], main = "Ordinary Kriging Variance")

cv2 = krige.cv(newresp 1,dat4,fit2,nfold=127)

```

A.14 Housing Prices Data Analysis

```
library(geoR)

dat = read.geodata("Housing Prices Data.txt",header=FALSE,

coords.col=2:3,data.col=4)

y = dat$data

x = dat$coords

### Preliminary Analysis

postscript(file="Prelim.ps",height=10,width=10,horizontal=F)

plot(dat,lowess=T,density=FALSE)

dev.off()

### Variogram Cloud

postscript(file="Cloud.ps",height=8,width=8,horizontal=F)

cloud = variog(dat,option="cloud")

plot(cloud,xlab="--h--",ylab=expression(paste(gamma, "(h)")))

dev.off()

trend = trend.spatial("1st",dat)

d.variogram = variog4(dat,option="bin",uvec=seq(0,70,by=10),trend=trend)

postscript(file="Directional2.ps",height=8,width=8,horizontal=F)

plot(d.variogram)

dev.off()

### Transform into isotropic space

a.angle = 3*pi/2

angle = c(cos(a.angle),-sin(a.angle),+sin(a.angle),cos(a.angle))

angle = matrix(angle,2,2)

a.ratio = 1/1.5
```

```

ratio = c(1,0,0,a.ratio)

ratio = matrix(ratio,2,2)

x1 = x[,1]

x2 = x[,2]

x3 = cbind(x1,x2)

new.coords = x3*%%angle%*%%ratio

dat2 = as.geodata(cbind(new.coords,y),header=FALSE,

coords.col=1:2,data.col=3)

d2.variogram = variog4(dat2,option="bin",uvec=seq(0,100,by=10),trend=trend)

postscript(file="Directional3.ps",height=8,width=8,horizontal=F)

plot(d2.variogram)

dev.off()

### Experimental Variogram after transformation of coordinates

experimental3 = variog(dat,option="bin",uvec=seq(0,100,by=10),trend=trend)

postscript(file="Experimental2.ps",height=8,width=8,horizontal=F)

plot(experimental3)

dev.off()

### To estimate the covariance parameters

ml = likfit(dat2,cov.model="exponential",ini.cov.pars=c(500,50),

trend=trend,nugget=200)

kc1 = krige.control(type.krige="ok",obj.model=ml)

loci = pred_grid(c(500,590),c(-660,-570),by=10)

pred1 = krige.conv(dat2,locations=loci,krige=kc1)

### To visualize the predictions

postscript(file="Contour.ps",height=8,width=8,horizontal=F)

contour(pred1)

```

```
dev.off()

postscript(file="Prediction2.ps",height=8,width=8,horizontal=F)

par(mfrow=c(1,2))

contour(pred1)

points(dat2,add=T,cex.max=2)

persp(pred1)

dev.off()
```

B References

1. ASCE, M., Harris, D.J. and Zirschky J.H., 1983, Geostatistical Analysis of Hazardous Waste Site Data: *Journal of Environmental Engineering*, v.38, p. 770-784.
2. Bivand, R.S., Pebesma, E.J. and Rubio V.G., 2008, *Applied Spatial Data Analysis with R*, Springer, New York.
3. Cox, D.R. and Hinkley, D.V., 1974, *Theoretical Statistics*: Chapman and Hall, London.
4. Cressie, N., 1993, *Statistics for Spatial Data*, Revised Edition: Wiley Series in Probability and Mathematical Statistics.
5. Diggle, P.J. and Ribeiro Jr., P.J., 2007, *Model-Based Geostatistics*: Springer, New York.
6. Gombay, E., 2010, *When the Law of Large Numbers Do Not Hold*: manuscript, University of Alberta.
7. Johnson, R.A. and Wichern D.W., 1998, *Applied Multivariate Statistical Analysis*: Prentice-Hall, Canada.
8. Journel, A.G. and Huijbregts, C.J., 1978, *Mining Geostatistics*: Academic Press, London.
9. Wackernagel, H., 1998, *Multivariate Geostatistics*: Springer-Verlag, Germany.
10. Zimmerman, D.L., 1993, Another Look at Anisotropy in Geostatistics: *Mathematical Geology*, v.25, p. 453-470.