Fall 2009

Spatial multivariate design in the plane and on stream networks

Jie Li
University of Iowa

SPATIAL MULTIVARIATE DESIGN IN THE PLANE AND ON STREAM NETWORKS

by

Jie Li

An Abstract

Of a thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Statistics in the
Graduate College of The
University of Iowa

December 2009

Thesis Supervisor: Professor Dale Zimmerman

ABSTRACT

In environmental studies, measurements of interest are often taken on multiple variables. The results of spatial data analyses can be substantially affected by the spatial configuration of the sites where measurements are taken. Hence, optimal designs which result in data guaranteeing efficient statistical inferences need to be studied.

We study optimal designs on two large classes of spatial regions with respect to three design criteria, which were prediction, covariance parameter estimation, and empirical prediction. The first class of regions includes those in the plane, where Euclidean distance is used. The performance of the optimal designs is compared to that of randomly chosen designs. Optimal designs for a small example and a relatively large example are obtained. For the small example, complete enumeration of all possible designs is computationally feasible. For the large example, the computational difficulty in searching for the optimal spatial sampling design is overcome by a simulated annealing algorithm.

The second class of spatial regions includes streams and rivers, where the distance is defined as distance along the stream network. A moving average construction is used to establish valid covariance and cross-covariance models using stream distance. Optimal designs for small and large examples are obtained. An application of our methodology to a real stream network is included.

We discuss the impact of asymmetry in the cross covariance function on the spatial multivariate design. The relationship between multivariate optimal design and univariate optimal design if the multivariate design is restricted to be completely collocated is studied. The efficiency lost if we consider the design that is optimal within the class of collocated designs is discussed.

Abstract Approved: _____
Thesis Supervisor

Title and Department

Date

SPATIAL MULTIVARIATE DESIGN IN THE PLANE AND ON STREAM NETWORKS

by

Jie Li

A thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Statistics in the
Graduate College of The
University of Iowa

December 2009

Thesis Supervisor: Professor Dale Zimmerman

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Jie Li

has been approved by the Examining Committee
for the thesis requirement for the Doctor of
Philosophy degree in Statistics at the December 2009
graduation.

Thesis Committee: _____
Dale Zimmerman, Thesis Supervisor

_____
Kung-Sik Chan

_____
Kate Cowles

_____
Richard Dykstra

_____
Witold Krajewski

To my family

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my major professor Dr. Dale Zimmerman for his inspiring guidance, constructive suggestions and enthusiastic encouragement during my graduate study. I am also very grateful to my committee members, Drs. Kung-Sik Chan, Kate Cowles, Richard Dykstra, and Witold Krajewski for their precious help. I am deeply appreciative of the professors in the department for their excellent teaching and the staff for their kind assistance in these years.

ABSTRACT

In environmental studies, measurements of interest are often taken on multiple variables. The results of spatial data analyses can be substantially affected by the spatial configuration of the sites where measurements are taken. Hence, optimal designs which result in data guaranteeing efficient statistical inferences need to be studied.

We study optimal designs on two large classes of spatial regions with respect to three design criteria, which were prediction, covariance parameter estimation, and empirical prediction. The first class of regions includes those in the plane, where Euclidean distance is used. The performance of the optimal designs is compared to that of randomly chosen designs. Optimal designs for a small example and a relatively large example are obtained. For the small example, complete enumeration of all possible designs is computationally feasible. For the large example, the computational difficulty in searching for the optimal spatial sampling design is overcome by a simulated annealing algorithm.

The second class of spatial regions includes streams and rivers, where the distance is defined as distance along the stream network. A moving average construction is used to establish valid covariance and cross-covariance models using stream distance. Optimal designs for small and large examples are obtained. An application of our methodology to a real stream network is included.

We discuss the impact of asymmetry in the cross covariance function on the spatial multivariate design. The relationship between multivariate optimal design and univariate optimal design if the multivariate design is restricted to be completely collocated is studied. The efficiency lost if we consider the design that is optimal within the class of collocated designs is discussed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure

CHAPTER 1
INTRODUCTION

1.1 Multivariate spatial sampling design

Data collected in conjunction with studies in mining engineering, soil and crop science, hydrology, and ecology are usually spatially correlated. The geostatistical method has proven to be a useful approach to analyze such data. However, in spatial data analysis, inferences are substantially affected by the spatial configuration of the sites where measurements are taken. Hence optimal designs which result in data guaranteeing efficient statistical inferences need to be studied.

In geostatistical analysis, spatial prediction is usually the ultimate objective. To achieve this objective, one variable may be predicted at a time, based on data from the same type of variable (kriging) or using additional data which is from auxiliary variables (cokriging). However, simultaneously predicting two or more variables (multivariate prediction) may also be of interest. For example, in mining, geologists may measure concentrations of both lead and zinc from the same ore body and use the data to predict both lead and zinc at unsampled locations simultaneously. In ecology, ecologists are concerned with the joint spatial pattern of several species in the same community, and it is desirable to predict the joint abundance of species at unsampled spatial locations. The variables are usually correlated with each other, both spatially (at nearby sites) and non-spatially (at the same site). Hence, multivariate spatial sampling design should consider both the correlation among values of the same variable at different spatial locations and the cross correlation among values of different variables.

1.2 Spatial autocovariance function for Euclidean distance and stream distance

Spatial optimal design depends on the nature of the study region. One important issue related to this is how to define the distance between two spatial locations. For studies

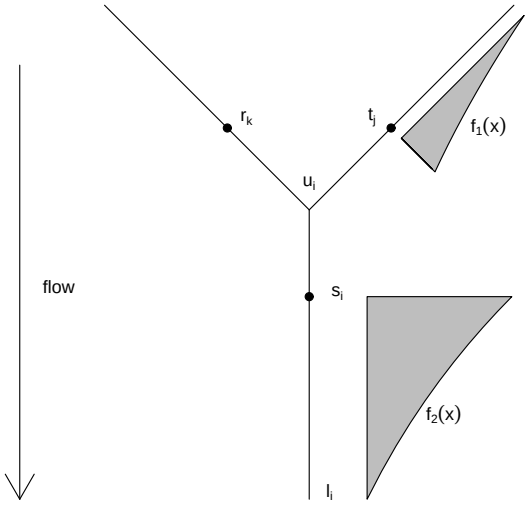

Figure 1.1: A very simple stream network

in mining engineering, soil and crop science, and ecology, distance is usually defined as the Euclidean distance, which can be computed by utilizing each location's latitude and longitude (provided that the extent of the study region is small enough that the curvature of the earth's surface may be ignored). For this case, the covariance model can be chosen from among those autocovariance functions already developed for Euclidean distance.

However, for studies related to hydrology, using Euclidean distance usually is not appropriate. Suppose there is a stream network as shown in Figure 1.1. Since the movement of materials in the water only occurs within the stream network, it is more appropriate to use distance defined only along the stream, or what we call stream distance. Stream distance between any two points is obtained by summing up the length of segments which connect them. For example, $dist(r_k, t_j) = dist(r_k, u_i) + dist(u_i, t_j)$ is the stream distance between r_k and t_j in Figure 1.1. One problem that arises with using the stream distance metric is that those autocovariance functions developed for use with a Euclidean distance metric are not necessarily valid when using stream distance, as pointed out by Ver Hoef, Peterson, and Theobald (2005). In their paper, they considered an idealized stream network starting with a single node splitting into two, and then each node splitting into two

more, etc., for $2^6-1 = 63$ nodes. They further supposed that the stream segments between nodes are all one unit long. They then considered the use of three standard autocovariance models including spherical, exponential, and linear-with-sill models over this region when using stream distance, rather than Euclidean distance. They showed that in general, using stream distance for an autocovariance function developed for Euclidean distance could result in a covariance matrix that is not positive definite.

To obtain valid (positive definite) autocovariance functions on stream networks, Ver Hoef and Peterson (2008) proposed a moving average construction method which incorporates both stream distance and flow direction. Incorporating flow direction can be achieved by choosing moving average functions whose tails go along with or against the flow direction, named as tail-down and tail-up moving average constructions, respectively. Incorporating flow direction allows for making the moving average construction method suitable for different types of variables. Some types of variables, such as water pollutant or other stream chemistry variables, move along with the water flow, so it is sensible to consider a covariance model that does not allow autocorrelation between locations if the water at one location does not flow into the other (flow-unconnected). For example, in Figure 1.1, location r_k and t_j do not share flow, so it makes sense that a stream chemistry variable at r_k be modeled as uncorrelated with that variable at t_j. But other variables, such as the occurrence of fish species, may move upstream. For these we need to consider a covariance model that allows autocorrelation even for two locations that are not flow connected. For example, in Figure 1.1, fish may swim from location r_k downstream to node u_i, and then upstream to t_j, so it is easy to understand that the occurrence of fish at r_k may be correlated with that at t_j.

A moving average function is a continuous function defined on $\mathcal{R}^1$ where $\int_{-\infty}^{\infty} |f(x)|dx$ is finite in order to create a stationary process. There is much freedom in choosing a moving average function, but typically we choose a function where most of its mass occurs at 0. To understand how the moving average construction works, here are some examples. First, create random variables as the integration of a moving average function over a white noise

random process,

$$Z(s) = \int_{-\infty}^{\infty} f(x - s|\boldsymbol{\theta})W(x)dx, \tag{1.1}$$

where $W(x)$ is a white noise random process, with $\mathrm{E}(W(x)) = 0$ and $\mathrm{var}(W(x)) = 1$, and $f(x|\boldsymbol{\theta})$ is the moving average function. The integral in Equation 1.1 is not the ordinary Riemann integral. In fact, it should be really written as $Z(s,\omega) = \int_{-\infty}^{\infty} f(x-s|\boldsymbol{\theta})W(x,\omega)dx = \int_{-\infty}^{\infty} f(x-s|\boldsymbol{\theta})dW(x,\omega)$, which is a Lebesgue integral. Then based on the moving average construction, the autocovariance function for $Z(s)$ can be expressed as

$$\begin{aligned}
C(h|\boldsymbol{\theta}) &= \mathrm{cov}(Z(s), Z(s+h)) \\
&= \mathrm{E}\left(\int_{-\infty}^{\infty} f(x-s|\boldsymbol{\theta})W(x)dx \int_{-\infty}^{\infty} f(y-s-h|\boldsymbol{\theta})W(y)dy\right) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{E}[f(x-s|\boldsymbol{\theta})f(y-s-h|\boldsymbol{\theta})W(x)W(y)]dxdy \\
&\quad \text{(Fubini Theorem)} \\
&= \int_{-\infty}^{\infty} f(x-s|\boldsymbol{\theta})f(x-s-h|\boldsymbol{\theta})dx \\
&= \int_{-\infty}^{\infty} f(x|\boldsymbol{\theta})f(x-h|\boldsymbol{\theta})dx, \qquad (1.2)
\end{aligned}$$

where in (1.2), Fubini theorem can be used because

$$\begin{aligned}
&\mathrm{E}\left(\int_{-\infty}^{\infty} |f(x-s|\boldsymbol{\theta})W(x)|dx \int_{-\infty}^{\infty} |f(y-s-h|\boldsymbol{\theta})W(y)|dy\right) \\
= &\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{E}|f(x-s|\boldsymbol{\theta})f(y-s-h|\boldsymbol{\theta})W(x)W(y)|dxdy \quad (\text{Tonelli}'\text{s Theorem}) \\
= &\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |f(x-s|\boldsymbol{\theta})f(y-s-h|\boldsymbol{\theta})| \cdot \mathrm{E}|W(x)W(y)|dxdy \\
\leq &\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |f(x-s|\boldsymbol{\theta})f(y-s-h|\boldsymbol{\theta})| \sqrt{\mathrm{E}W^2(x)\mathrm{E}W^2(y)}dxdy \\
= &\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |f(x-s|\boldsymbol{\theta})f(y-s-h|\boldsymbol{\theta})|dxdy \\
< &\ \infty
\end{aligned}$$

For the tail-up construction, we choose a moving average function $f_1(x)$ as shown in Figure 1.1, where the tail of $f_1(x)$ goes upstream. For example, let $f_1(x) = \theta_1 \exp(-x/\theta_2)I(x \geq 0)$, where $\theta_1 > 0$ and $\theta_2 > 0$. Then using (1.1) and (1.2), the autocovariance function is given by

$$C(h|\boldsymbol{\theta}) = \frac{\theta_2}{2}\theta_1^2 \exp(-h/\theta_2) \tag{1.3}$$

for any two flow-connected points, where h is the stream distance between those two points. For the tail-down construction, choose a moving average function $f_2(x)$ as shown in Figure 1.1, where the tail of $f_2(x)$ goes downstream. For example, let $f_2(x) = \theta_1 \exp(x/\theta_2) I(x \leq 0)$, where $\theta_1 > 0$ and $\theta_2 > 0$. Then the autocovariance function is given by

$$C(h|\boldsymbol{\theta}) = \frac{\theta_2}{2}\theta_1^2 \exp(-h/\theta_2) \tag{1.4}$$

for any two points in the stream network, regardless of whether they are flow-connected or flow-unconnected. Positive definite covariance matrices result from the use of these autocovariance functions or from any autocovariance function obtained via the moving average construction.

Ver Hoef and Peterson (2008) introduced the moving average construction for the univariate case only. In this thesis, because we will be considering multivariate design on stream networks, we will extend it to the multivariate case.

1.3 Optimal design for prediction, covariance estimation, and empirical prediction in the univariate setting

Spatial optimal design also depends on study objectives. For example, in the univariate setting, kriging is often the ultimate objective of a study, so most authors have considered sampling design that is optimal for predicting a spatial random process Z at some set of unobserved locations. Invariably, these design considerations are made under an assumption that the spatial covariance structure is completely known. However, in practice, the covariance structure usually has to be estimated, and in some studies, understanding it is the main objective. Hence, sampling designs that are optimal for covariance parameter estimation have also been considered for the univariate case. To still address the issue of prediction, several authors have considered optimal sampling design for prediction at unobserved locations when the covariance parameters are estimated — so-called empirical prediction. So the three most commonly considered objectives in spatial design problems are prediction, covariance parameter estimation, and empirical prediction, and lots of research has been done with respect to these objectives in the univariate planar

setting.

In the univariate planar setting, Bras and Rodriguez-Iturbe (1976), McBratney, Webster, and Burgess (1981), Yfantis, Flatman, and Behar (1987), Barnes (1989), Cressie, Gotway, and Grondona (1990), and Zimmerman (2006) considered spatial optimal design with respect to prediction. Suppose that a continuous, spatially-varying random process $Z = \{Z(\mathbf{s}) : \mathbf{s} \in \mathcal{D}\}$ is observed at locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$ in a study region $\mathcal{D}$. Let $\mathbf{z} = (Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n))$ represent these observations, which can be considered as a spatially incomplete sample of one realization of the random process Z. Furthermore, assume that the mean structure of this random process is a linear function of unknown parameters; that is, $E[Z(\mathbf{s})] = \sum_{i=0}^{p} \beta_i \mu_i(\mathbf{s})$ where $\mu_i' s$ are known functions of the observed spatial coordinates of $\mathbf{s}$. Also define the covariance function by $C(\mathbf{s}, \mathbf{u}) = \text{cov}[Z(\mathbf{s}), Z(\mathbf{u})]$, which is assumed, for now, to be completely known. Then for an arbitrary point $\mathbf{s}_0 \in \mathcal{D}$, the best linear unbiased predictor (BLUP) of $Z(\mathbf{s}_0)$, or the so-called universal kriging predictor, is given by

$$\hat{Z}(\mathbf{s}_0, \boldsymbol{\theta}) = [\mathbf{c} + \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}(\mathbf{X}(\mathbf{s}_0) - \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{c})]'\boldsymbol{\Sigma}^{-1}\mathbf{z}. \tag{1.5}$$

In this equation, $\mathbf{c}$ is a vector whose ith element is $C(\mathbf{s}_i, \mathbf{s}_0)$; $\mathbf{X}$ is the model matrix whose ijth element is $\mu_{j-1}(\mathbf{s}_i)$; $\boldsymbol{\Sigma}$ is the covariance matrix whose ijth element is $C(\mathbf{s}_i, \mathbf{s}_j)$; and $\mathbf{X}(\mathbf{s}_0)$ is a vector whose jth element is $\mu_{j-1}(\mathbf{s}_0)$. Then the prediction error variance, or the so-called universal kriging variance, associated with $\hat{Z}(\mathbf{s}_0)$ is given by

$$\sigma_K^2(\mathbf{s}_0, \boldsymbol{\theta}) = C(\mathbf{s}_0, \mathbf{s}_0) - \mathbf{c}'\boldsymbol{\Sigma}^{-1}\mathbf{c} + (\mathbf{X}(\mathbf{s}_0) - \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{c})'(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}(\mathbf{X}(\mathbf{s}_0) - \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{c}). \tag{1.6}$$

The two most common design criteria considered by the aforementioned authors are either the average kriging variance or the maximum kriging variance, i.e.,

$$\frac{1}{|\mathcal{S}|}\sum_{\mathbf{s}_i \in \mathcal{S}} \sigma_K^2(\mathbf{s}_i) \quad \text{or} \quad \max_{\mathbf{s}_i \in \mathcal{S}} \sigma_K^2(\mathbf{s_i}), \tag{1.7}$$

where $\mathcal{S} \subset \mathcal{D}$ is a set (usually finite in practice) of prediction locations, and $|\mathcal{S}|$ is the cardinality of $\mathcal{S}$. An n-point design is optimal with respect to one of these criteria if it minimizes the criterion over all possible n-point designs, each point taken from a set of candidate locations. Zimmerman (2006) discussed optimal design with respect to $\max_{\mathbf{s} \in \mathrm{S}} \sigma_K^2(\mathbf{s})$ on a 5×5 grid (toy example) and a 100×100 grid (large example). In both examples, he found that the points in the optimal design for this criterion tend to be relatively uniformly

dispersed over the study region.

In practice, inference is performed as though the covariance function is known only up to the value of a parameter vector $\boldsymbol{\theta}$, so it is natural to seek a design that is optimal for estimating $\boldsymbol{\theta}$ in some sense. It is known that such an optimal design may depend on the particular estimation method and the specific measure of the estimator's quality that are to be used. In the univariate setting, Müller and Zimmerman (1999) proposed to maximize the determinant of the information matrix associated with a nonlinear generalized least squares estimator of $\boldsymbol{\theta}$. Zhu and Stein (2005) proposed a related approach, which is to assume that the underlying process is Gaussian and maximize the determinant of the Fisher information matrix associated with either the maximum likelihood (ML) estimator or residual maximum likelihood (REML) estimator of $\boldsymbol{\theta}$. The rationale for using these criteria is that under appropriate regularity conditions, the inverse Fisher information matrix is the asymptotic covariance matrix of the corresponding estimator of $\boldsymbol{\theta}$. Hence if the sample size is sufficiently large, the inverse of the information matrix would provide a reasonable approximation to the mean square error (MSE) matrix

$$\mathbf{M}(\boldsymbol{\theta}) = E[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'].$$

For a Gaussian random process Z, the Fisher information matrix associated with the ML/REML estimator can be obtained by fairly simple formulas. Zimmerman (2006) discussed the optimal design with respect to covariance parameter estimation using ML and REML estimation for covariance parameters and found that in all the optimal designs, the points are clustered to some degree. It appears that the presence of some small lags in the design is very useful for precise estimation of covariance parameters but not for prediction with known covariance parameters. Actually, Zimmerman (2006) pointed out that these two design objectives are largely antithetical.

Using the Fisher information matrix as a design criterion for covariance parameter estimation is based on standard asymptotic theory. In reality, we usually suffer from sample sizes too small to guarantee the asymptotic properties. Both Zhu and Stein (2005) and Zimmerman (2006) investigated whether the inverse Fisher information matrix is a reasonable design criterion for datasets of sizes often encountered in applications. As

pointed out in these papers, the usefulness of the inverse Fisher information matrix as a design criterion relies not on its quality as an approximation to the MSE matrix but on the monotonicity of its relationship with MSE. Simulation studies done in their papers demonstrated that the inverse Fisher information matrix and MSE rank designs in very similar orders. So even for small sample sizes, using the Fisher information matrix as the design criterion appears to be a reasonable thing to do.

When considering spatial sampling design for empirical prediction, one needs to account for covariance parameter estimation uncertainty in the point prediction. Zimmerman (2006) considered designs that minimize the average approximate mean square errors of the empirical best linear unbiased predictors (E-BLUP). The E-BLUP $\hat{Z}(\mathbf{s}_0, \hat{\boldsymbol{\theta}})$ is given by the same expression as (1.5) except with $\boldsymbol{\Sigma}$ and $\mathbf{c}$ evaluated at $\hat{\boldsymbol{\theta}}$, an estimate of $\boldsymbol{\theta}$, rather than at the assumed known $\boldsymbol{\theta}$. It is known that the E-BLUP is unbiased under very weak conditions (Kackar and Harville 1984). However, there is no known explicit expression for the prediction error variance associated with E-BLUP, i.e.,

$$\sigma_K^2(s_0, \hat{\boldsymbol{\theta}}) = \text{var}(\hat{Z}(\mathbf{s}_0, \hat{\boldsymbol{\theta}}) - Z(\mathbf{s}_0)).$$

Harville and Jeske (1992) and Zimmerman and Cressie (1992) proposed an approximation to it:

$$\sigma_K^2(s_0, \hat{\boldsymbol{\theta}}) \doteq \sigma_K^2(s_0, \boldsymbol{\theta}) + \text{tr}[\mathbf{A}(\mathbf{s}_0, \boldsymbol{\theta})\mathbf{H}(\boldsymbol{\theta})] \tag{1.8}$$

where $\mathbf{A}(\mathbf{s}_0, \boldsymbol{\theta}) = \text{var}[\frac{\partial \hat{Z}(\mathbf{s}_0, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}]$ and $\mathbf{H}(\boldsymbol{\theta})$ is a matrix that either equals or approximates the MSE matrix $\mathbf{M}(\boldsymbol{\theta})$. If $\boldsymbol{\theta}$ is estimated by ML/REML, then a natural choice for $\mathbf{H}(\boldsymbol{\theta})$ is the inverse Fisher information matrix for the ML/REML estimator. Accordingly, the design criterion with respect to empirical prediction can either be the average empirical kriging variance or the maximum empirical kriging variance. Zimmerman (2006) used maximum empirical kriging variance as the criterion and the characteristics of the optimal design were found to lie somewhere between those for prediction and covariance parameter estimation. That is, the optimal design exhibits some clustering but is otherwise rather regularly spaced. Zhu and Stein (2006) derived another design criterion that accounts for estimation uncertainty in both point and interval prediction using asymptotic approximations, and

similar results were found.

1.4 Overview

In this thesis, we extend existing methods for spatial optimal design in the univariate planar setting to the multivariate planar setting. Also, we develop spatial optimal design on stream networks for multivariate settings.

Chapter 2 discusses spatial multivariate optimal design with respect to prediction, covariance parameter estimation, and empirical prediction in the plane. Designs for a small example (toy example) and a relatively large example are considered. For the small example, complete enumeration of all possible designs is computationally feasible. Therefore, the optimal designs are obtained. For the large example, the computational difficulty in searching for the optimal spatial sampling design is overcome by a simulated annealing algorithm (SAA), which was discussed by Lark (2002) and Zhu and Stein (2005). The design produced by SAA is obtained for each study objective. The performance of the designs obtained by our strategy is compared to that of randomly chosen designs.

Chapter 3 discusses spatial multivariate design on stream networks with respect to the same three objectives. A moving average construction is used to establish valid covariance and cross-covariance models using stream distance. Designs for small and large examples are considered. An application of our methodology to a real stream network is included in Section 3.5.

The examples in Chapter 2 and 3 are done under the assumption that the cross-covariance function is symmetric. Chapter 4 discusses the impact of asymmetry in the cross-covariance function on the spatial multivariate design. Chapter 5 discuss the relationship between multivariate optimal design and univariate optimal design if we restrict the design to be completely collocated. In Chapter 6, we further discuss the efficiency lost if we consider the design that is optimal within the class of collocated designs.

As a summary, the important features of the optimized criteria and optimal designs are discussed in Chapter 7.

CHAPTER 2
OPTIMAL DESIGNS IN THE PLANE

2.1 Model

Consider a continuous, spatially-varying, vector-valued random field, $\{\mathbf{Z}(\mathbf{s}) : \mathbf{s} \in D\}$, where $\mathbf{Z}(\mathbf{s}) \in \Re^m$ and $D \subset \Re^d$. Assume the following linear model:

$$\mathbf{Z}(\mathbf{s}) = \boldsymbol{\mu}(\mathbf{s}) + \boldsymbol{\delta}(\mathbf{s}) \tag{2.1}$$

where $\boldsymbol{\mu}(\mathbf{s})$ is the mean vector of $\mathbf{Z}(\mathbf{s})$, and $\boldsymbol{\delta}(\mathbf{s})$ is a random vector with a multivariate normal distribution. Let $\mathbf{Z}_i \equiv [Z_i(\mathbf{s}_1^i), \ldots, Z_i(\mathbf{s}_{n_i}^i)]'$ denote the data for the i^{th} variable observed at spatial locations $\{\mathbf{s}_1^i, \ldots, \mathbf{s}_{n_i}^i\} \in D$ and let $\boldsymbol{\delta}_i \equiv [\delta_i(\mathbf{s}_1^i), \ldots, \delta_i(\mathbf{s}_{n_i}^i)]'$ denote the random vector for the i^{th} variable at these locations, and define $\mathbf{Z} = [\mathbf{Z}_1', \ldots, \mathbf{Z}_m']'$ and $\boldsymbol{\delta} = [\boldsymbol{\delta}_1', \ldots, \boldsymbol{\delta}_m']'$. Also, assume that $\mu_i(\mathbf{s}) = [\mathbf{x}_i(\mathbf{s})]'\boldsymbol{\beta}_i$, where $\boldsymbol{\beta}_i$ is a parameter vector of dimension $\pi_i \times 1$ and $\mathbf{x}_i(\mathbf{s})$ is a vector of explanatory variables associated with the i^{th} spatial response variable at point $\mathbf{s}$. Finally, let $\mathbf{X}_i$ be an $n_i \times \pi_i$ matrix whose k^{th} row is $[\mathbf{x}_i(\mathbf{s}_k^i)]'$; $k = 1, \ldots, n_i$; $i = 1, \ldots, m$, where m is the number of response variables. Then the linear model in (2.1) can be written as,

$$\begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \\ \vdots \\ \mathbf{Z}_m \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{X}_m \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_m \end{pmatrix} + \begin{pmatrix} \boldsymbol{\delta}_1 \\ \boldsymbol{\delta}_2 \\ \vdots \\ \boldsymbol{\delta}_m \end{pmatrix} \tag{2.2}$$

or $\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\delta}$. The covariance function of the random field is defined as $C_{ii}(\mathbf{h}) = \text{cov}(Z_i(\mathbf{s}+\mathbf{h}), Z_i(\mathbf{s}))$, and the cross-covariance function of the random field is defined as $C_{ij}(\mathbf{h}) = \text{cov}(Z_i(\mathbf{s}+\mathbf{h}), Z_j(\mathbf{s}))$.

2.2 Criteria

2.2.1 Criteria for prediction

Suppose we are most interested in predicting the value of the random field $\mathbf{Z}(\mathbf{s})$ at any arbitrary point $\mathbf{s}_0 \in D$, i.e., predicting $\mathbf{Z}(\mathbf{s}_0)$. Assuming that we know the covariance

and cross-covariance functions, under the model described above, the best linear unbiased predictor (BLUP) for $\mathbf{Z}(\mathbf{s}_0)$ is

$$\hat{\mathbf{Z}}(\mathbf{s}_0) = [\mathbf{C} + \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}(\mathbf{X}(\mathbf{s}_0) - \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{C})]'\boldsymbol{\Sigma}^{-1}\mathbf{Z}.$$

Here, $\mathbf{C} = \begin{pmatrix} C_{11}(\mathbf{s}_1^1, \mathbf{s}_0) & C_{12}(\mathbf{s}_1^1, \mathbf{s}_0) & \cdots & C_{1m}(\mathbf{s}_1^1, \mathbf{s}_0) \\ \vdots & \vdots & & \vdots \\ C_{11}(\mathbf{s}_{n_1}^1, \mathbf{s}_0) & C_{12}(\mathbf{s}_{n_1}^1, \mathbf{s}_0) & \cdots & C_{1m}(\mathbf{s}_{n_1}^1, \mathbf{s}_0) \\ C_{21}(\mathbf{s}_1^2, \mathbf{s}_0) & C_{22}(\mathbf{s}_1^2, \mathbf{s}_0) & \cdots & C_{2m}(\mathbf{s}_1^2, \mathbf{s}_0) \\ \vdots & \vdots & & \vdots \\ C_{21}(\mathbf{s}_{n_2}^2, \mathbf{s}_0) & C_{22}(\mathbf{s}_{n_2}^2, \mathbf{s}_0) & \cdots & C_{2m}(\mathbf{s}_{n_2}^2, \mathbf{s}_0) \\ \vdots & \vdots & & \vdots \\ C_{m1}(\mathbf{s}_1^m, \mathbf{s}_0) & C_{m2}(\mathbf{s}_1^m, \mathbf{s}_0) & \cdots & C_{mm}(\mathbf{s}_1^m, \mathbf{s}_0) \\ \vdots & \vdots & & \vdots \\ C_{m1}(\mathbf{s}_{n_m}^m, \mathbf{s}_0) & C_{m2}(\mathbf{s}_{n_m}^m, \mathbf{s}_0) & \cdots & C_{mm}(\mathbf{s}_{n_m}^m, \mathbf{s}_0) \end{pmatrix}$; $\mathbf{X}$ is the design matrix in linear model (2.2); $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \cdots & \boldsymbol{\Sigma}_{1m} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \cdots & \boldsymbol{\Sigma}_{2m} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{\Sigma}_{m1} & \boldsymbol{\Sigma}_{m2} & \cdots & \boldsymbol{\Sigma}_{mm} \end{pmatrix}$, where $\boldsymbol{\Sigma}_{ij} = \text{cov}[\mathbf{Z}_i, \mathbf{Z}_j]$ is a $n_i \times n_j$ matrix; $\mathbf{X}(\mathbf{s}_0) = \begin{pmatrix} \mathbf{x}_1(\mathbf{s}_0) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_2(\mathbf{s}_0) & \cdots & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{x}_m(\mathbf{s}_0) \end{pmatrix}$; and $\mathbf{Z} = [\mathbf{Z}_1', \mathbf{Z}_2', \cdots, \mathbf{Z}_m']'$ is the data vector. The prediction error covariance matrix associated with $\hat{\mathbf{Z}}(\mathbf{s}_0)$, or the mean square prediction error matrix, is given by

$$\mathbf{M}(\mathbf{s}_0, \boldsymbol{\theta}) = \boldsymbol{\Sigma}_0 - \mathbf{C}'\boldsymbol{\Sigma}^{-1}\mathbf{C} + (\mathbf{X}(\mathbf{s}_0) - \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{C})'(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}(\mathbf{X}(\mathbf{s}_0) - \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{C}), \quad (2.3)$$

where $\boldsymbol{\Sigma}_0 = \begin{pmatrix} \text{cov}(Z_1(\mathbf{s}_0), Z_1(\mathbf{s}_0)) & \text{cov}(Z_1(\mathbf{s}_0), Z_2(\mathbf{s}_0)) & \cdots & \text{cov}(Z_1(\mathbf{s}_0), Z_m(\mathbf{s}_0)) \\ \text{cov}(Z_2(\mathbf{s}_0), Z_1(\mathbf{s}_0)) & \text{cov}(Z_2(\mathbf{s}_0), Z_2(\mathbf{s}_0)) & \cdots & \text{cov}(Z_2(\mathbf{s}_0), Z_m(\mathbf{s}_0)) \\ \vdots & \vdots & & \vdots \\ \text{cov}(Z_m(\mathbf{s}_0), Z_1(\mathbf{s}_0)) & \text{cov}(Z_m(\mathbf{s}_0), Z_2(\mathbf{s}_0)) & \cdots & \text{cov}(Z_m(\mathbf{s}_0), Z_m(\mathbf{s}_0)) \end{pmatrix}$.

Suppose $S \subseteq D$ is the collection of all candidate locations for the observations. In

the univariate case, if we want to choose a subset of S where the design will be optimal for prediction, then there are two design criteria, the average kriging variance and the maximum kriging variance, most commonly used. Here in the multivariate situation, as a natural generalization of them, we can consider the average generalized prediction error variance and the maximum generalized prediction error variance as the design criteria for prediction. In other words, in order to find the optimal multivariate design with respect to prediction, we can minimize one of the following two criteria:

$$\frac{1}{|S|}\sum_{\mathbf{s}_i \in S} |\mathbf{M}(\mathbf{s}_i, \boldsymbol{\theta})| \quad \text{or} \quad \max_{\mathbf{s}_i \in S} |\mathbf{M}(\mathbf{s}_i, \boldsymbol{\theta})|$$

Here $|S|$ is the cardinality of S, and $|\mathbf{M}(\mathbf{s}_i, \boldsymbol{\theta})|$ is the determinant of $\mathbf{M}(\mathbf{s}_i, \boldsymbol{\theta})$.

2.2.2 Criterion for covariance parameter estimation

As mentioned in Chapter 1, it is more reasonable in practice to assume that the covariance structure is known up to some unknown parameters $\boldsymbol{\theta}$, and the optimal design in this case may depend on which estimator $\hat{\boldsymbol{\theta}}$ we use, and how we measure the quality of the estimator $\hat{\boldsymbol{\theta}}$. It is easy to extend the results of the univariate case to the multivariate case: the MSE matrix $\mathbf{M}(\boldsymbol{\theta}) = E[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})']$ can be approximated by the inverse of the Fisher information matrix, if the sample size is sufficiently large. Under model (2.2), assuming $\mathbf{Z}$ is a Gaussian random process, the ij^{th} element of the information matrix associated with the ML estimator, $\mathbf{I}_{\mathrm{ML}}(\boldsymbol{\theta})$ is given by

$$\frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j})$$

while the ij^{th} element of $\mathbf{I}_{\mathrm{REML}}(\boldsymbol{\theta})$, the information matrix associated with the REML estimator, is given by

$$\frac{1}{2}\mathrm{tr}(\mathbf{P}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}\mathbf{P}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j})$$

where $\mathbf{P} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}$. Notice that the expressions for $\mathbf{I}_{\mathrm{ML}}(\boldsymbol{\theta})$ and $\mathbf{I}_{\mathrm{REML}}(\boldsymbol{\theta})$ are the same as in the univariate case (Mardia and Marshall 1984, and McCullagh and Nelder 1989, p.247). Optimal design can be obtained by minimizing $|\mathbf{M}(\boldsymbol{\theta})|$, or in other words, by minimizing one of the following two:

$$1/|\mathbf{I}_{\mathrm{ML}}(\boldsymbol{\theta})| \text{ or } 1/|\mathbf{I}_{\mathrm{REML}}(\boldsymbol{\theta})|$$

depending on whether we use maximum likelihood or restricted maximum likelihood to estimate $\boldsymbol{\theta}$.

2.2.3 Criteria for empirical prediction

In practice, the problem of most interest is to predict the unobserved value of $\mathbf{Z}(\mathbf{s}_0)$ when the covariance parameters are unknown. Neither of the previous two design criteria fully address this issue. In order to predict $\mathbf{Z}(\mathbf{s}_0)$ in this situation, we need to estimate the parameters $\boldsymbol{\theta}$ and then plug the estimated parameters into the expression of the BLUP, which is known as the empirical best linear predictor or E-BLUP. For example, if $\hat{\mathbf{Z}}(\mathbf{s}_0, \boldsymbol{\theta}) = \begin{pmatrix} \hat{Z}_1(\mathbf{s}_0, \boldsymbol{\theta}) \\ \hat{Z}_2(\mathbf{s}_0, \boldsymbol{\theta}) \\ \vdots \\ \hat{Z}_m(\mathbf{s}_0, \boldsymbol{\theta}) \end{pmatrix}$ is the known-covariance-parameter BLUP at $\mathbf{s}_0$, and $\hat{\boldsymbol{\theta}}$ is the estimator of $\boldsymbol{\theta}$, then $\hat{\mathbf{Z}}(\mathbf{s}_0, \hat{\boldsymbol{\theta}})$ is the E-BLUP evaluated at $\hat{\boldsymbol{\theta}}$. We therefore need a criterion that measures the quality of the E-BLUP as a predictor of $\mathbf{Z}(\mathbf{s}_0)$

Just as in the univariate case, the E-BLUP is unbiased under very weak conditions. The following theorem can be used to derive useful criteria for multivariate empirical prediction.

Theorem 2.1 *Assume* $\mathbf{Z}$ *follows linear model (2.2), and let the parameter space* Ω *be the set of* $\boldsymbol{\theta}$*-vectors for which* $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta}) = \mathrm{var}(\mathbf{Z})$ *is positive definite, where* $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)^T$ *is a* $p \times 1$ *vector. Let* $\boldsymbol{\tau} = \boldsymbol{\Lambda}'\boldsymbol{\beta} + \mathbf{u}$ *be a predictable linear combination, where* $\boldsymbol{\Lambda}' = \mathrm{diag}(\boldsymbol{\lambda_1}', \boldsymbol{\lambda_2}', \ldots, \boldsymbol{\lambda_m}')$, $\mathrm{cov}(\boldsymbol{\delta}, \mathbf{u}) = \mathbf{K}$, *and* $\mathbf{W} = \mathbf{W}(\boldsymbol{\theta}) = \mathrm{var}(\mathbf{u})$*; let* $\mathbf{p}(\mathbf{Z}, \boldsymbol{\theta})$ *be the BLUP of* $\boldsymbol{\tau}$*; let* $\hat{\boldsymbol{\theta}}$ *be an estimator of* $\boldsymbol{\theta}$*; and let* $\hat{\boldsymbol{\tau}} = \mathbf{p}(\mathbf{Z}, \hat{\boldsymbol{\theta}})$ *be the empirical BLUP of* $\boldsymbol{\tau}$*. Assume that*

$$\begin{pmatrix} \mathbf{Z} \\ \mathbf{u} \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{X}\boldsymbol{\beta} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma} & \mathbf{K} \\ \mathbf{K} & \mathbf{W} \end{pmatrix} \right)$$

and $\hat{\boldsymbol{\tau}}$ *is unbiased. Also let* $m(\boldsymbol{\theta}) = \mathrm{var}[\mathbf{p}(\mathbf{Z}, \boldsymbol{\theta}) - \boldsymbol{\tau}]$ *and* $m_E(\boldsymbol{\theta}) = \mathrm{var}[\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}]$.

(1) *Suppose* $\hat{\boldsymbol{\tau}} - \mathbf{p}(\mathbf{Z}, \boldsymbol{\theta})$ *depends on* $\mathbf{Z}$ *only through the value of* $\mathbf{Z} - \mathbf{X}\hat{\boldsymbol{\beta}}$*, where* $\hat{\boldsymbol{\beta}}$ *is the solution to the Aitken equations. Then* $m_E(\boldsymbol{\theta}) = m(\boldsymbol{\theta}) + \mathrm{E}((\hat{\boldsymbol{\tau}} - \mathbf{p}(\mathbf{Z}, \boldsymbol{\theta}))(\hat{\boldsymbol{\tau}} - \mathbf{p}(\mathbf{Z}, \boldsymbol{\theta}))')$

(2) *$m_E(\boldsymbol{\theta})$ can be approximated by*

$$m_E(\boldsymbol{\theta}) \doteq m(\boldsymbol{\theta}) + \mathrm{E}((\frac{\partial \mathbf{p}(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})'(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})'(\frac{\partial \mathbf{p}(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}))$$

(3) *Assume $\frac{\partial \mathbf{p}(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}$ are independent,then based on part **(1)** and **(2)**, an approximation of $m_E(\boldsymbol{\theta})$ is given by*

$$m_E(\boldsymbol{\theta}) \doteq m(\boldsymbol{\theta}) + \begin{pmatrix} tr(\mathbf{A}_{11}(\boldsymbol{\theta})\mathbf{H}(\boldsymbol{\theta})) & tr(\mathbf{A}_{12}(\boldsymbol{\theta})\mathbf{H}(\boldsymbol{\theta})) & \cdots & tr(\mathbf{A}_{1m}(\boldsymbol{\theta})\mathbf{H}(\boldsymbol{\theta})) \\ tr(\mathbf{A}_{21}(\boldsymbol{\theta})\mathbf{H}(\boldsymbol{\theta})) & tr(\mathbf{A}_{22}(\boldsymbol{\theta})\mathbf{H}(\boldsymbol{\theta})) & \cdots & tr(\mathbf{A}_{2m}(\boldsymbol{\theta})\mathbf{H}(\boldsymbol{\theta})) \\ \vdots & \vdots & & \vdots \\ tr(\mathbf{A}_{m1}(\boldsymbol{\theta})\mathbf{H}(\boldsymbol{\theta})) & tr(\mathbf{A}_{m2}(\boldsymbol{\theta})\mathbf{H}(\boldsymbol{\theta})) & \cdots & tr(\mathbf{A}_{mm}(\boldsymbol{\theta})\mathbf{H}(\boldsymbol{\theta})) \end{pmatrix}, \tag{2.4}$$

where $\mathbf{A}_{ij}(\boldsymbol{\theta}) = \mathrm{cov}(\frac{\partial p_i(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \frac{\partial p_j(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})$ and $\mathbf{H}(\boldsymbol{\theta}) = E[(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})']$ or an approximation of it.

Proof:

(1) Since $\boldsymbol{\tau}$ is a predictable linear combination, we may write $\boldsymbol{\Lambda}' = \mathbf{A}'\mathbf{X}$ for some matrix $\mathbf{A}$, and we have

$$\begin{aligned} \mathbf{p}(\mathbf{Z},\boldsymbol{\theta}) - \boldsymbol{\tau} &= \boldsymbol{\Lambda}'\hat{\boldsymbol{\beta}} + \mathbf{K}'\boldsymbol{\Sigma}^{-1}(\mathbf{Z}-\mathbf{X}\hat{\boldsymbol{\beta}}) - \boldsymbol{\Lambda}'\boldsymbol{\beta} - \mathbf{u} \\ &= \mathbf{A}'\mathbf{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) + \mathbf{K}'\boldsymbol{\Sigma}^{-1}(\mathbf{Z}-\mathbf{X}\hat{\boldsymbol{\beta}}) - \mathbf{u}. \end{aligned}$$

Based on the assumption, $\hat{\boldsymbol{\tau}} - \mathbf{p}(\mathbf{Z},\boldsymbol{\theta}) = g(\mathbf{Z}-\mathbf{X}\hat{\boldsymbol{\beta}})$, furthermore, $\begin{pmatrix} \mathbf{Z}-\mathbf{X}\hat{\boldsymbol{\beta}} \\ \mathbf{p}(\mathbf{y},\boldsymbol{\theta}) - \boldsymbol{\tau} \end{pmatrix}$

have a multivariate normal joint distribution, since

$$\begin{aligned}
& \mathrm{cov}(\mathbf{p}(\mathbf{Z},\boldsymbol{\theta})-\boldsymbol{\tau}, \mathbf{Z}-\mathbf{X}\hat{\boldsymbol{\beta}}) \\
= \;& \mathbf{A}'\mathbf{X}\mathrm{cov}(\hat{\boldsymbol{\beta}}, \mathbf{Z}-\mathbf{X}\hat{\boldsymbol{\beta}}) + \mathbf{K}'\boldsymbol{\Sigma}^{-1}\mathrm{var}(\mathbf{Z}-\mathbf{X}\hat{\boldsymbol{\beta}}) - \mathrm{cov}(\mathbf{u}, \mathbf{Z}-\mathbf{X}\hat{\boldsymbol{\beta}}) \\
= \;& \mathbf{A}'\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}(\mathbf{I}-\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}'\boldsymbol{\Sigma}^{-1})' + \\
& \mathbf{K}'\boldsymbol{\Sigma}^{-1}(\mathbf{I}-\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}'\boldsymbol{\Sigma}^{-1})\boldsymbol{\Sigma}(\mathbf{I}-\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}'\boldsymbol{\Sigma}^{-1})' \\
& -\mathbf{K}'(\mathbf{I}-\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}'\boldsymbol{\Sigma}^{-1})' \\
= \;& \mathbf{A}'\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}'(\mathbf{I}-\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}'\boldsymbol{\Sigma}^{-1})' + \\
& \mathbf{K}'(\mathbf{I}-\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}'\boldsymbol{\Sigma}^{-1})' - \mathbf{K}'(\mathbf{I}-\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-}\mathbf{X}'\boldsymbol{\Sigma}^{-1})' \\
= \;& \mathbf{0},
\end{aligned}$$

so $\mathbf{p}(\mathbf{Z},\boldsymbol{\theta})-\boldsymbol{\tau}$ and $\mathbf{Z}-\mathbf{X}\hat{\boldsymbol{\beta}}$ are independent, hence $\mathbf{p}(\mathbf{Z},\boldsymbol{\theta})-\boldsymbol{\tau}$ and $\hat{\boldsymbol{\tau}}-\mathbf{p}(\mathbf{Z},\boldsymbol{\theta})$ are independent. Therefore,

$$\begin{aligned}
m_E(\boldsymbol{\theta}) &= \mathrm{var}(\hat{\boldsymbol{\tau}}-\boldsymbol{\tau}) \\
&= \mathrm{E}((\hat{\boldsymbol{\tau}}-\boldsymbol{\tau})(\hat{\boldsymbol{\tau}}-\boldsymbol{\tau})') \quad (\hat{\tau} \text{ is unbiased}) \\
&= \mathrm{E}((\hat{\boldsymbol{\tau}}-\mathbf{p}(\mathbf{Z},\boldsymbol{\theta})+\mathbf{p}(\mathbf{Z},\boldsymbol{\theta})-\boldsymbol{\tau})(\hat{\boldsymbol{\tau}}-\mathbf{p}(\mathbf{Z},\boldsymbol{\theta})+\mathbf{p}(\mathbf{Z},\boldsymbol{\theta})-\boldsymbol{\tau})') \\
&= \mathrm{E}((\hat{\boldsymbol{\tau}}-\mathbf{p}(\mathbf{Z},\boldsymbol{\theta}))(\hat{\boldsymbol{\tau}}-\mathbf{p}(\mathbf{Z},\boldsymbol{\theta}))' + (\mathbf{p}(\mathbf{Z},\boldsymbol{\theta})-\boldsymbol{\tau})(\mathbf{p}(\mathbf{Z},\boldsymbol{\theta})-\boldsymbol{\tau})' + \\
&\quad (\mathbf{p}(\mathbf{Z},\boldsymbol{\theta})-\boldsymbol{\tau})(\hat{\boldsymbol{\tau}}-\mathbf{p}(\mathbf{Z},\boldsymbol{\theta}))' + (\hat{\boldsymbol{\tau}}-\mathbf{p}(\mathbf{Z},\boldsymbol{\theta}))(\mathbf{p}(\mathbf{Z},\boldsymbol{\theta})-\boldsymbol{\tau})') \\
&= m(\boldsymbol{\theta}) + E((\hat{\boldsymbol{\tau}}-\mathbf{p}(\mathbf{Z},\boldsymbol{\theta}))(\hat{\boldsymbol{\tau}}-\mathbf{p}(\mathbf{Z},\boldsymbol{\theta}))').
\end{aligned}$$

(2) Expand $\hat{\boldsymbol{\tau}}-\mathbf{p}(\mathbf{Z},\boldsymbol{\theta})$ in a Taylor series about $\boldsymbol{\theta}$

$$\begin{aligned}
\hat{\boldsymbol{\tau}}-\mathbf{p}(\mathbf{Z},\boldsymbol{\theta}) &= \mathbf{p}(\mathbf{Z},\hat{\boldsymbol{\theta}})-\mathbf{p}(\mathbf{Z},\boldsymbol{\theta}) \\
&= (\frac{\partial \mathbf{p}(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})'(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}) + O(||\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}||^2),
\end{aligned}$$

hence $m_E(\boldsymbol{\theta}) \doteq m(\boldsymbol{\theta}) + \mathrm{E}((\frac{\partial \mathbf{p}(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})'(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})'(\frac{\partial \mathbf{p}(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}))$, where $(\frac{\partial \mathbf{p}(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}) = \begin{pmatrix} \frac{\partial p_1(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} & \frac{\partial p_2(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} & \dots & \frac{\partial p_m(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \end{pmatrix}$ is a $p \times m$ matrix.

(3)

$$(\frac{\partial \mathbf{p}(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})'(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})'(\frac{\partial \mathbf{p}(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})$$

$$= \begin{pmatrix} (\frac{\partial p_1(\mathbf{y},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})' \\ \frac{\partial p_2(\mathbf{y},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\ \vdots \\ \frac{\partial p_m(\mathbf{y},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \end{pmatrix} (\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})' \begin{pmatrix} \frac{\partial p_1(\mathbf{y},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} & \frac{\partial p_2(\mathbf{y},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} & \cdots & \frac{\partial p_m(\mathbf{y},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \end{pmatrix},$$

so the ij^{th} element of $\mathrm{E}((\frac{\partial \mathbf{p}(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})'(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})'(\frac{\partial \mathbf{p}(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}))$ is $\mathrm{E}((\frac{\partial p_i(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})'(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})'(\frac{\partial p_j(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}))$. We can show that

$$\begin{aligned}
& \mathrm{E}((\frac{\partial p_i(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})'(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})'(\frac{\partial p_j(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})) \\
= \ & \mathrm{tr}(\mathrm{E}((\frac{\partial p_i(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})'(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})'(\frac{\partial p_j(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}))) \\
= \ & \mathrm{E}(\mathrm{tr}((\frac{\partial p_i(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})'(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})'(\frac{\partial p_j(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}))) \\
= \ & \mathrm{E}(\mathrm{tr}((\frac{\partial p_j(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})'(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})'(\frac{\partial p_i(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}))) \\
= \ & \mathrm{E}(\mathrm{tr}((\frac{\partial p_i(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})(\frac{\partial p_j(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})'(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})')) \\
= \ & \mathrm{tr}(\mathrm{E}((\frac{\partial p_i(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})(\frac{\partial p_j(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})'(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})')) \\
= \ & \mathrm{tr}(\mathrm{E}((\frac{\partial p_i(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})(\frac{\partial p_j(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})')\mathrm{E}((\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})')) \\
& \text{(by the assumption that } \frac{\partial \mathbf{p}(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \text{ and } \hat{\boldsymbol{\theta}}-\boldsymbol{\theta} \text{ are independent).}
\end{aligned}$$

Notice that $\mathrm{E}(\frac{\partial \mathbf{p}(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}) = \frac{\partial}{\partial \boldsymbol{\theta}}\mathrm{E}(\mathbf{p}(\mathbf{Z},\boldsymbol{\theta})) = \frac{\partial}{\partial \boldsymbol{\theta}}(\boldsymbol{\lambda}'\boldsymbol{\beta}) = 0$, so

$$\mathbf{A}_{ij}(\boldsymbol{\theta}) = \mathrm{cov}(\frac{\partial p_i(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \frac{\partial p_j(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}) = \mathrm{E}((\frac{\partial p_i(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})(\frac{\partial p_j(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})').$$

Hence

$$\mathrm{E}((\frac{\partial p_i(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})'(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})'(\frac{\partial p_j(\mathbf{Z},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})) = \mathrm{tr}(\mathbf{A}_{ij}(\boldsymbol{\theta})\mathbf{H}(\boldsymbol{\theta})),$$

and equation (2.4) gives an approximation to $m_E(\boldsymbol{\theta})$. □

Applying Theorem 2.1 to our case, we obtain

$$\mathbf{M}(\mathbf{s_0},\hat{\boldsymbol{\theta}}) \doteq \mathbf{M}(\mathbf{s_0}) + \begin{pmatrix} tr(\mathbf{A}_{11}(\boldsymbol{\theta})\mathbf{H}(\boldsymbol{\theta})) & \cdots & tr(\mathbf{A}_{1m}(\boldsymbol{\theta})\mathbf{H}(\boldsymbol{\theta})) \\ \vdots & & \vdots \\ tr(\mathbf{A}_{m1}(\boldsymbol{\theta})\mathbf{H}(\boldsymbol{\theta})) & \cdots & tr(\mathbf{A}_{mm}(\boldsymbol{\theta})\mathbf{H}(\boldsymbol{\theta})) \end{pmatrix}, \qquad (2.5)$$

where $\mathbf{M}(\mathbf{s_0})$ is the mean square prediction error matrix associated with BLUP, $\mathbf{A}_{ij}(\boldsymbol{\theta}) = \mathrm{cov}(\frac{\partial \hat{Z}_i(\mathbf{s}_0,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \frac{\partial \hat{Z}_j(\mathbf{s}_0,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}})$, and $\mathbf{H}(\boldsymbol{\theta})$ can be approximated by $\mathbf{I}_{ML}(\boldsymbol{\theta})$ or $\mathbf{I}_{REML}(\boldsymbol{\theta})$ if we use

maximum likelihood or residual maximum likelihood to estimate $\boldsymbol{\theta}$. To get the optimal design, one can minimize either of the following two:

$$\frac{1}{|S|}\sum_{\mathbf{s}_i\in S}|\mathbf{M}(\mathbf{s}_i,\hat{\boldsymbol{\theta}})| \quad \text{or} \quad \max_{\mathbf{s}_i\in S}|\mathbf{M}(\mathbf{s}_i,\hat{\boldsymbol{\theta}})|.$$

Based on equation (2.5), the criteria for E-BLUP give weight to both prediction quality under known covariance structure ($\mathbf{M}(\mathbf{s_0})$) and covariance parameter estimation quality ($\mathbf{H}(\boldsymbol{\theta})$).

2.3 Toy example

2.3.1 The design problem

In the following case studies, for the purpose of illustration, consider the bivariate case, that is, $\mathbf{Z}(\mathbf{s}) = \begin{pmatrix} Z_1(\mathbf{s}) \\ Z_2(\mathbf{s}) \end{pmatrix}$ at any point $\mathbf{s} \in D$. Unless noted otherwise, we assume $\mathbf{Z}(\mathbf{s})$ follows model (2.2), has a constant mean function over space for each variable and a symmetric separable exponential covariance model, i.e.,

$$E(\mathbf{Z}(\mathbf{s})) \equiv \boldsymbol{\mu}(\mathbf{s}) = \boldsymbol{\mu}, \tag{2.6}$$

$$C_{ij}(\mathbf{h}) = \text{cov}(Z_i(\mathbf{s}+\mathbf{h}), Z_j(\mathbf{s})) = \text{cov}(Z_j(\mathbf{s}+\mathbf{h}), Z_i(\mathbf{s})) = C_{ji}(\mathbf{h}), \tag{2.7}$$

$$C_{ii}(\mathbf{h}) = \sigma^2\rho^{\|\mathbf{h}\|}, C_{ij}(\mathbf{h}) = \sigma^2\rho_c\rho^{\|\mathbf{h}\|} \tag{2.8}$$

where $\|\mathbf{h}\|$ is the Euclidean distance between any two points. We consider six combinations of ρ and ρ_c, three of which are $(\rho, \rho_c) = (\rho, 0.50)$, with $\rho = 0.20, 0.50, 0.80$; the other three being $(\rho, \rho_c) = (0.50, \rho_c)$, with $\rho_c = 0.20, 0.50, 0.80$.

Consider a 4×4 square point grid, where the distance between any two adjacent points is one unit. Our design problem is to choose from among the 16 grid points, 4 points for each variable such that the corresponding criterion can be minimized. The size of the grid and the size of the samples are determined mainly based on the computational effort involved. There are $C_4^{16} \cdot C_4^{16} = 3,312,400$ designs in total for this design problem, but not all of them are distinct, due to symmetries of the designs with respect to rotations or reflections of the points. By enumerating all 3,312,400 designs, we can determine the optimal designs with respect to each of prediction, covariance parameter estimation and

empirical prediction.

For each of the three criteria, we also obtain the median criterion value from all designs. Comparison of these median criterion values to the optimized values allows us to measure how much improvement we obtain by optimizing the design.

2.3.2 Optimal design for prediction

Table 2.1 contains the optimized criteria for prediction on the 4×4 grid, and Figure 2.1 shows the optimal designs corresponding to each ρ and ρ_c combination, when using $\max_{i \in S} |\mathbf{M}(\mathbf{s}_i, \boldsymbol{\theta})|$ as the design criterion. From Table 2.1, the optimized criterion decreases as ρ or ρ_c increase, while keeping the other parameter constant. This is as expected, as when ρ or ρ_c increase, the spatial correlation between any two locations for a single variable or the cross-correlation between the two variables increases. Hence, more information can be borrowed from observed data when predicting a specific unsampled point, which leads to smaller prediction error variance. In all the optimal designs, the four design points for each variable are widely dispersed, which is the same as the univariate design situation. As pointed out in Zimmerman (2006), this is intuitively reasonable, as it prevents any prediction site from being too far from the closest design point. However, there are some important differences with respect to the degree of collocation, which is defined, here and throughout, as the ratio of the number of collocated points to the total number of candidate points in the corresponding design. The optimal design changes from completely collocated to completely non-collocated when ρ increases, while the optimal design changes from completely non-collocated to completely collocated when ρ_c increases.

The percentages in the parentheses in Table 2.1 represent the improvement in the criterion for the optimal design relative to the corresponding median criterion value from all designs. We can see improvement is considerable (all are above 27%), indicating that the optimal design does provide valuable improvement in efficiency for statistical inferences.

Using $\frac{1}{|S|} \sum_{\mathbf{s}_i \in S} |\mathbf{M}(\mathbf{s}_i)|$ as the design criterion yields similar results, which are not presented in this thesis.

ρ	0.20	0.50	0.80
$\rho_c = 0.50$	0.931(27%)	0.441(47%)	0.067(65%)
ρ_c	0.20	0.50	0.80
$\rho = 0.50$	0.526(48%)	0.441(47%)	0.222(50%)

Table 2.1: $\min\{\max_{\mathbf{s}_i \in S} |\mathbf{M}(\mathbf{s}_i, \boldsymbol{\theta})|\}$ for prediction on the 4×4 grid

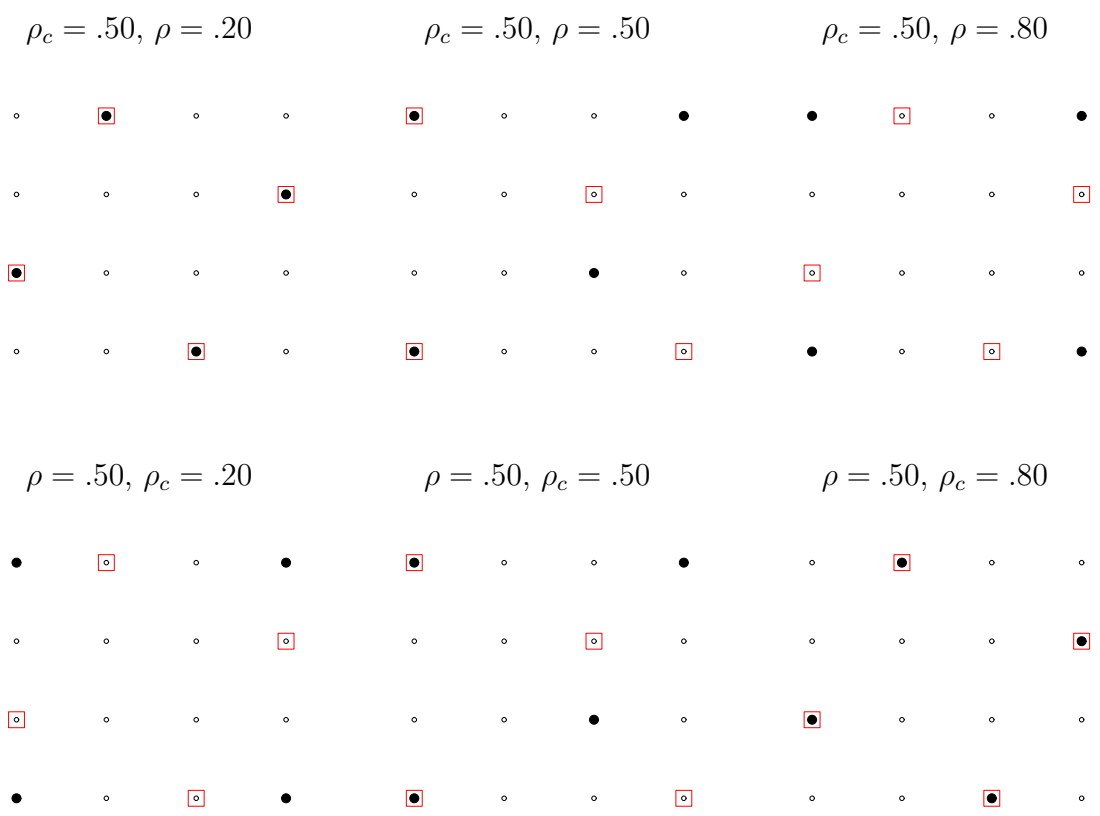

Figure 2.1: Optimal designs for prediction on the 4×4 grid; black solid dots are for variable 1 and red empty squares are for variable 2

2.3.3 Optimal design for covariance parameter estimation

Table 2.2 contains the optimized criteria for covariance estimation, and Figure 2.2 shows the optimal designs corresponding to each ρ and ρ_c combination, when using $1/|\mathbf{I}_{REML}(\boldsymbol{\theta})|$ as the design criterion. As can be seen from Table 2.2, the optimized criterion decreases as ρ_c increases, but does not have a monotone trend when ρ increases. This indicates that larger ρ_c results in more informative design than smaller ρ_c for parameter estimation. The spatial configuration of the optimal designs indicates some important characteristics for the optimal design in this case. First, all the points in the optimal design are completely collocated, which means that the design points for the two variables are exactly the same. Second, the optimal designs seem to be rather invariant to the values of ρ and ρ_c. Third, in most case there are two clusters in all the optimal designs and the two clusters are far apart from each other. When $\rho = 0.8$ and $\rho_c = 0.5$, all the points are elongated into a linear strand. These are quite different from the optimal designs for prediction. In Section 2.3.2, we also observed a completely collocated design when the effect of ρ_c is dominant, but those collocated points are widely spread out. As pointed out in Zimmerman (2006) for a univariate situation, it is very useful to have some small lags in the design for precise estimation of the covariance parameters, but not for prediction with known covariance parameters.

Table 2.2 also shows that the optimal design tends to be much better than the median design. In all cases, the criterion improves by more than 87% compared to the median value of all designs.

The spatial configuration of the optimal designs corresponding to ML estimation look somewhat different from those for REML estimation. Figure 2.3 shows the optimal designs for ML estimation. All the points are completely collocated just as for REML estimation. When ρ is small the design is a cluster around the corner, but as ρ increases the cluster tends to elongate into a linear strand, which is exactly the same trend observed in the univariate case (Zimmerman 2006). We will focus on the REML estimation criterion for covariance estimation in the rest of this thesis as optimal designs for ML estimation

ρ	0.20	0.50	0.80
$\rho_c = 0.50$	0.024(92%)	0.026(89%)	0.025(87%)
ρ_c	0.20	0.50	0.80
$\rho = 0.50$	0.042(88%)	0.026(89%)	0.006(93%)

Table 2.2: $\min\{1/|\mathbf{I}_{REML}(\boldsymbol{\theta})|\}$ for covariance estimation on the 4×4 grid

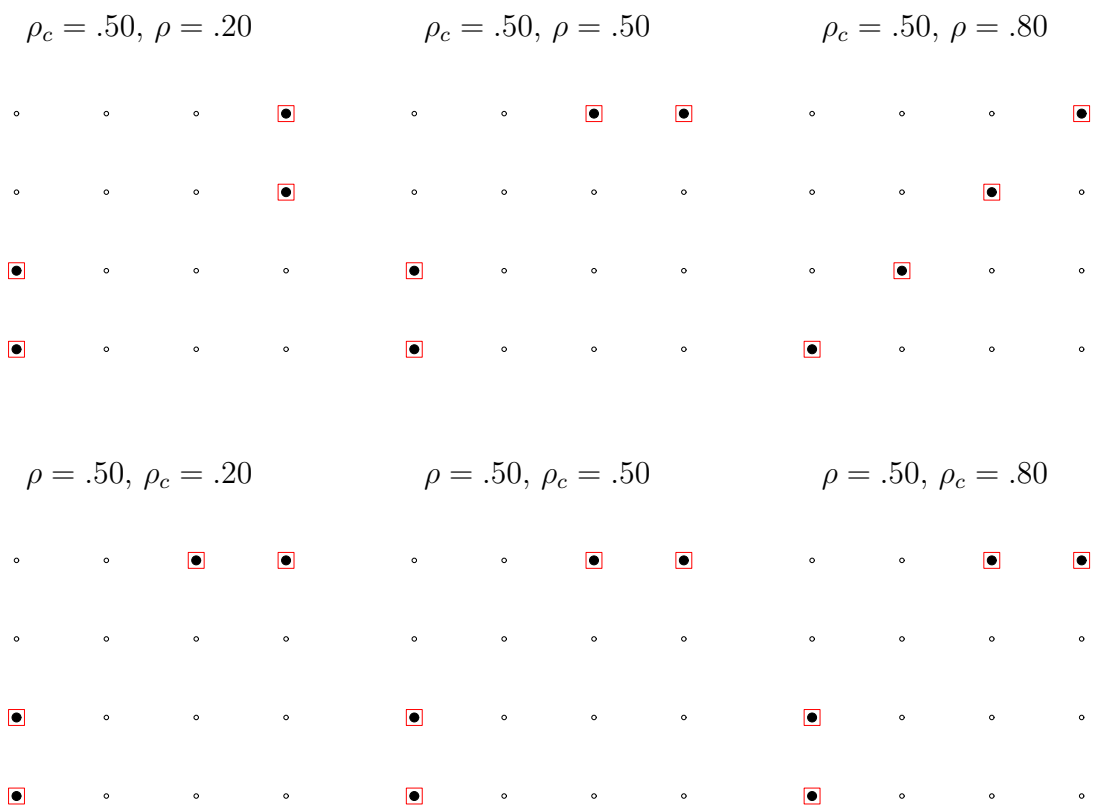

Figure 2.2: Optimal designs for covariance estimation (REML estimation) on the 4×4 grid; black solid dots are for variable 1 and red empty squares are for variable 2

have many similar characteristics as optimal designs for REML estimation.

2.3.4 Optimal design for empirical prediction

Table 2.3 contains the optimized criteria for empirical prediction, and Figure 2.4 shows the optimal designs corresponding to each ρ and ρ_c combination. From Table

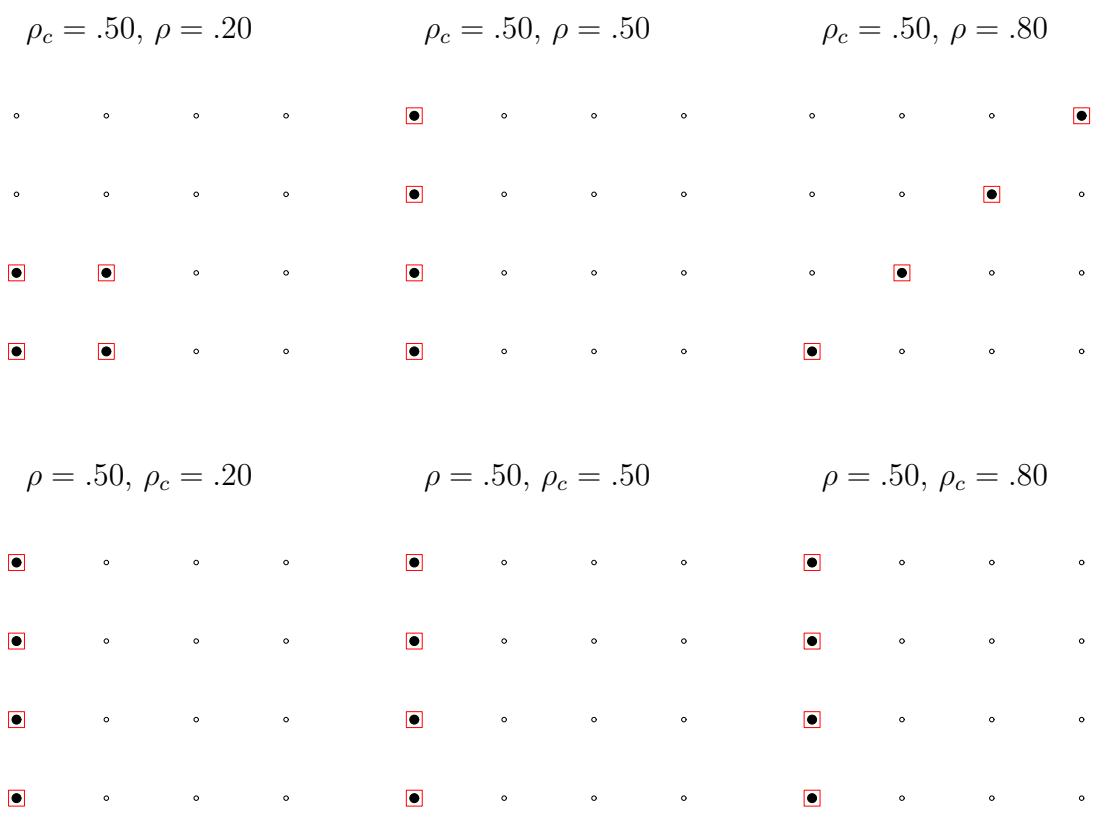

Figure 2.3: Optimal designs for covariance estimation (ML estimation) on the 4×4 grid; black solid dots are for variable 1 and red empty squares are for variable 2

ρ	0.20	0.50	0.80
$\rho_c = 0.50$	2.719(64%)	4.300(69%)	65.209(72%)
ρ_c	0.20	0.50	0.80
$\rho = 0.50$	5.724(70%)	4.300(69%)	1.788(74%)

Table 2.3: $\min\{\max_{i \in S} |\mathbf{M}(\mathbf{s}_i, \hat{\boldsymbol{\theta}})|\}$ for empirical prediction on the 4×4 grid

2.3, the optimized criterion increases as ρ increases and decreases as ρ_c increases. The optimal design is not as easy to characterize as in the previous cases, and it appears to vary more as the covariance parameters change. Usually it is not completely collocated, and it is different from the corresponding design that is optimal for prediction. These characteristics show that the criterion for empirical prediction is a combination of that for prediction and covariance parameter estimation in some sense, by the nature of the empirical predictor.

The improvements in the criterion relative to the median of all designs are above 60% in every case, an amount which lies between that for prediction and covariance parameter estimation. Thus, optimal design for empirical prediction yields even larger relative gains in efficiency than optimal design for prediction with known covariance parameters.

2.4 Large example using simulated annealing algorithm

2.4.1 The design problem

There are obvious limitations to studying optimal spatial design using only the toy example on a 4×4 grid. Therefore, we also consider a substantially larger design. Now let S be a 25×25 square grid with unit spacing, so the total number of points is 625. Let the sample size for each variable be 15. In this case, it is impossible to enumerate all the possible combinations ($C_{15}^{625} \cdot C_{15}^{625}$ designs) to find the optimal design, even though not all of them are distinct. Instead, we use a simulated annealing algorithm (SAA) to improve an initially chosen design. The algorithm contains the following steps (for more details

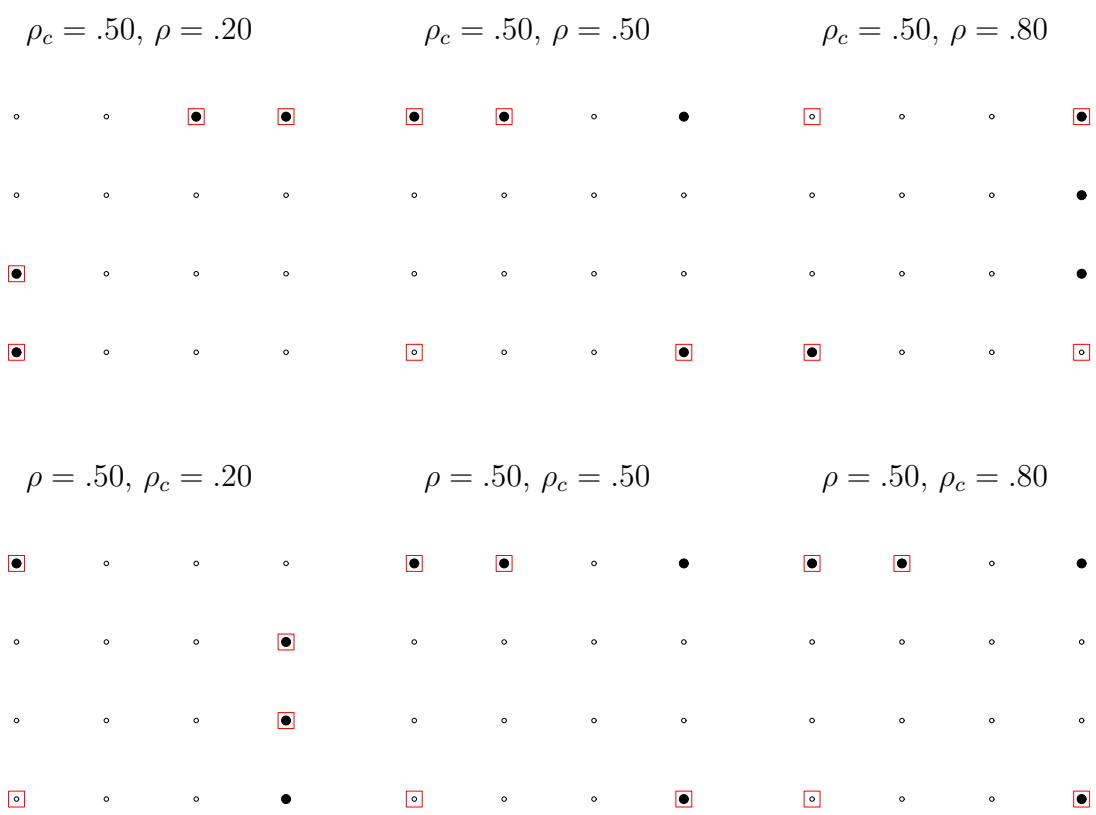

Figure 2.4: Optimal designs for empirical prediction on the 4×4 grid; black solid dots are for variable 1 and red empty squares are for variable 2

about the application of SAA to spatial design see Lark (2002)):

1. Randomly choose a design S_0, then compute $\Phi(S_0)$, where Φ is the corresponding criterion function that we wish to minimize. Set the initial values for the "cooling factor" and "distance factor", τ_0 and h_{max_0}, respectively.

2. For each point in design S_0, randomly select a direction, and then move this point h units in the selected direction, where $h \sim unif[0, h_{max_0}]$. Here h_{max_0} is the distance of the neighbor that each point in S_0 can at most move to. If the perturbation takes the point to a location outside the design region, then the point is returned to its original location and a new perturbation is generated at random until the perturbed point falls into the design region. Call this new design S_1.

3. Calculate $\Phi(S_1)$. Then the transition $S_0 \to S_1$ is accepted with the following probability:
$$P_\tau(S_0 \to S_1) = \begin{cases} 1, & \text{if } \Phi(S_1) \leq \Phi(S_0); \\ \exp\{\frac{\Phi(S_0)-\Phi(S_1)}{\tau_0}\}, & \text{if } \Phi(S_1) > \Phi(S_0). \end{cases}$$

4. Repeat steps 2-3 for N_{sim} (sufficiently large) times.

5. At the end of the N_{sim} repetitions, reduce the "cooling factor" and "distance factor" by $\tau_{k+1} = \alpha_\tau \tau_k$, $h_{max_{k+1}} = \alpha_h h_{max_k}$.

6. Repeat steps 2-5 for N_t times.

In this algorithm, τ_0 is set such that 95% or more of the perturbations will be accepted before the first cooling step; α_τ is set such that the acceptance rate will be below 0.1% at the final step; h_{max_0} is usually set at half the length of the region to be sampled; and α_h is set such that the final h_{max} is 0.5 unit.

For our specific case, the design region contains finitely many points, hence the region is discrete rather than continuous. Furthermore, we can only move the point towards four directions: north, south, east and west, instead of any random direction. Also, the distance to be moved is generated from a discrete uniform distribution, i.e., the distance must be an integer multiple of the unit distance. In all cases, we choose $N_{sim} = 100$ and $N_t = 1000$.

ρ	0.20	0.50	0.80
$\rho_c = 0.50$	0.866(3%)	0.856(5%)	0.467(45%)
ρ_c	0.20	0.50	0.80
$\rho = 0.50$	1.061(5%)	0.856(5%)	0.420(15%)

Table 2.4: $\min\{\max_{i \in S} |\mathbf{M}(\mathbf{s}_i, \boldsymbol{\theta})|\}$ for prediction using SAA

In our study, it took a few hours for the algorithm to converge to a (local) minimum, given a specific set of values for ρ and ρ_c, while using complete enumeration might take weeks or even longer. So the simulated annealing algorithm is much more feasible computationally. The results obtained by SAA can be affected by the initial design S_0, hence, different start-up designs (5 for each case) were chosen randomly and the results presented below are the best result we obtained. Although different start-up designs will lead to different final designs, the results are robust in the sense that the criterion values obtained and the spatial configuration of the design points do not differ much.

2.4.2 Designs produced by SAA for prediction

Table 2.4 contains the criterion values produced by SAA for prediction, and Figure 2.5 shows the designs produced by SAA corresponding to each ρ and ρ_c combination. The criterion and the degree of collocation in the design show the same trends as what we observed for prediction in the toy example. That is, the criterion decreases as ρ or ρ_c increase, while keeping the other parameter constant. The points in the designs are widely spread out, and the degree of collocation decreases as ρ increases and increases as ρ_c increases, which are the same trends as we observed in the toy example.

Table 2.4 shows positive improvements in the design criterion, but they are relatively small compared to their counterparts in the toy example. This is understandable, because (a) we do not enumerate all the possible designs to get the optimal one by using SAA, and (b) the designs have more points.

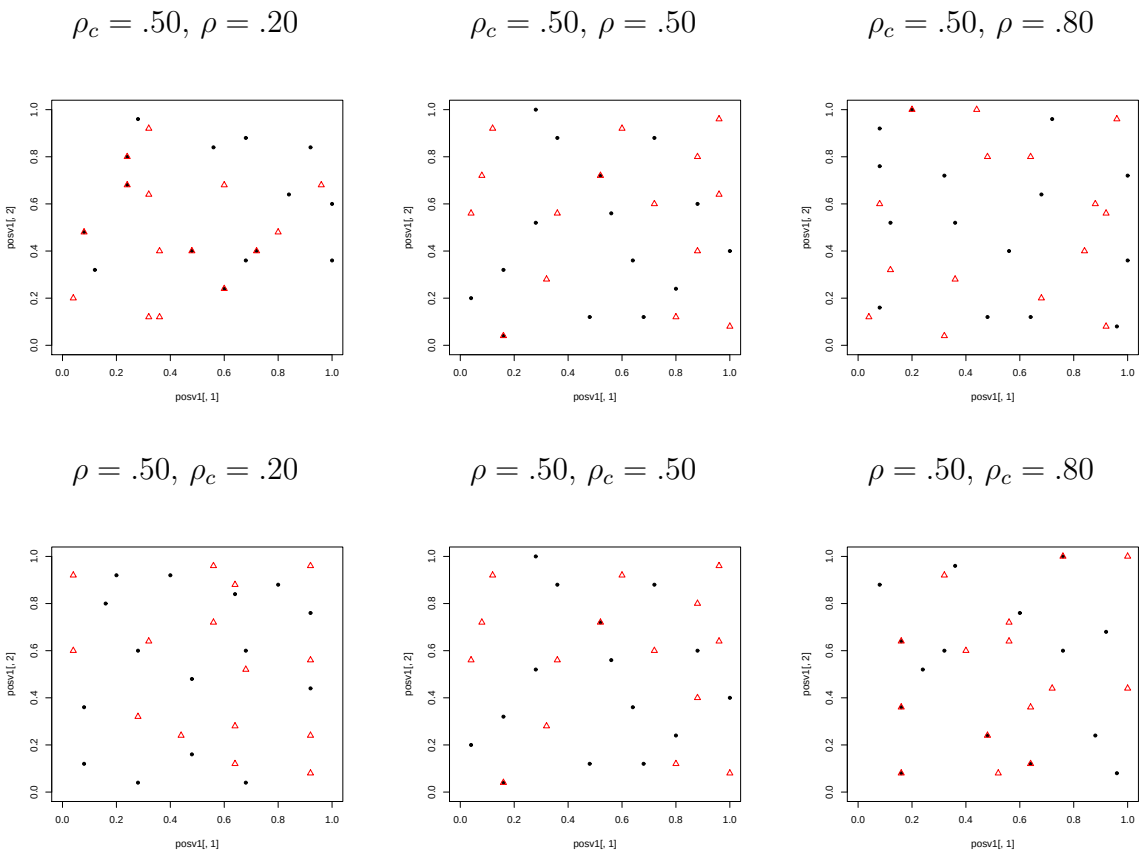

Figure 2.5: Designs for prediction using SAA; black solid dots are for variable 1 and red empty triangles are for variable 2

ρ	0.20	0.50	0.80
$\rho_c = 0.50$	0.000897(99.8%)	0.000404(95%)	0.000069(72%)
ρ_c	0.20	0.50	0.80
$\rho = 0.50$	0.000582(94%)	0.000404(95%)	0.000063(99%)

Table 2.5: $\min\{1/|\mathbf{I}_{REML}(\boldsymbol{\theta})|\}$ for covariance estimation using SAA

2.4.3 Designs produced by SAA for covariance parameter estimation

Table 2.5 contains the criterion values produced by SAA for covariance estimation, and Figure 2.6 shows the designs produced by SAA corresponding to each ρ and ρ_c combination. The criterion decreases when ρ or ρ_c increases. The configurations of the designs are different from the toy example; in particular, we do not have complete collocation here. Compared to the designs for prediction, there are obvious clusters in the designs produced by SAA for covariance parameter estimation while the points in the designs produced by SAA for prediction are widely spread out.

Table 2.5 also shows great improvements in the criterion for the designs produced by SAA relative to the corresponding median criterion value from 10,000 randomly chosen designs. This indicates that the simulated annealing algorithm works particularly well for the covariance parameter estimation case.

2.4.4 Designs produced by SAA for empirical prediction

Table 2.6 contains the criterion values produced by SAA for empirical prediction, and Figure 2.7 shows the designs produced by SAA corresponding to each ρ and ρ_c combination. The criterion decreases as ρ or ρ_c increases. Just like what was observed in the toy example, the designs for empirical prediction appear to sit in between those for prediction and covariance estimation. There are some small clusters but meanwhile those clusters are broadly scattered over the study region.

The scale of improvement in the criterion also lies between that for prediction and

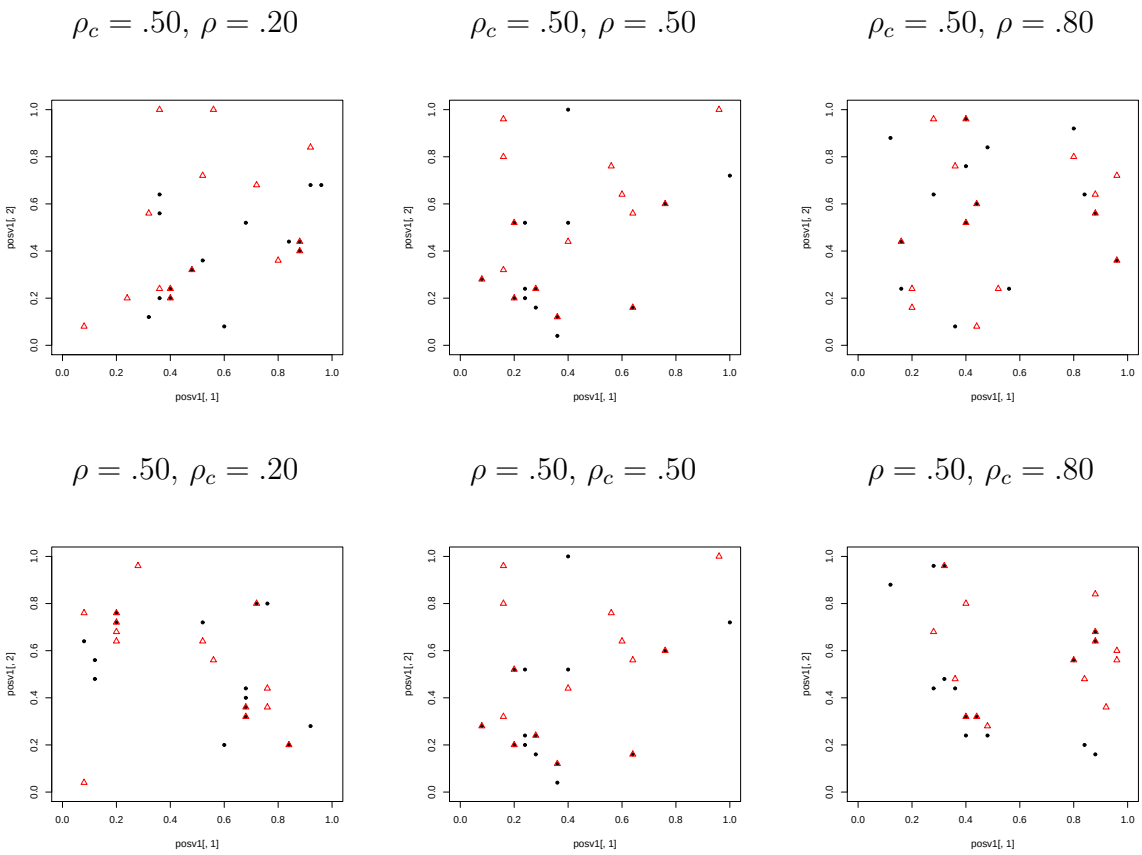

Figure 2.6: Designs for covariance estimation using SAA; black solid dots are for variable 1 and red empty triangles are for variable 2

ρ	0.20	0.50	0.80
$\rho_c = 0.50$	1.031(94%)	0.960(27%)	0.582(47%)
ρ_c	0.20	0.50	0.80
$\rho = 0.50$	1.213(23%)	0.960(27%)	0.459(53%)

Table 2.6: $\min\{\max_{i \in S} |\mathbf{M}(\mathbf{s}_i, \hat{\boldsymbol{\theta}})|\}$ for empirical prediction using SAA

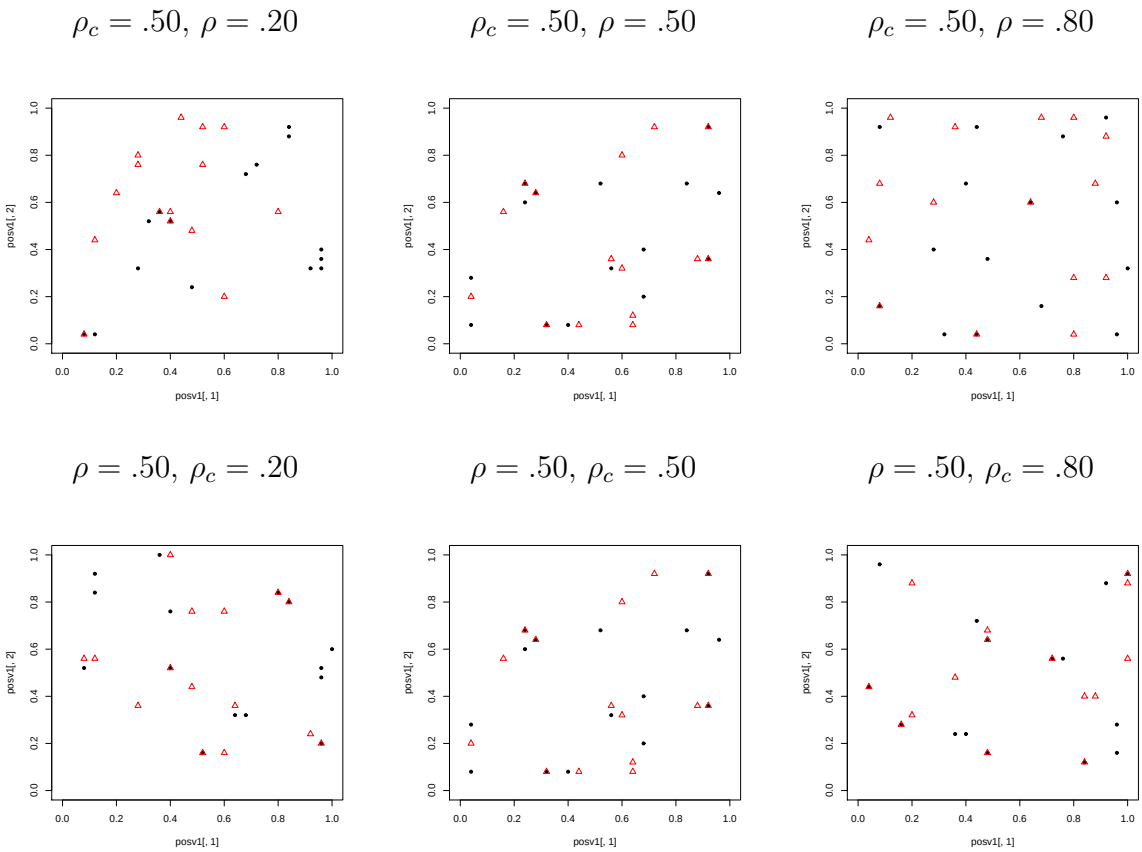

Figure 2.7: Designs for empirical prediction using SAA; black solid dots are for variable 1 and red empty triangles are for variable 2

covariance parameter estimation. Overall, there are substantial improvements in the design criterion compared to the randomly chosen designs with the least one being 23%.

CHAPTER 3
OPTIMAL DESIGNS ON STREAM NETWORKS

In many nations, streams and rivers are the most important natural water resources. Hence, lots of money and time has been spent to characterize and monitor streams and rivers. Since riverine transport of materials and movements of fish occur almost entirely within the stream network, Euclidean distance is not as appropriate for spatial modeling on streams as it is in the planar case. Alternatively, stream distance would seem to be more appropriate. However, one problem associated with the use of stream distance is that covariance models developed for spatial processes in the plane may not be valid for processes on stream networks.

3.1 Covariance model

We wish to use the moving average construction method which incorporates both stream distance and flow direction, as described by Ver Hoef and Peterson (2008), to construct valid multivariate covariance models for random processes using stream distance. Although univariate models on streams have been developed previously, multivariate models have not. Specifically, in this thesis, we will focus on developing a valid bivariate exponential covariance model to be used for optimal bivariate design on streams. As mentioned in Section 1.2, we will use a tail-up and a tail-down construction in order to incorporate flow direction. The tail-up construction does not allow autocorrelation between any two sites that are not flow connected (flow connected means there is water flowing from one of the sites to the other) and the tail of the moving average function goes upstream. The tail-down construction does allow autocorrelation between two sites that are not flow connected and the tail of the moving average function goes downstream.

Let $W_k(\mathbf{x})$ be a zero mean white noise random process with $\mathbf{x} \in \Re^d$, $k = 0, 1, 2$ which has the following property:

Property 3.1 *Let Ω denote the space spanned by all the segments of a stream network.*

Suppose A and B are two subsets of Ω, and let $|A|$ stand for the measure of set A. Then:

(1) $\text{var}[\int_A W_k(\mathbf{x})d\mathbf{x}] = |A|$, $k = 0, 1, 2$;

(2) $\text{cov}[\int_A W_k(\mathbf{x})d\mathbf{x}, \int_B W_k(\mathbf{x})d\mathbf{x}] = 0$, $k = 0, 1, 2$, *when* $A \bigcap B = \emptyset$;

(3) $W_k(\mathbf{x})$ *is independent of* $W_m(\mathbf{x})$ *when* $k \neq m$.

Consider a very simple stream network as shown in Figure 1.1. Let $f_1^1(x|\boldsymbol{\theta}_1) = \theta_{11}e^{-\theta_{12}x}I(x \geq 0)$, $f_1^2(x|\boldsymbol{\theta}_2) = \theta_{21}e^{-\theta_{22}x}I(x \geq 0)$, where $\theta_{ij} > 0, i = 1, 2, j = 1, 2$. We call $f_1^1(x)$ and $f_1^2(x)$ moving average functions. Let $-1 \leq \rho_k \leq 1$ and define white noise process $Y_k(\mathbf{x}|\rho_k) = \sqrt{1-\rho_k^2}W_k(\mathbf{x}) + \rho_k W_0(\mathbf{x})$, $k = 1, 2$. For the tail-up construction, we construct variables Z_1 and Z_2 using moving average functions as follows:

$$Z_1(s_i) = \int_{s_i}^{u_i} f_1^1(x_i - s_i|\boldsymbol{\theta}_1)Y_1(x_i|\rho_1)dx_i + \sum_{j \in U_{s_i}} \Big(\prod_{k \in B_{s_i,[j]}} \sqrt{w_k} \Big) \int_{l_j}^{u_j} f_1^1(x_j - s_i|\boldsymbol{\theta}_1)Y_1(x_j|\rho_1)dx_j,$$

$$Z_2(s_i) = \int_{s_i}^{u_i} f_1^2(x_i - s_i|\boldsymbol{\theta}_2)Y_2(x_i|\rho_2)dx_i + \sum_{j \in U_{s_i}} \Big(\prod_{k \in B_{s_i,[j]}} \sqrt{w_k} \Big) \int_{l_j}^{u_j} f_1^2(x_j - s_i|\boldsymbol{\theta}_2)Y_2(x_j|\rho_2)dx_j.$$

Here, l_i and u_i stand for the most downstream and the most upstream locations on the i^{th} segment. U_{s_i} is the index set of stream segments upstream of s_i, excluding i; B_{s_i,t_j} is the set of stream segments between s_i and t_j, including the segment for the upstream location (t_j) but excluding the segment for the downstream location (s_i); and w_k is the weight we put on the k^{th} segment according to flow volume. At each fork upstream of the i^{th} segment, we require that $w_j + w_k = 1$ where j and k are the segments immediately upstream from the fork.

Now we are ready to derive the bivariate exponential covariance model. Suppose $dist(s_i, t_j) = h$, and let $\text{cov}_{kl}(h) = \text{cov}(Z_k(s_i), Z_l(t_j))$ be the covariance between $Z_k(s_i)$ and $Z_l(t_j)$.

(1) If s_i and t_j are not flow connected, we have

$$\begin{aligned}
& \text{cov}\Big(\int_{s_i}^{u_i} f_1^1(x_i - s_i|\boldsymbol{\theta}_1)W_1(x_i)dx_i, \int_{t_j}^{u_j} f_1^1(x_j - t_j|\boldsymbol{\theta}_1)W_1(x_j)dx_j\Big) \\
= \; & \text{E}\Big(\int_{s_i}^{u_i} f_1^1(x_i - s_i|\boldsymbol{\theta}_1)W_1(x_i)dx_i \int_{t_j}^{u_j} f_1^1(x_j - t_j|\boldsymbol{\theta}_1)W_1(x_j)dx_j\Big) \\
= \; & \int_{s_i}^{u_i}\int_{t_j}^{u_j} f_1^1(x_i - s_i|\boldsymbol{\theta}_1)f_1^1(x_j - t_j|\boldsymbol{\theta}_1)\text{E}(W_1(x_i)W_1(x_j))dx_i dx_j \\
= \; & 0
\end{aligned}$$

by Property 3.1(2). Let $A = \{\text{segment } (s_i, u_i) \bigcup \text{segments } (l_j, u_j) \text{ for } j \in U_{s_i}\}$, let $B = \{\text{segment } (t_j, u_j) \bigcup \text{segments } (l_k, u_k) \text{ for } k \in U_{t_j}\}$, then since s_i and t_j are not flow connected and we are using tail-up construction, then $A \bigcap B = \emptyset$. Based on the above, it is easy to show that $\text{cov}_{11}(h) = 0$.

(2) If s_i and t_j are exactly the same point, i.e., if $h = 0$, then

$$\begin{aligned}
& \text{cov}_{11}(h) \\
= \; & \text{cov}(Z_1(s_i), Z_1(t_j)) \\
= \; & \text{var}(Z_1(s_i)) \\
= \; & \int_{s_i}^{u_i} \theta_{11}^2 e^{-2\theta_{12}(x_i - s_i)} I((x_i - s_i) \geq 0)[(1-\rho_1^2) + \rho_1^2] dx_i \\
& + \sum_{j \in U_{s_i}} \Big(\prod_{k \in B_{s_i,[j]}} \sqrt{w_k} \Big) \int_{l_j}^{u_j} \theta_{11}^2 e^{-2\theta_{12}(x_i - s_i)} I((x_i - s_i) \geq 0)[(1-\rho_1^2) + \rho_1^2] dx_j \\
= \; & \int_0^{\infty} \theta_{11}^2 e^{-2\theta_{12} x} dx \\
= \; & \frac{\theta_{11}^2}{2\theta_{12}}.
\end{aligned}$$

(3) If s_i and t_j are flow connected and $h > 0$, then

$$\begin{aligned}
& \text{cov}_{11}(h) \\
= \; & (1-\rho_1^2) \Big(\prod_{k \in B_{s_i,t_j}} \sqrt{w_k} \Big) \int_0^{\infty} \theta_{11} e^{-\theta_{12} x} \theta_{11} e^{-\theta_{12}(x-h)} I(x \geq 0) I((x-h) \geq 0) dx \\
& + \rho_1^2 \Big(\prod_{k \in B_{s_i,t_j}} \sqrt{w_k} \Big) \int_0^{\infty} \theta_{11} e^{-\theta_{12} x} \theta_{11} e^{-\theta_{12}(x-h)} I(x \geq 0) I((x-h) \geq 0) dx \\
= \; & \Big(\prod_{k \in B_{s_i,t_j}} \sqrt{w_k} \Big) \int_0^{\infty} \theta_{11}^2 e^{-\theta_{12} x} e^{-\theta_{12}(x-h)} I((x-h) \geq 0) dx \\
= \; & \Big(\prod_{k \in B_{s_i,t_j}} \sqrt{w_k} \Big) \int_h^{\infty} \theta_{11}^2 e^{-\theta_{12} x} e^{-\theta_{12}(x-h)} dx \\
= \; & \Big(\prod_{k \in B_{s_i,t_j}} \sqrt{w_k} \Big) \int_0^{\infty} \theta_{11}^2 e^{-\theta_{12}(u+h)} e^{-\theta_{12} u} du \\
= \; & \Big(\prod_{k \in B_{s_i,t_j}} \sqrt{w_k} \Big) e^{-\theta_{12} h} \frac{\theta_{11}^2}{2\theta_{12}}.
\end{aligned}$$

The three results above yield the covariance function for the first variable. The covariance

function for the second variable can be obtained similarly. That is,

$$\mathrm{cov}_{22}(h) = \begin{cases} 0, & \text{if } s_i \text{ and } t_j \text{ are not flow connected;} \\ \frac{\theta_{21}^2}{2\theta_{22}}, & \text{if } s_i \text{ and } t_j \text{ are flow connected and } d(s_i, t_j) = 0; \\ \prod_{j \in B_{s_i,t_j}} \sqrt{w_j} e^{-\theta_{22} h} \frac{\theta_{21}^2}{2\theta_{22}}, & \text{if } s_i \text{ and } t_j \text{ are flow connected and } d(s_i, t_j) = h > 0. \end{cases}$$

The crosscovariance between variables Z_1 and Z_2 at any two points s_i and t_j, respectively, can be derived as follows:

(1) By Property 3.1(3),

$$\mathrm{cov}(\int f_1^1(x|\boldsymbol{\theta}_1) W_1(x) dx, \int f_1^2(y|\boldsymbol{\theta}_1) W_0(y) dy) = 0,$$
$$\mathrm{cov}(\int f_1^1(x|\boldsymbol{\theta}_1) W_1(x) dx, \int f_1^2(y|\boldsymbol{\theta}_1) W_2(y) dy) = 0, \text{ and}$$
$$\mathrm{cov}(\int f_1^1(x|\boldsymbol{\theta}_1) W_0(x) dx, \int f_1^2(y|\boldsymbol{\theta}_1) W_2(y) dy) = 0.$$

If s_i and t_j are not flow connected, by the similar reasoning as in **(1)** for the above covariance function, we have

$$\mathrm{cov}(\int_{s_i}^{u_i} f_1^1(x_i - s_i|\boldsymbol{\theta}_1) W_0(x_i) dx_i, \int_{t_j}^{u_j} f_1^2(x_j - t_j|\boldsymbol{\theta}_1) W_0(x_j) dx_j) = 0.$$

Similarly, let $A = \{\text{segment } (s_i, u_i) \bigcup \text{segments } (l_j, u_j) \text{ for } j \in U_{s_i}\}$, let $B = \{\text{segment } (t_j, u_j) \bigcup \text{segments } (l_k, u_k) \text{ for } k \in U_{t_j}\}$, then since s_i and t_j are not flow connected and we are using tail-up construction, then $A \bigcap B = \emptyset$. Based on the above, it is easy to show that $\mathrm{cov}_{12}(h) = 0$.

(2) If s_i and t_j are exactly the same point, i.e., $h = 0$, then

$$\begin{aligned} & \mathrm{cov}_{12}(h) \\ = \; & \mathrm{cov}(Z_1(s_i), Z_2(s_i)) \\ = \; & \rho_1 \rho_2 \int_{s_i}^{u_i} \theta_{11} e^{-\theta_{12}(x_i - s_i)} \theta_{21} e^{-\theta_{22}(x_i - s_i)} I((x_i - s_i) \geq 0) dx_i \\ & + \rho_1 \rho_2 \sum_{j \in U_{s_i}} (\prod_{k \in B_{s_i,[j]}} \sqrt{w_k}) \int_{l_j}^{u_j} \theta_{11} e^{-\theta_{12}(x_j - s_i)} \theta_{21} e^{-\theta_{22}(x_j - s_i)} I((x_j - s_i) \geq 0) dx_j \\ = \; & \rho_1 \rho_2 \int_0^{\infty} \theta_{11} \theta_{21} e^{-(\theta_{12} + \theta_{22})x} dx \\ = \; & \rho_1 \rho_2 \frac{\theta_{11} \theta_{21}}{\theta_{12} + \theta_{22}}. \end{aligned}$$

(3) If s_i and t_j are flow connected and $h > 0$, then

$$\begin{aligned}
& \mathrm{cov}_{12}(h) \\
= \; & \rho_1\rho_2\Big(\prod_{k\in B_{s_i,t_j}}\sqrt{w_k}\Big)\int_0^\infty \theta_{11}e^{-\theta_{12}x}\theta_{21}e^{-\theta_{22}(x-h)}I(x\geq 0)I(x-h\geq 0)dx \\
= \; & \rho_1\rho_2\Big(\prod_{k\in B_{s_i,t_j}}\sqrt{w_k}\Big)\int_h^\infty \theta_{11}\theta_{21}e^{-\theta_{12}x}e^{-\theta_{22}(x-h)}dx \\
= \; & \Big(\prod_{k\in B_{s_i,t_j}}\sqrt{w_k}\Big)\rho_1\rho_2\int_0^\infty \theta_{11}\theta_{21}e^{-\theta_{12}(u+h)}e^{-\theta_{22}u}du \\
= \; & \prod_{j\in B_{s_i,t_j}}\sqrt{w_j}e^{-\theta_{12}h}\rho_1\rho_2\frac{\theta_{11}\theta_{21}}{\theta_{12}+\theta_{22}}.
\end{aligned}$$

We can add some further assumptions and do a reparameterization to simplify the model as follows:

$$\theta_{11} = \theta_{21},\ \theta_{12} = \theta_{22} = \frac{\theta_{11}^2}{2},\ \rho_1\rho_2 = \rho_c,\ e^{-\theta_{12}} = e^{-\theta_{22}} = \rho. \tag{3.1}$$

The model may then be simplified as

$$cov_{11}(h) = cov_{22}(h) = \prod_{j\in B_{s_i,t_j}}\sqrt{w_j}\cdot\rho^h,\ cov_{12}(h) = \prod_{j\in B_{s_i,t_j}}\sqrt{w_j}\cdot\rho_c\rho^h \tag{3.2}$$

where h is the stream distance.

The exponential covariance model for the tail-down construction can be obtained by choosing $f_2^1(x|\boldsymbol{\theta}_1) = \theta_{11}e^{\theta_{12}x}I(x\leq 0)$, $f_2^2(x|\boldsymbol{\theta}_2) = \theta_{21}e^{\theta_{22}x}I(x\leq 0)$, and then constructing variables Z_1 and Z_2 as follows:

$$\begin{aligned}
Z_1(s_i) &= \int_{l_i}^{s_i} f_2^1(x_i - s_i|\boldsymbol{\theta}_1)Y_1(x_i|\rho_1)dx_i + \sum_{j\in D_{s_i}}\int_{l_j}^{u_j} f_2^1(x_j - s_i|\boldsymbol{\theta}_1)Y_1(x_j|\rho_1)dx_j, \\
Z_2(s_i) &= \int_{l_i}^{s_i} f_2^2(x_i - s_i|\boldsymbol{\theta}_2)Y_2(x_i|\rho_2)dx_i + \sum_{j\in D_{s_i}}\int_{l_j}^{u_j} f_2^2(x_j - s_i|\boldsymbol{\theta}_2)Y_2(x_j|\rho_2)dx_j,
\end{aligned}$$

where D_{s_i} is the index set of all stream segments downstream of s_i into which s_i flows, excluding i. The covariance model can be expressed as

$$\mathrm{cov}_{11}(h) = \begin{cases} e^{-\theta_{12}h}\frac{\theta_{11}^2}{2\theta_{12}}, & \text{if } s_i \text{ and } t_j \text{ are flow connected} \\ e^{-\theta_{12}h}, & \text{if } s_i \text{ and } t_j \text{ are not flow connected,} \end{cases}$$

$$\mathrm{cov}_{12}(h) = \begin{cases} \rho_1\rho_2 e^{-\theta_{12}h}\frac{\theta_{11}^2}{2\theta_{12}}, & \text{if } s_i \text{ and } t_j \text{ are flow connected} \\ \rho_1\rho_2 e^{-\theta_{12}h}, & \text{if } s_i \text{ and } t_j \text{ are not flow connected.} \end{cases}$$

Simplifying the model according to equations (3.1), we get a covariance model identical to that for the planar case, illustrated in equation (2.8), except that this is a special case where $\sigma^2 = 1$. Note that this is a pure stream-distance model, or in other words,

$C_{ij}(Z_i(s_1), Z_j(t_2)) = C_{ij}(Z_i(t_2), Z_j(r_3))$ as long as the stream distance between s_1 and t_2 is the same as that between t_2 and r_3, even when the latter pair of points are not flow connected. Ver Hoef, Peterson, and Theobald (2005) pointed this out for the exponential moving average model, but it is not true for several other models, for example, the spherical model.

Notice that in the tail-up construction, there is weight on the second part of the right hand side of the equation for Z_1 and Z_2; but for the tail-down construction, there is no such weight. The reason is that we want the process to be stationary. In the tail-up construction, the random variable at s_i is constructed by integrating over the segment on which s_i lies and the segments in U_{s_i}. Suppose that in Figure 1.1, upstream of t_j there are no further splits, while upstream of r_k there are many splits. If we put no weighting on subbranches at all, then the total area under the moving average function will be much greater for r_k than for t_j. If Z is the random variable we construct, then $\text{cov}(Z(s_i), Z(r_k))$ will be much larger than $\text{cov}(Z(s_i), Z(t_j))$. When we use the tail-down construction, however, we do not have this "splitting" issue.

3.2 Design criteria

The same criteria that we used in Chapter 2 for optimal designs with respect to prediction, covariance parameter estimation and empirical prediction in the planar case, we now use for optimal designs on stream networks.

3.3 A small design on a stream network

3.3.1 The design problem

We first start with a relatively small design on stream network. Consider the stream network shown in Figure 3.1, where the length of each segment is one unit, and there are three candidate design points on each segment. The distance of three points on the same segment relative to the most downstream point on that segment are .01, .50, and .99. We select three points for each variable out of these 21 candidate points to optimize the

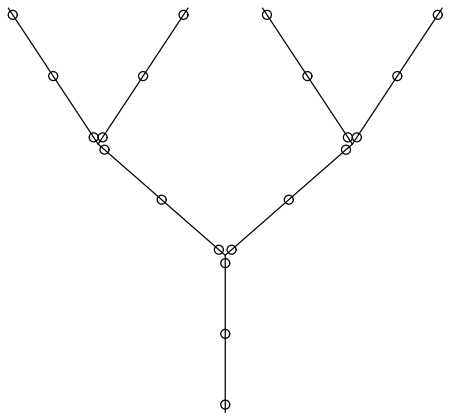

Figure 3.1: A small design problem on a stream network

corresponding criterion. Thus, in order to find the optimal design, we need to enumerate $C_3^{21} \cdot C_3^{21} = 1,768,900$ designs, not all of which are distinct, however, due to the symmetry in this stream network.

3.3.2 Optimal design for prediction

Table 3.1 contains the optimized criteria for prediction with the tail-up construction, and Figure 3.2 shows the optimal designs corresponding to each ρ and ρ_c combination. From Table 3.1, we observe that the optimized criterion decreases as ρ or ρ_c increase, while keeping the other parameter constant. The optimal design is almost invariant to changes in ρ and ρ_c. Usually, there are two design points for each variable located on one of the furthest upstream branches, the segments of which flow into the same node. There is one point for each variable located on the most downstream segment, which are collocated except for the combinations of ($\rho = 0.20$, $\rho_c = 0.50$) and ($\rho = 0.50$, $\rho_c = 0.20$). There are substantial improvements in the criterion over median criterion value of all designs, the least improvement being 38%.

ρ	0.20	0.50	0.80
$\rho_c = 0.50$	1.015(38%)	0.551(61%)	0.189(78%)
ρ_c	0.20	0.50	0.80
$\rho = 0.50$	0.680(60%)	0.551(61%)	0.291(63%)

Table 3.1: $\min\{\max_{i\in S} |\mathbf{M}(\mathbf{s}_i, \boldsymbol{\theta})|\}$ for prediction with tail-up construction

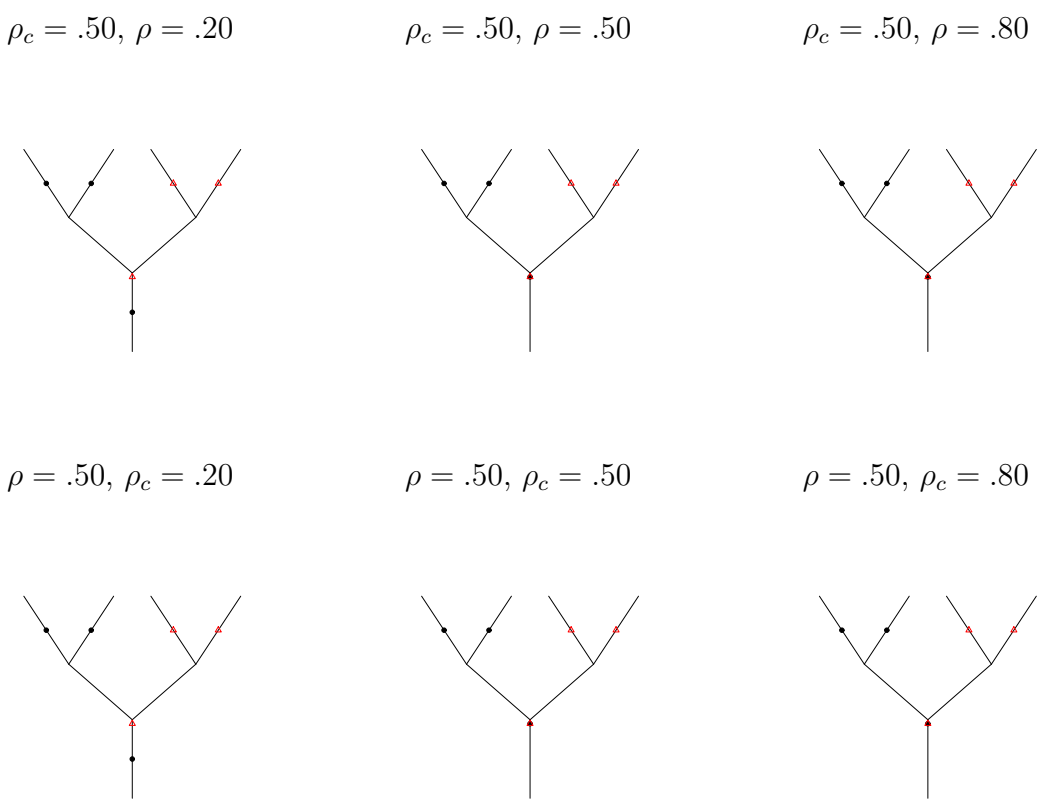

Figure 3.2: Optimal designs for prediction with tail-up construction; black solid dots are for variable 1 and red empty triangles are for variable 2

ρ	0.20	0.50	0.80
$\rho_c = 0.50$	0.916(42%)	0.380(62%)	0.058(74%)
ρ_c	0.20	0.50	0.80
$\rho = 0.50$	0.440(63%)	0.380(62%)	0.215(61%)

Table 3.2: $\min\{\max_{i \in S} |\mathbf{M}(\mathbf{s}_i, \boldsymbol{\theta})|\}$ for prediction with tail-down construction

Table 3.2 contains the optimized criteria for prediction with the tail-down construction, and Figure 3.3 shows the optimal designs corresponding to each ρ and ρ_c combination. Just as for the tail-up construction, the optimized criterion decreases as ρ or ρ_c increases, while keeping the other parameter constant. The optimal designs look similar to those for the tail-up construction except that in the tail-down construction, there are design points located at the furthest upstream location. This is no unexpected, since the tail-down construction results in a pure distance covariance model, which is the same as the planar exponential covariance model, and in Section 2.3.2 we observed that the design points were widely spread out. There are substantial improvements in the criterion over median criterion value of all designs, the least improvement being 42%.

3.3.3 Optimal design for covariance parameter estimation

Table 3.3 contains the optimized criteria for covariance estimation with the tail-up construction, and Figure 3.4 shows the optimal designs corresponding to each ρ and ρ_c combination. From Table 3.3, the optimized criterion decreases as ρ or ρ_c increases, while keeping the other parameter constant. One interesting thing is that the optimal design is invariant to the choice of ρ and ρ_c, with all the points collocated. Also, we see that the optimal design is translation invariant, by which we mean the only thing that matters in the optimal design is the relative positions of the points, not their absolute positions; if we move the points up to the most upstream segments but still keep their relative distances the same, the design obtained has the same criterion value as the optimal design. There are substantial improvements in the criterion for the optimal design, all of which are above

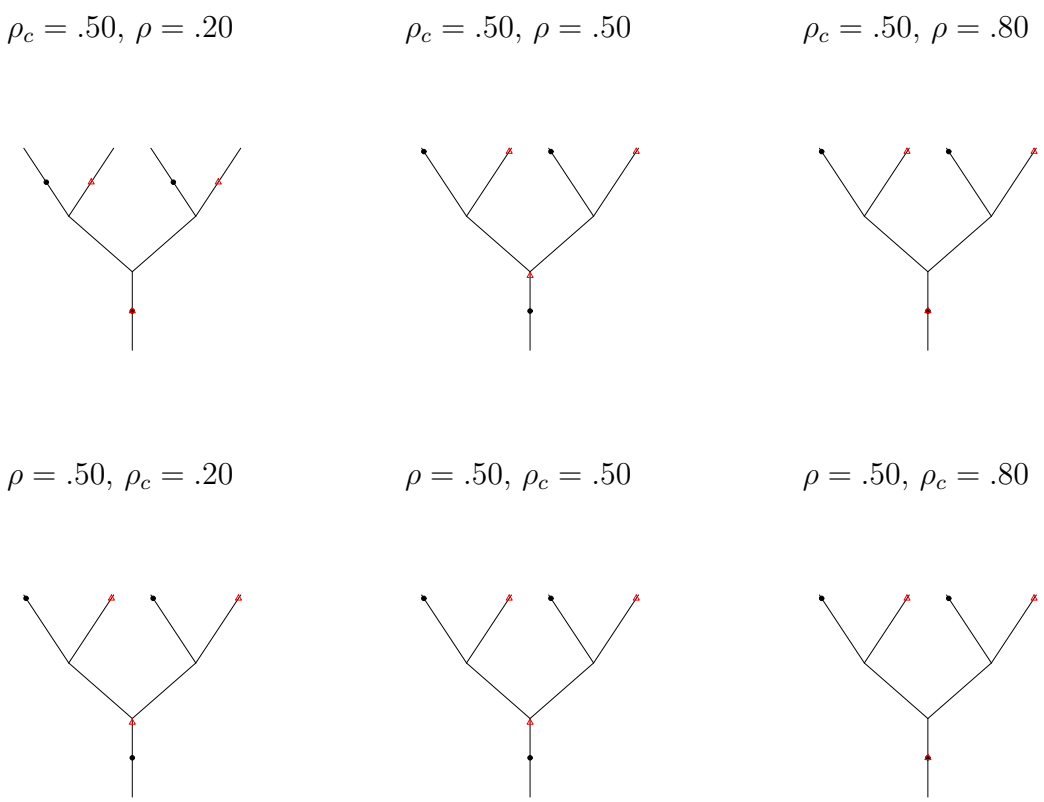

Figure 3.3: Optimal designs for prediction with tail-down construction; black solid dots are for variable 1 and red empty triangles are for variable 2

ρ	0.20	0.50	0.80
$\rho_c = 0.50$	0.0504(98%)	0.0412(96%)	0.0095(95%)
ρ_c	0.20	0.50	0.80
$\rho = 0.50$	0.0676(96%)	0.0412(96%)	0.0095(98%)

Table 3.3: $\min\{1/|\mathbf{I}_{REML}(\boldsymbol{\theta})|\}$ for covariance estimation with tail-up construction

All ρ and ρ_c combination in Table 3.3

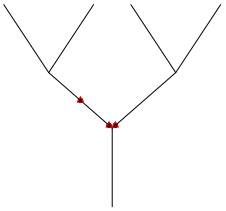

Figure 3.4: Optimal designs for covariance estimation with tail-up construction; black solid dots are for variable 1 and red empty triangles are for variable 2

95%.

Table 3.4 contains the optimized criteria for covariance estimation with the tail-down construction, and Figure 3.5 shows the optimal designs corresponding to each ρ and ρ_c combination. The first interesting thing is that the optimized criterion increases as ρ increases, but it decreases as ρ_c increases. Second, all points in the optimal designs are collocated. Third, they are almost invariant to the choice of ρ and ρ_c.

One thing worth pointing out is that we have complete collocation in the optimal designs for both the planar and stream network toy examples. Also, there are substantial improvements in the criterion over the median criterion value of all designs, all of which are above 93%.

ρ	0.20	0.50	0.80
$\rho_c = 0.50$	0.0321(96%)	0.0692(94%)	0.0747(93%)
ρ_c	0.20	0.50	0.80
$\rho = 0.50$	0.113(93%)	0.0692(94%)	0.0159(97%)

Table 3.4: $\min\{1/|\mathbf{I}_{REML}(\boldsymbol{\theta})|\}$ for covariance estimation with tail-down construction

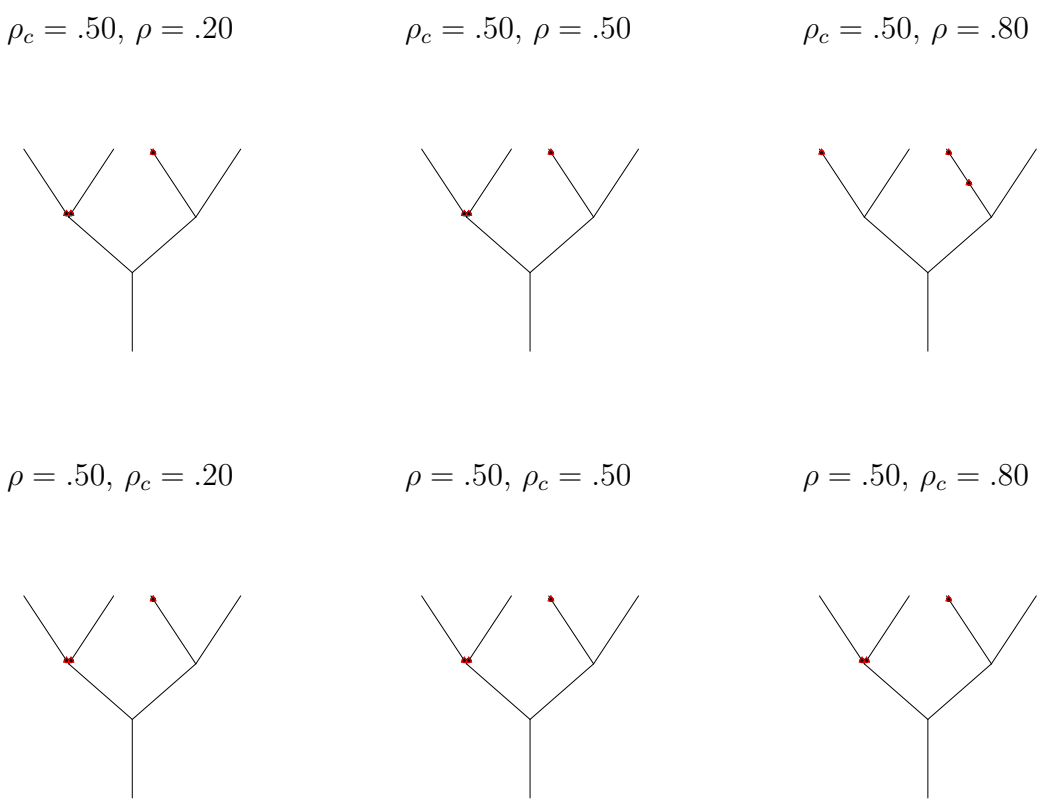

Figure 3.5: Optimal designs for covariance estimation with tail-down construction; black solid dots are for variable 1 and red empty triangles are for variable 2

ρ	0.20	0.50	0.80
$\rho_c = 0.50$	4.700(86%)	3.891(80%)	1.701(83%)
ρ_c	0.20	0.50	0.80
$\rho = 0.50$	4.751(80%)	3.891(80%)	1.648(85%)

Table 3.5: $\min\{\max_{i \in S} |\mathbf{M}(\mathbf{s}_i, \hat{\boldsymbol{\theta}})|\}$ for empirical prediction with tail-up construction

3.3.4 Optimal design for empirical prediction

Table 3.5 contains the optimized criteria for empirical prediction with the tail-up construction, and Figure 3.6 shows the optimal designs corresponding to each ρ and ρ_c combination. From Table 3.5, the optimized criterion decreases as ρ or ρ_c increase while keeping the other parameter constant. The optimal designs, except for the case when $\rho = 0.20$ and $\rho_c = 0.50$, have two points collocated at one of the nodes. That the optimized criteria and the degree of collocation in the optimal designs for empirical prediction sit between those for prediction and covariance estimation matches the fact that the empirical prediction criterion is a combination of those for prediction and covariance parameter estimation. Table 3.5 shows great improvement in the criterion over median criterion value of all designs, by an amount which is between that for prediction and covariance parameter estimation.

Table 3.6 contains the optimized criteria for empirical prediction with the tail-down construction, and Figure 3.7 shows the optimal designs corresponding to each ρ and ρ_c combination. From Table 3.6, the optimized criterion increases as ρ increases and decreases as ρ_c decreases, which actually is the same trend observed in the tail-down covariance estimation case. The optimal design is either completely collocated or has two points collocated. Again, this is consistent with the fact that the empirical prediction criterion is a combination of those for prediction and covariance parameter estimation. Table 3.6 shows great improvements in the criterion over median criterion value of all designs, by an amount which is between that for prediction and covariance parameter estimation.

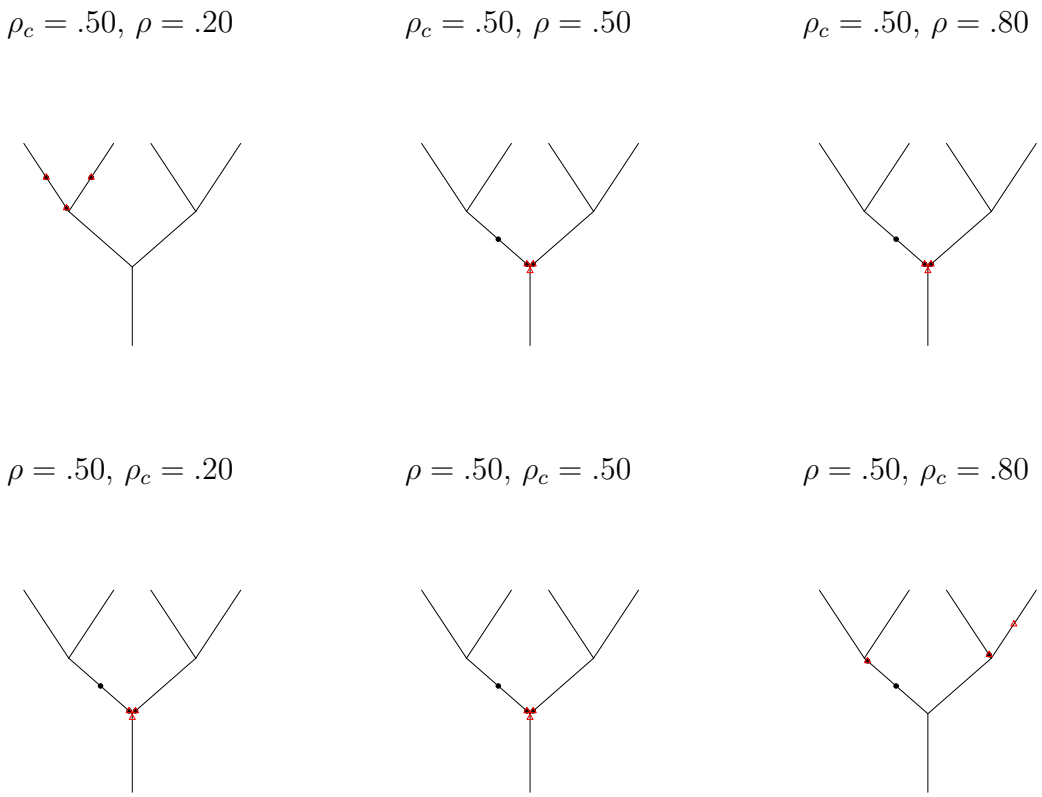

Figure 3.6: Optimal designs for empirical prediction with tail-up construction; black solid dots are for variable 1 and red empty triangles are for variable 2

ρ	0.20	0.50	0.80
$\rho_c = 0.50$	5.104(80%)	8.569(86%)	118.59(90%)
ρ_c	0.20	0.50	0.80
$\rho = 0.50$	10.904(87%)	8.569(86%)	4.124(87%)

Table 3.6: $\min\{\max_{i\in S} |\mathbf{M}(\mathbf{s}_i, \hat{\boldsymbol{\theta}})|\}$ for empirical prediction with tail-down construction

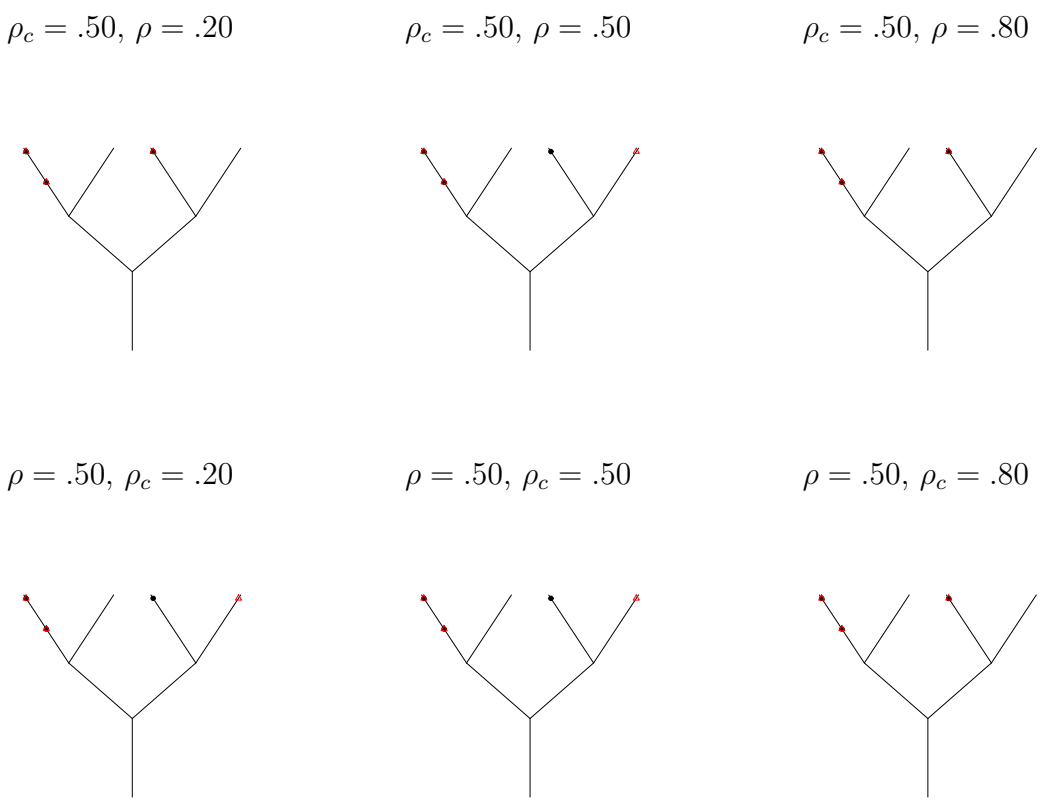

Figure 3.7: Optimal designs for empirical prediction with tail-down construction; black solid dots are for variable 1 and red empty triangles are for variable 2

3.4 Large example using simulated annealing algorithm

3.4.1 The design problem

As another, much larger example, consider a stream network which has the same configuration as in Figure 3.1, but assume that we have 10, rather than 3, candidate design points on each segment. The distances of those points relative to the most downstream node on that segment are 0.05, 0.15, ... 0.95. Hence, we have 70 points in total, and we choose 5 points for each variable. It is not feasible computationally to enumerate all the possible designs ($C_5^{70} \cdot C_5^{70}$ designs) to get the optimal one, even though not all of them are distinct due to symmetry. Instead, we use a simulated annealing algorithm as described in Chapter 2. Compared to the planar case, the only difference in using SAA within the stream network context is that we can only move the point along the stream network. Or in other words, a point on an arbitrary segment can only be moved toward its parent, child or sibling segment if it has one. The following results show that the change in criterion and the designs produced by SAA basically share the same characteristics as their counterparts in the toy example. Again, The results obtained by SAA can be affected by the initial design S_0, hence, different start-up designs (5 for each case) were chosen randomly and the results presented below are the best result we obtained. Although different start-up designs will lead to different final designs, the results are robust in the sense that the criterion values obtained and the spatial configuration of the design points do not differ much.

3.4.2 Designs produced by SAA for prediction

Table 3.7 contains the criterion values produced by SAA for prediction with the tail-up construction, and Figure 3.8 shows the designs produced by SAA corresponding to each ρ and ρ_c combination. Just as for the toy example, the criterion decreases as ρ and ρ_c increases. Using SAA, we still see substantial improvements in the criterion compared to randomly chosen designs, with the least one being 42%.

ρ	0.20	0.50	0.80
$\rho_c = 0.50$	0.696(42%)	0.371(67%)	0.135(83%)
ρ_c	0.20	0.50	0.80
$\rho = 0.50$	0.459(67%)	0.371(67%)	0.178(69%)

Table 3.7: $\min\{\max_{i \in S} |\mathbf{M}(\mathbf{s}_i, \boldsymbol{\theta})|\}$ for prediction with tail-up construction using SAA

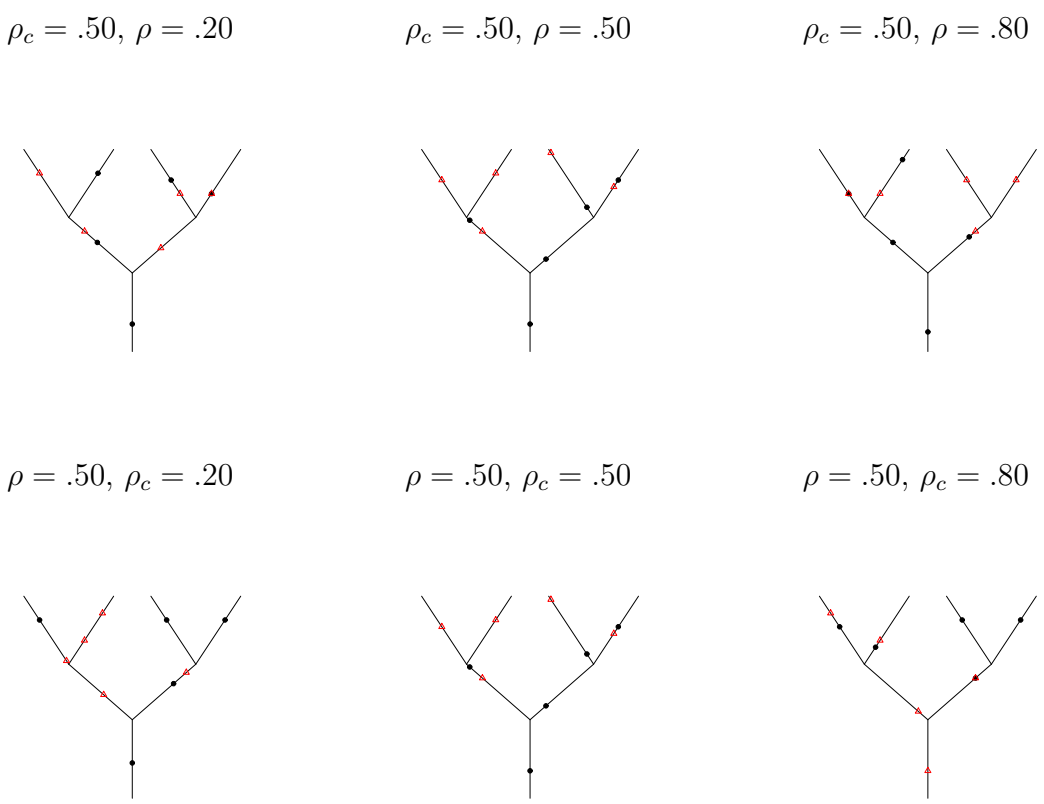

Figure 3.8: Designs for prediction with tail-up construction using SAA; black solid dots are for variable 1 and red empty triangles are for variable 2

ρ	0.20	0.50	0.80
$\rho_c = 0.50$	0.475(59%)	0.146(80%)	0.0209(87%)
ρ_c	0.20	0.50	0.80
$\rho = 0.50$	0.195(79%)	0.146(80%)	0.0800(80%)

Table 3.8: $\min\{\max_{i \in S} |\mathbf{M}(\mathbf{s}_i, \boldsymbol{\theta})|\}$ for prediction with tail-down construction using SAA

Table 3.8 contains the criterion values produced by SAA for prediction with the tail-down construction, and Figure 3.9 shows the designs produced by SAA corresponding to each ρ and ρ_c combination. As we observed in the toy example, the criterion decreases as ρ and ρ_c increases. Compared with the tail-up construction, we see more symmetry in the designs for tail-down construction. Using SAA, we still see substantial improvements in the criterion compared to randomly chosen designs, with the least one being 59%.

3.4.3 Designs produced by SAA for covariance parameter estimation

Table 3.9 contains the criterion values produced by SAA for covariance estimation with the tail-up construction, and Figure 3.10 shows the designs produced by SAA corresponding to each ρ and ρ_c combination. Just as for the toy example, the criterion decreases as ρ_c or ρ_c increases. We see that at least four of the five design points are collocated in the designs produced by SAA. There are substantial improvements in the criterion compared to random designs. The improvements are smaller compared to those in the toy example, which is because we cannot enumerate all the designs by using SAA.

Table 3.10 contains the criterion values produced by SAA for covariance estimation with the tail-down construction, and Figure 3.11 shows the designs produced by SAA corresponding to each ρ and ρ_c combination. The criterion increases as ρ increases and decreases as ρ_c increases. We have at least four collocated points in the designs produced by SAA. Just as for the tail-up construction, there are substantial improvements in the criterion relative to random designs, which are smaller compared to those in the toy example.

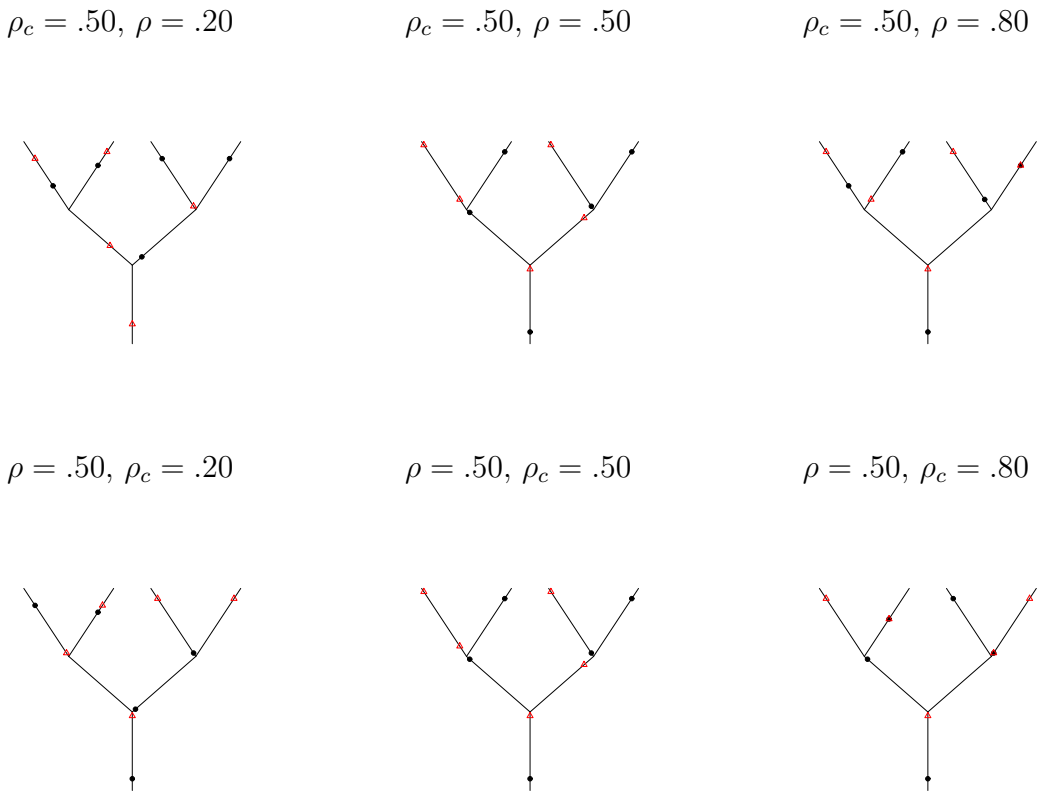

Figure 3.9: Designs for prediction with tail-down construction using SAA; black solid dots are for variable 1 and red empty triangles are for variable 2

ρ	0.20	0.50	0.80
$\rho_c = 0.50$	0.00494(88%)	0.00433(88%)	0.00113(82%)
ρ_c	0.20	0.50	0.80
$\rho = 0.50$	0.00939(81%)	0.00433(88%)	0.00157(89%)

Table 3.9: $\min\{1/|\mathbf{I}_{REML}(\boldsymbol{\theta})|\}$ for covariance estimation with tail-up construction using SAA

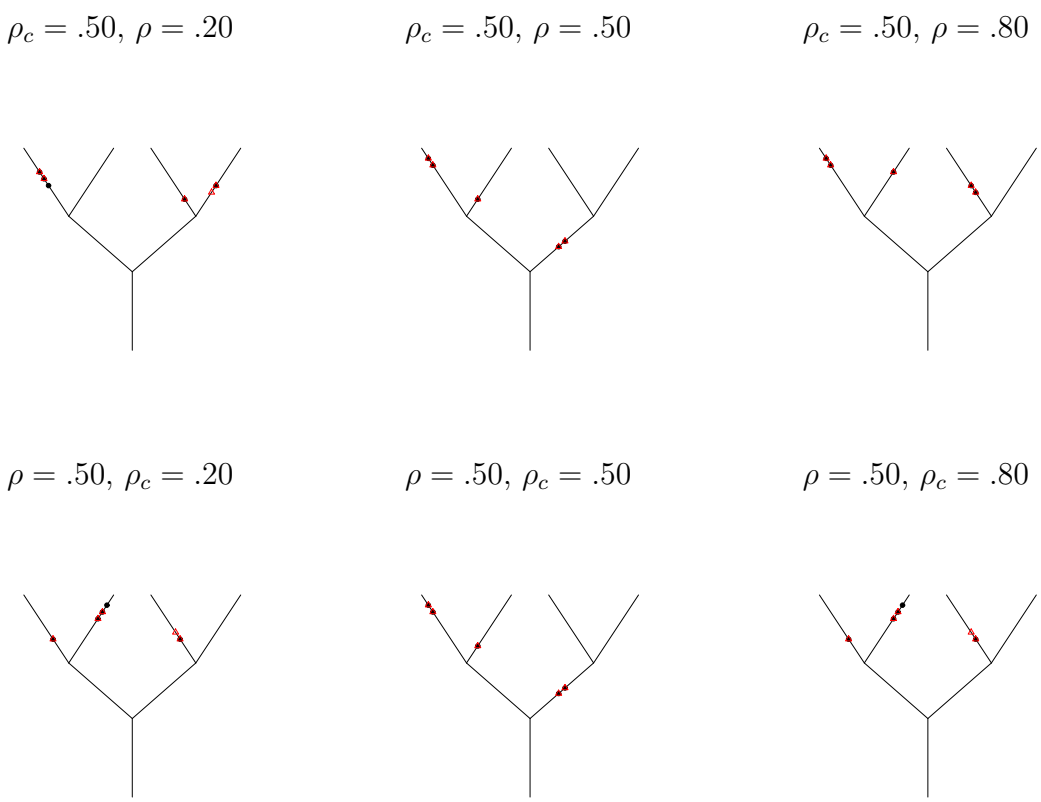

Figure 3.10: Designs for covariance estimation with tail-up construction using SAA; black solid dots are for variable 1 and red empty triangles are for variable 2

ρ	0.20	0.50	0.80
$\rho_c = 0.50$	0.00595(86%)	0.00908(87%)	0.0149(80%)
ρ_c	0.20	0.50	0.80
$\rho = 0.50$	0.0176(81%)	0.00908(87%)	0.00233(92%)

Table 3.10: $\min\{1/|\mathbf{I}_{REML}(\boldsymbol{\theta})|\}$ for covariance estimation with tail-down construction using SAA

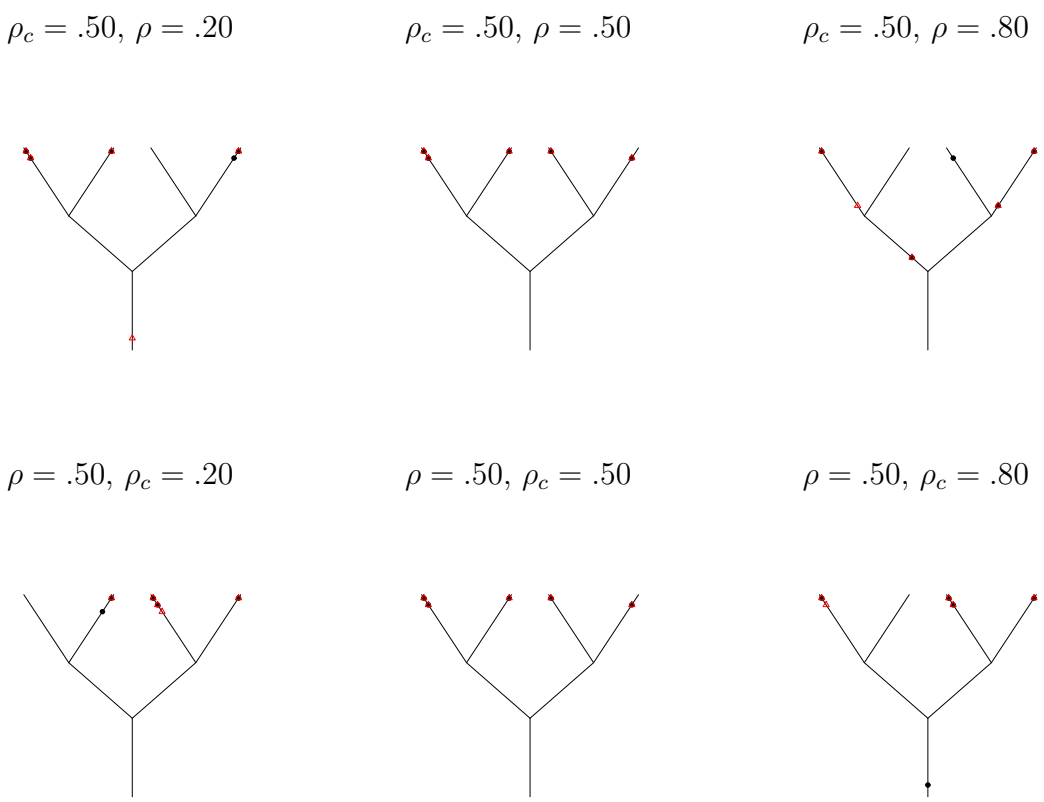

Figure 3.11: Designs for covariance estimation with tail-down construction using SAA; black solid dots are for variable 1 and red empty triangles are for variable 2

Combined with the results we got from the toy example, we can see that a high degree of collocation is the most substantial characteristic in the designs produced by SAA for covariance parameter estimation on the stream network we used in this case study.

ρ	0.20	0.50	0.80
$\rho_c = 0.50$	1.753(44%)	1.255(57%)	0.582(73%)
ρ_c	0.20	0.50	0.80
$\rho = 0.50$	1.640(55%)	1.255(57%)	0.518(65%)

Table 3.11: $\min\{\max_{i \in S} |\mathbf{M}(\mathbf{s}_i, \hat{\boldsymbol{\theta}})|\}$ for empirical prediction with tail-up construction using SAA

3.4.4 Designs produced by SAA for empirical prediction

Table 3.11 contains the criterion values produced by SAA for empirical prediction with the tail-up construction, and Figure 3.12 shows the designs produced by SAA corresponding to each ρ and ρ_c combination. The criterion shows the same trends as we observed from the toy example, that is, it decreases as ρ and ρ_c increase. The designs have a much larger degree of collocation than those for prediction, and the design points are more spread out than those in the designs for covariance parameter estimation. The improvements in the criterion relative to random designs are all above 44%, which sits between those for prediction and covariance parameter estimation.

Table 3.12 contains the criterion values produced by SAA for empirical prediction with the tail-down construction, and Figure 3.13 shows the designs produced by SAA corresponding to each ρ and ρ_c combination. The criterion increases as ρ increases and decreases as ρ_c increases, where we observe the same trends in the corresponding toy example. The designs have a much larger degree of collocation than those for prediction, and the design points are more spread out than those in the designs for covariance parameter estimation. The improvements in the criterion are all above 58%, which sits between those for prediction and covariance parameter estimation.

Combined with the results we have for the toy example, we see that the designs produced by SAA for empirical prediction combine characteristics of those for prediction and covariance estimation.

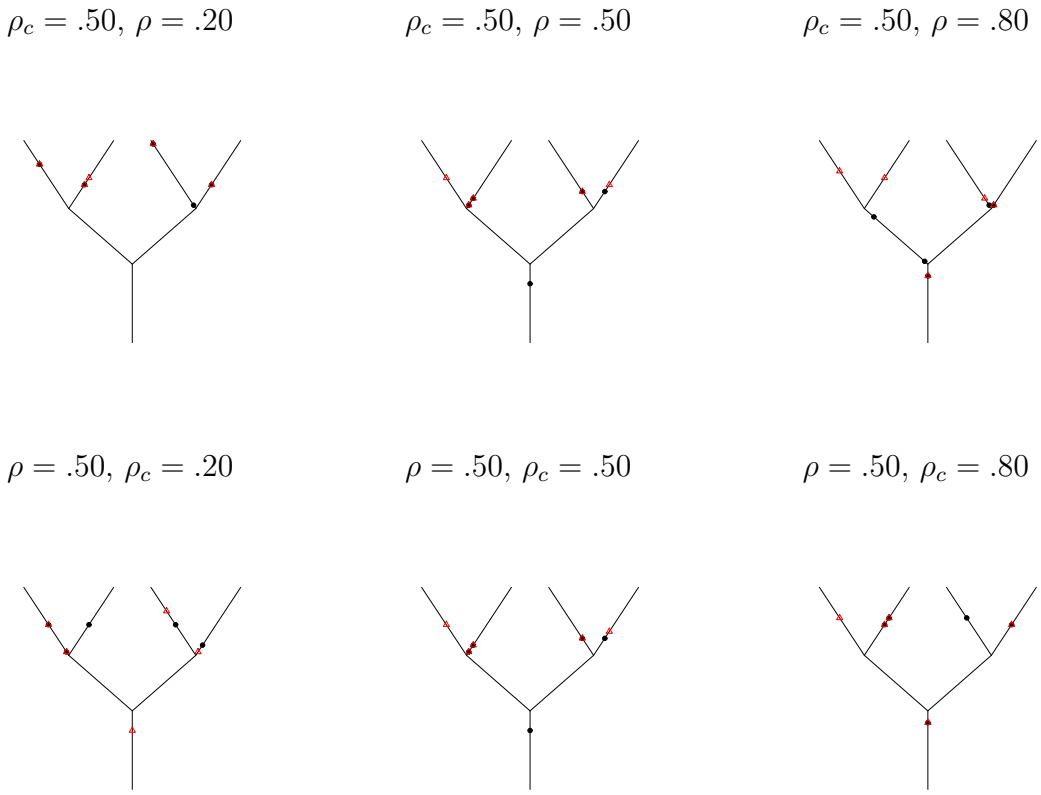

Figure 3.12: Designs for empirical prediction with tail-up construction using SAA; black solid dots are for variable 1 and red empty triangles are for variable 2

ρ	0.20	0.50	0.80
$\rho_c = 0.50$	1.623(58%)	2.184(78%)	49.346(76%)
ρ_c	0.20	0.50	0.80
$\rho = 0.50$	2.455(80%)	2.184(78%)	0.847(84%)

Table 3.12: $\min\{\max_{i \in S} |\mathbf{M}(\mathbf{s}_i, \hat{\boldsymbol{\theta}})|\}$ for empirical prediction with tail-down construction using SAA

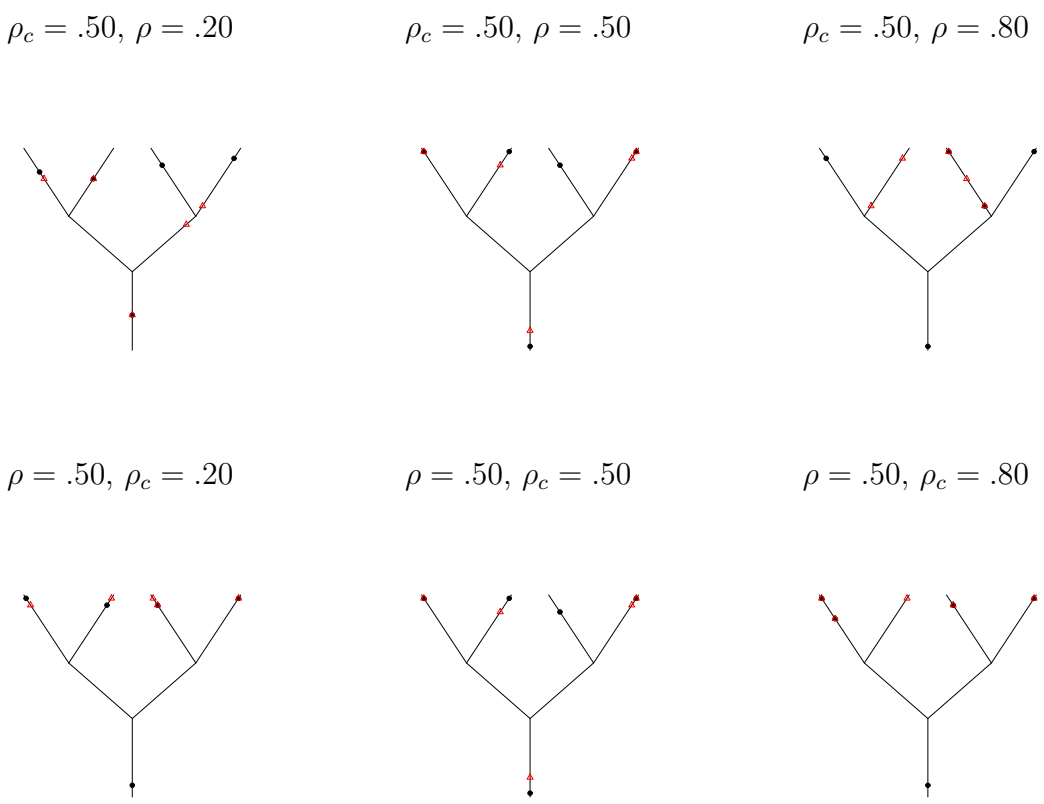

Figure 3.13: Designs for empirical prediction with tail-down construction using SAA; black solid dots are for variable 1 and red empty triangles are for variable 2

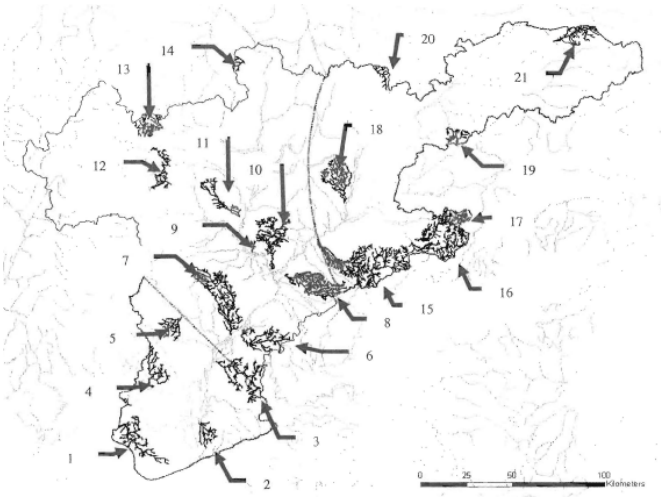

Figure 3.14: Nushagak drainage survey areas

3.5 An application

Now we will apply our methodology to a real stream network. Figure 3.14 shows the local stream networks selected for intensive sampling by the Alaska Department of Fish and Game in the Nushagak drainage survey areas in Alaska. Figure 3.14 shows that there were 21 different local stream networks of various sizes selected in this study. We will focus on developing optimal multivariate designs on the stream network shown in the left panel of Figure 3.15. Since for our optimal design criteria, it is the relative distances between the design points, not their absolute positions, that affect the corresponding value of the criterion function, we can further simplify the stream network of interest to the one that is displayed in the right panel of Figure 3.15, which looks the same as the stream network we considered in the simulation studies—a standard binary tree. The results may be extended easily to more complicated stream networks.

There are two study objectives in this application. The first one is to predict the occurrence of two fish species at unsampled points; the second one is to predict the occurrence of a fish species and the concentration of a certain water pollutant. Here, prediction refers to empirical prediction. The study has some requirements imposed by the Alaska Department of Fish and Game. First, it requires that there must be one design point near

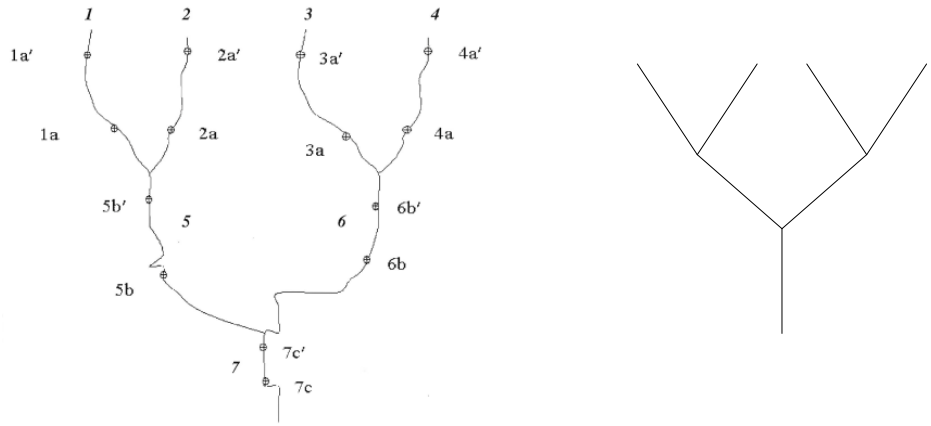

Figure 3.15: Local stream network (left) and its simplified version (right)

the middle of each stream segment. Thus, seven points of the design are already determined. To this existing seven point design, we need to add two new points, of which at least one is located upstream of all seven existing design points. Last, the study requires the design to be completely collocated because this is a very wild and difficult-to-access area in Alaska.

We assume there are 20 candidate points on each segment of the stream network shown in the right panel of Figure 3.15, so there are 140 candidate points in total. For the first study objective, we use a tail-down construction to model the occurrence of two fish species. Figure 3.16 shows the optimal design for this case. The solid dots represent the seven existing design points, while the open triangles represent the two additional design points. We tried different combination of ρ and ρ_c, but it turned out the optimal locations of the two additional design points were invariant to the choice of ρ and ρ_c. The optimal design results from placing the two additional points in two furthermost upstream locations on distinct segments of different branches.

For the second study objective, we use a tail-down construction for modelling the occurrence of the fish species (variable 1) and a tail-up construction for the water pollutant (variable 2). Let $f_1(x) = \theta_{21}e^{\theta_{22}x}I(x \leq 0)$ and $f_2(x) = \theta_{11}e^{-\theta_{12}x}I(x \geq 0)$, where $\theta_{ij} > 0$ for $i, j = 1, 2$. Let $Y_k(\mathbf{x}|\gamma_k) = \sqrt{1-\gamma_k^2}W_k(\mathbf{x}) + \gamma_k W_0(\mathbf{x})$, where $-1 \leq \gamma_k \leq 1$, $k = 1, 2$, W_l

$$\rho_c = 0.50,\ \rho = 0.20, 0.50, 0.80;\ \rho = 0.50,\ \rho_c = 0.20, 0.50, 0.80$$

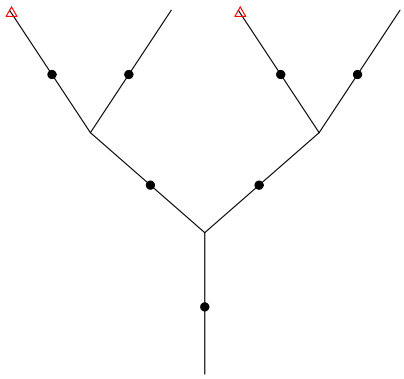

Figure 3.16: Optimal design for empirically predict the occurrence of two fish species; black solid dots are for variable 1 and red empty triangles are for variable 2

is a zero mean while noise process, $l = 0, 1, 2$. Construct the two variables Z_1 and Z_2 as follows:

$$Z_1(s_i) = \int_{l_i}^{s_i} f_2(x_i - s_i|\boldsymbol{\theta}_2) Y_2(x_i|\gamma_2) dx_i + \sum_{j \in D_{s_i}} \int_{l_j}^{u_j} f_2(x_j - s_i|\boldsymbol{\theta}_2) Y_2(x_j|\gamma_2) dx_j,$$

$$Z_2(s_i) = \int_{s_i}^{u_i} f_1(x_i - s_i|\boldsymbol{\theta}_1) Y_1(x_i|\gamma_1) dx_i + \sum_{j \in U_{s_i}} \Big(\prod_{k \in B_{s_i,[j]}} \sqrt{w_k} \Big) \int_{l_j}^{u_j} f_1(x_j - s_i|\boldsymbol{\theta}_1) Y_1(x_j|\gamma_1) dx_j.$$

Then it follows that the covariance and cross-covariance functions are given by

$$\mathrm{cov}_{11}(h) = \begin{cases} e^{-\theta_{12} h}, & \text{if } s_i \text{ and } t_j \text{ are not flow connected} \\ e^{-\theta_{12} h} \frac{\theta_{11}^2}{2\theta_{12}}, & \text{if } s_i \text{ and } t_j \text{ are flow connected,} \end{cases}$$

$$\mathrm{cov}_{22}(h) = \begin{cases} 0, & \text{if } s_i \text{ and } t_j \text{ are not flow connected} \\ \frac{\theta_{21}^2}{2\theta_{22}}, & \text{if } s_i \text{ and } t_j \text{ are flow connected and } d(s_i, t_j) = 0 \\ \prod_{j \in B_{s_i,t_j}} \sqrt{w_j} e^{-\theta_{22} h} \frac{\theta_{21}^2}{2\theta_{22}}, & \text{if } s_i \text{ and } t_j \text{ are flow connected and } d(s_i, t_j) = h > 0, \end{cases}$$

$$\mathrm{cov}_{12}(h) = \gamma_1 \gamma_2 \frac{\theta_{11}\theta_{21}}{\theta_{22} - \theta_{12}} e^{-\theta_{12} h} [1 - e^{-(\theta_{22} - \theta_{12}) h}].$$

In order to make $\mathrm{cov}_{12}(h)$ well-defined, we have to assume that $\theta_{12} \neq \theta_{22}$ under this construction. We can reparameterize the model as follows:

$$\frac{\theta_{11}^2}{2\theta_{12}} = \sigma_1^2,\ e^{-\theta_{12}} = \rho_1,\ \frac{\theta_{21}^2}{2\theta_{22}} = \sigma_2^2,\ e^{-\theta_{22}} = \rho_2,\ \gamma_1\gamma_2 \frac{\theta_{11}\theta_{21}}{\theta_{22} - \theta_{12}} = \rho_c.$$

$\rho_1 = 0.2,\ \rho_2 = 0.8,\ \rho_c = 0.5$ $\rho_1 = 0.8,\ \rho_2 = 0.2,\ \rho_c = 0.5$

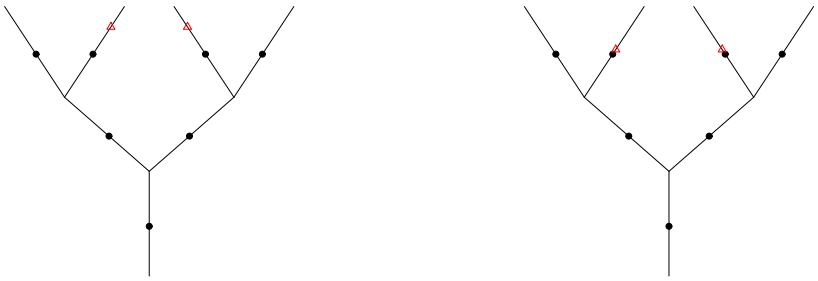

Figure 3.17: Optimal design for empirical prediction of the occurrence of a fish species and the concentration of a water pollutant; black solid dots are for variable 1 and red empty triangles are for variable 2

Then it follows that

$$\mathrm{cov}_{11}(h) = \begin{cases} \rho_1^h, & \text{if } s_i \text{ and } t_j \text{ are not flow connected} \\ \sigma_1^2 \rho_1^h, & \text{if } s_i \text{ and } t_j \text{ are flow connected,} \end{cases}$$

$$\mathrm{cov}_{22}(h) = \begin{cases} 0, & \text{if } s_i \text{ and } t_j \text{ are not flow connected} \\ \sigma_2^2, & \text{if } s_i \text{ and } t_j \text{ are flow connected and } d(s_i, t_j) = 0 \\ \sigma_2^2 \prod_{j \in B_{s_i,t_j}} \sqrt{w_j} \rho_2^h, & \text{if } s_i \text{ and } t_j \text{ are flow connected and } d(s_i, t_j) = h > 0, \end{cases}$$

$$\mathrm{cov}_{12}(h) = \rho_c[\rho_1^h - \rho_2^h].$$

Here, ρ_1 and ρ_2 are the spatial correlation factors for Z_1 and Z_2, respectively. Now $\theta_{12} \neq \theta_{22}$ is equivalent to $\rho_1 \neq \rho_2$. Thus, we must assume the spatial correlation factors for Z_1 and Z_2 are different under this construction.

Hence, we chose different spatial correlation factors for fish species and water pollutant. Optimal designs for two cases of (ρ_1, ρ_2) are shown in Figure 3.17. When the spatial correlation among occurrences the fish species is weaker than that of the water pollutant, we should locate the two additional design points at locations that are relatively further upstream compared to the results when the spatial correlation of the water pollutant is weaker.

ρ_c	0.20	0.50	0.80
$\rho_1 = 0.2, \rho_2 = 0.8$	0.549	0.517	0.458
ρ_c	0.20	0.50	0.80
$\rho_1 = 0.8, \rho_2 = 0.2$	0.293	0.292	0.248

Table 3.13: $\min\{\max_{i \in S} |\mathbf{M}(\mathbf{s}_i, \boldsymbol{\theta})|\}$ for prediction with one-tail-up-one-tail-down construction

3.5.1 Toy examples using one tail-up and one tail-down construction

In the above application, we showed how to construct a one-tail-up-one-tail-down covariance model and use empirical prediction as the criterion in this specific application. Here we will further investigate this model by carrying out simulations using all the three criteria on a toy example which is the same as the one used in Chapter 3.3.

For variable 1, we use a tail-down construction, and for variable 2, we use a tail-up construction. As mentioned in the above construction, we need to assume that the spatial correlation factors for the two variables under the two constructions are different. We consider six combination of (ρ_1, ρ_2, ρ_c), three of which are $(\rho_1, \rho_2) = (0.2, 0.8)$ and $\rho_c = 0.2, 0.5, 0.8$; the other three being $(\rho_1, \rho_2) = (0.8, 0.2)$ and $\rho_c = 0.2, 0.5, 0.8$.

3.5.1.1 Optimal design for prediction

Table 3.13 contains the optimized criteria and Figure 3.18 contains the optimal designs. The dots are points for variable 1 (tail-down construction) and the triangles are points for variable 2 (tail-up construction). Again, the optimized criterion decreases when ρ_c increases. There are no collocated points in the optimal design. The points for the tail-up construction are all widely spread out. For the variable using tail-down construction, there are two points clustered at the node for the combination of $\rho_1 = 0.2, \rho_2 = 0.8, \rho_c = 0.2, 0.5, 0.8$ and $\rho_1 = 0.8, \rho_2 = 0.2, \rho_c = 0.8$.

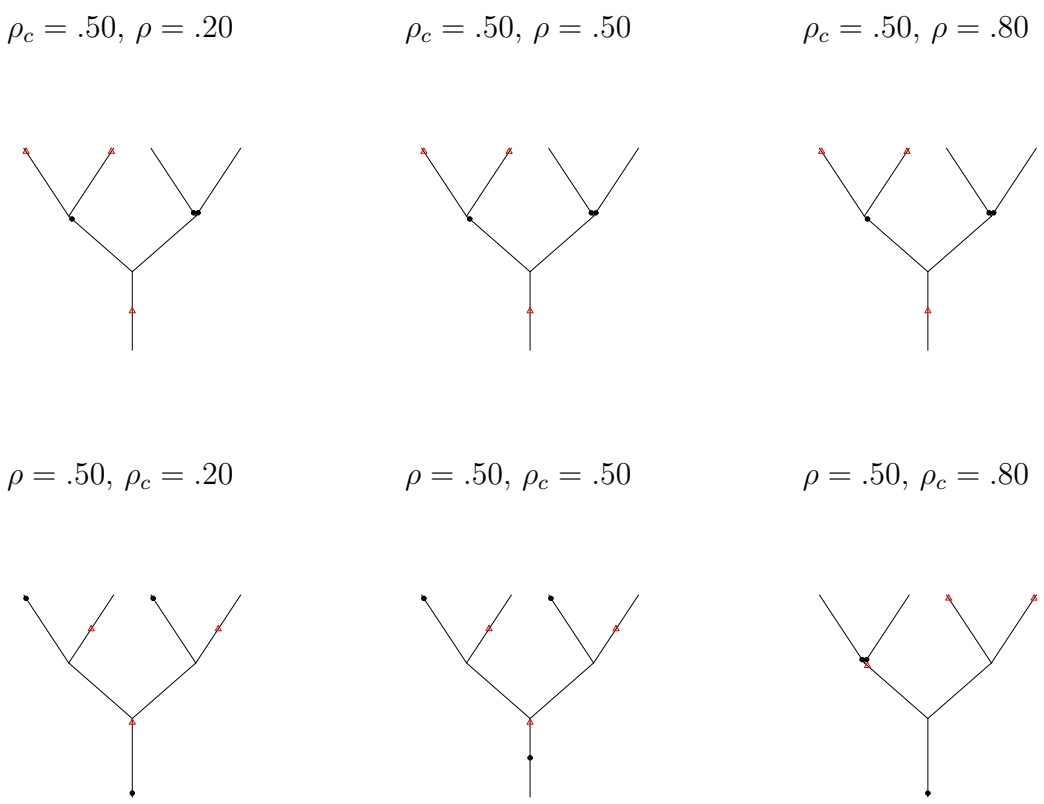

Figure 3.18: Optimal designs for prediction with one-tail-up-one-tail-down construction; black solid dots are for variable 1 and red empty triangles are for variable 2

ρ_c	0.20	0.50	0.80
$\rho_1 = 0.2, \rho_2 = 0.8$	0.014	0.005	0.00006
ρ_c	0.20	0.50	0.80
$\rho_1 = 0.8, \rho_2 = 0.2$	0.013	0.004	0.00003

Table 3.14: $\min\{1/|\mathbf{I}_{REML}(\boldsymbol{\theta})|\}$ for covariance parameter estimation with one-tail-up-one-tail-down construction

ρ_c	0.20	0.50	0.80
$\rho_1 = 0.2, \rho_2 = 0.8$	9.705	9.089	6.414
ρ_c	0.20	0.50	0.80
$\rho_1 = 0.8, \rho_2 = 0.2$	15.695	12.979	5.013

Table 3.15: $\min\{\max_{i \in S} |\mathbf{M}(\mathbf{s}_i, \hat{\boldsymbol{\theta}})|\}$ for empirical prediction with one-tail-up-one-tail-down construction

3.5.1.2 Optimal design for covariance parameter estimator

Table 3.14 contains the optimized criteria and Figure 3.19 contains the optimal designs. The optimized criterion decreases as ρ_c increases. Except for the combination of $\rho_1 = 0.8, \rho_2 = 0.2, \rho_c = 0.2$, there are at least two points collocated in the optimal design. Also, points are clustered on a single segment or around a node.

3.5.1.3 Optimal design for empirical prediction

Table 3.15 contains the optimized criteria and Figure 3.20 contains the optimal designs. Again, there is a decreasing trend in the optimized criterion when ρ_c increases. All of the optimal designs have two points collocated. These two collocated points lie on the same segment, while the two uncollocated points lie on different segments.

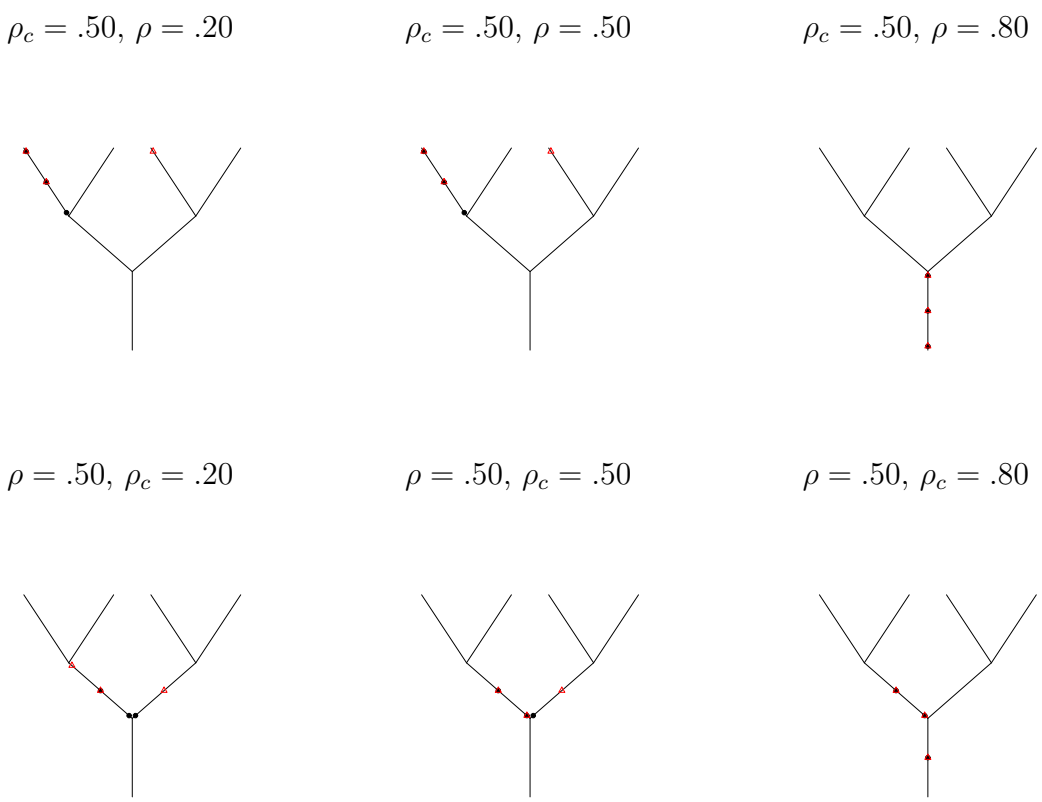

Figure 3.19: Optimal designs for covariance parameter estimation with one-tail-up-one-tail-down construction; black solid dots are for variable 1 and red empty triangles are for variable 2

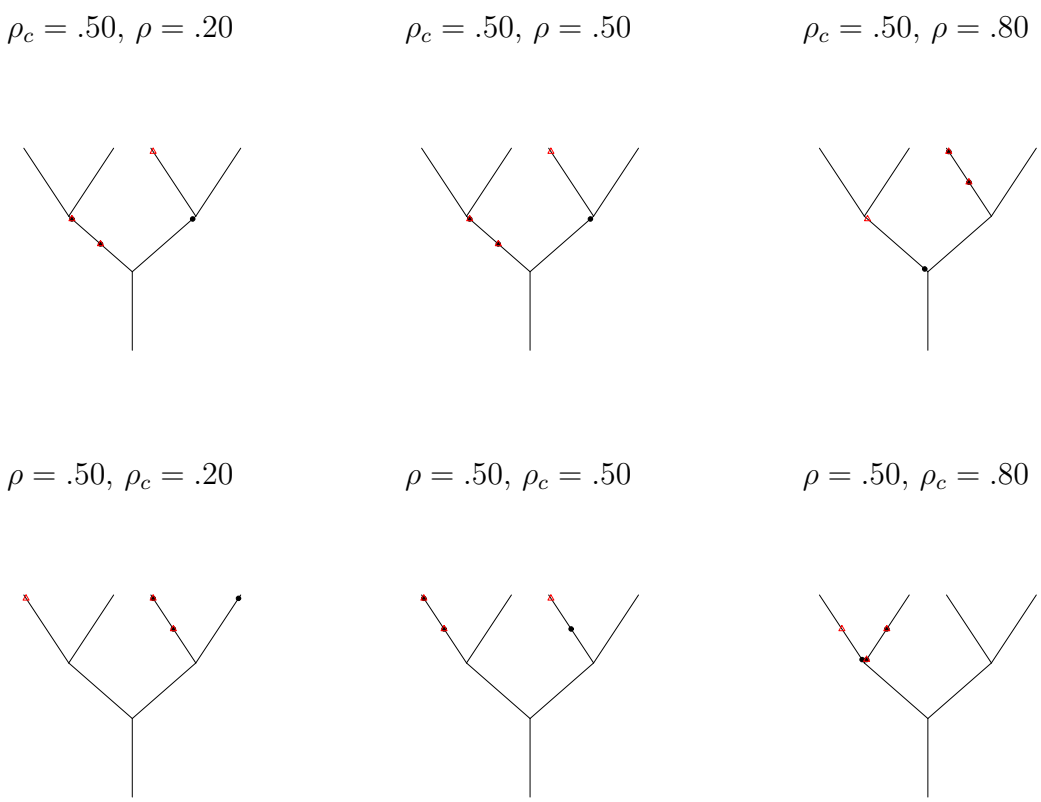

Figure 3.20: Optimal designs for empirical prediction with one-tail-up-one-tail-down construction; black solid dots are for variable 1 and red empty triangles are for variable 2

CHAPTER 4
OPTIMAL DESIGN WITH SHIFT PARAMETER

In the design problems considered so far, in both the plane and stream network, we made an assumption that the covariance function is symmetric as shown in (2.7). However, this is somewhat restrictive and difficult to verify in practice (Ver Hoef and Cressie 1993). For example, in New Zealand, while zinc and copper have both been detected in sediments of urban rivers and estuaries, the deposition of heavier metals will tend to occur further upstream. Hence the covariance between zinc and copper where zinc is in the upstream position relative to copper is different from the covariance between them where zinc is in the downstream position relative to copper (Sain and Cressie 2007). To broaden the scope of our investigation to allow for such asymmetries, we add a spatial shift parameter Δ to the covariance model. For the planar case, the covariance model becomes

$$\mathrm{cov}_{ij}(Z_i(\mathbf{s}_1), Z_j(\mathbf{s}_2)) = \begin{cases} \sigma^2\rho^{\|\mathbf{s}_1-\mathbf{s}_2\|}, & \text{if } i = j; \\ \sigma^2\rho_c\rho^{|x_{\mathbf{s}_1}-x_{\mathbf{s}_2}+\Delta|+|y_{\mathbf{s}_1}-y_{\mathbf{s}_2}+\Delta|}, & \text{if } i \neq j. \end{cases} \tag{4.1}$$

For the stream network, this objective can be achieved by choosing shift parameters $\boldsymbol{\Delta}_k$ and letting $Y_k(\mathbf{x}|\rho_k) = \sqrt{1-\rho_1^2}W_k(\mathbf{x}) + \rho_1 W_0(\mathbf{x}-\boldsymbol{\Delta}_k)$. Adding a shift parameter definitely affects the spatial configuration of the design; in particular, it affects the relative position of the design points for variable 1 to variable 2. We illustrate the effect of this shift parameter by using planar prediction as an example in the following.

Consider the design problem in the plane using a simulated annealing algorithm, and choosing $\Delta = 0.5$, as described above. Consider how the optimal design with respect to prediction changes as ρ_c increases. Table 4.1 contains the optimized criterion and Figure 4.1 shows the optimal designs. Compared to the case where $\Delta = 0$ (see Table 2.4), after adding Δ the optimized criteria are slightly larger and the degree of collocation decreases. Especially when $\rho_c = 0.8$, we can see an obvious pairing pattern in Figure 4.1, whereas we saw more collocated points in Figure 2.5. We also observed a decrease in degree of collocation in the optimal design for prediction when $\Delta = 2$, $\rho = 0.5$ and $\rho_c = 0.2, 0.5, 0.8$

ρ_c	0.20	0.50	0.80
$\rho = 0.50$	1.069	0.857	0.454

Table 4.1: $\min\{\max_{i \in S} |\mathbf{M}(\mathbf{s}_i)|\}$ for prediction using SAA with $\Delta = 0.5$

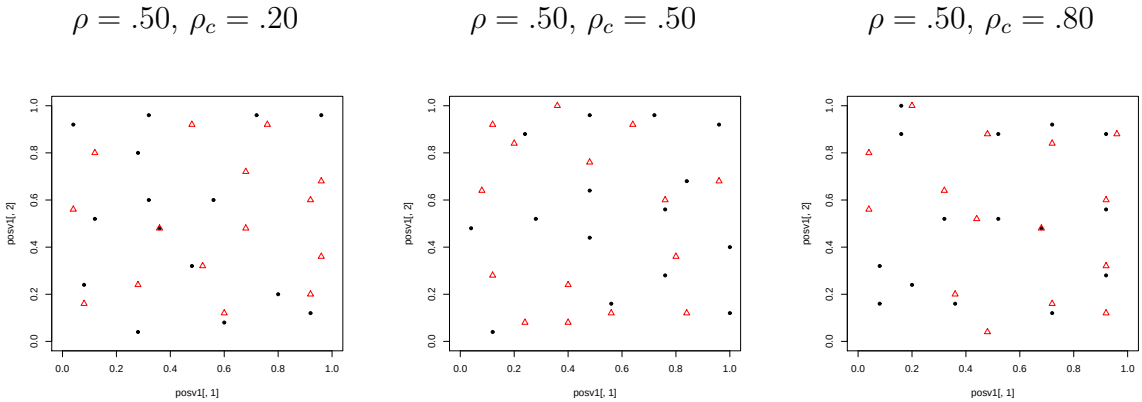

Figure 4.1: Optimal designs for prediction using SAA with $\Delta = 0.5$; black solid dots are for variable 1 and red empty triangles are for variable 2

(see Figure 4.2). Compared to the situation where $\Delta = 0.5$, here we see an increase in the separation of distance within pairs. We also considered the case for covariance parameter estimation when $\Delta = 2$, $\rho = 0.5$ and $\rho_c = 0.2, 0.5, 0.8$ (see Figure 4.3), and found that there were still clusters like we observed in Chapter 2.4.3 but none of the design points was collocated. Hence, our conclusion above is supported again, that is, adding a shift parameter into the covariance structure will result in a decrease in the degree of collocation.

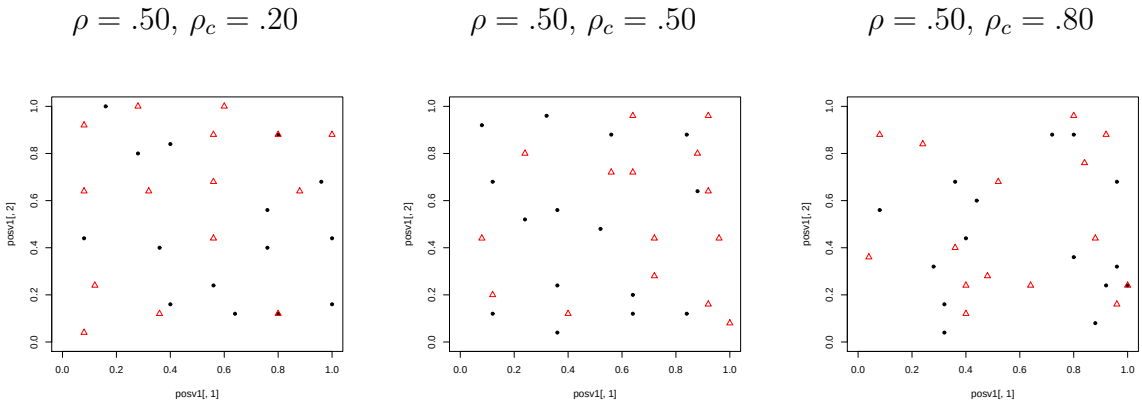

Figure 4.2: Optimal designs for prediction using SAA with $\Delta = 2$; black solid dots are for variable 1 and red empty triangles are for variable 2

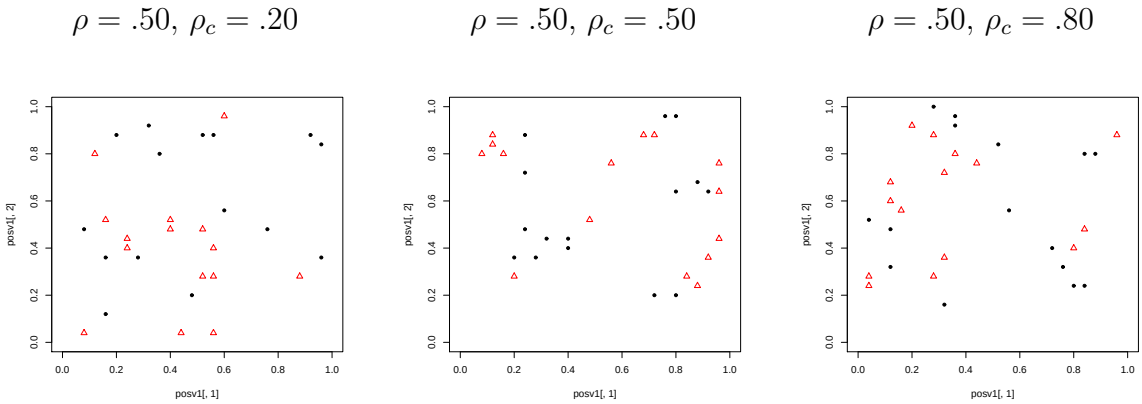

Figure 4.3: Optimal designs for covariance parameter estimation using SAA with $\Delta = 2$; black solid dots are for variable 1 and red empty triangles are for variable 2

CHAPTER 5
COLLOCATION OF OPTIMAL DESIGN IN THE PLANE

From the numerical studies presented thus far, we see that there often are some collocated design points in the optimal design. We are particularly interested in the occurrence of collocation in the optimal design. Suppose that we wish to monitor two air pollutants, for example, sulfur dioxide and sulfureted hydrogen, across all the cities in the United States. Then we need to build up monitor sites in cities in order to record the variable values. It is very expensive to build up a new monitor site as the construction fee and the daily maintenance fee will cost a lot. If we just add some new device to an existing monitor site to collect the new variable, instead of building up a new monitor site to collect the new variable, we may save a lot of money. So from an economic point of view, the design should have as many collocated points as possible. In this chapter, we will study the relationship between the optimal collocated design and the optimal univariate design.

5.1 Optimal collocated designs for prediction

Theorem 5.1 *Suppose* $\mathbf{Z}(\mathbf{s})$ *is a spatially-varying, Gaussian random process, where* $\mathbf{Z}(\mathbf{s}) = \begin{pmatrix} Z_1(\mathbf{s}) \\ Z_2(\mathbf{s}) \end{pmatrix} \in \Re^2$, $\mathbf{s} \in D \subset \Re^d$. *Suppose* $\mathbf{Z}(\mathbf{s})$ *has a constant mean and a separable covariance structure, which is defined as for any two points* $\mathbf{s}_i$ *and* $\mathbf{s}_j$, *where* $dist(\mathbf{s}_i, \mathbf{s}_j) = h$, $\text{cov}\begin{pmatrix} (Z_1(\mathbf{s}_i), Z_1(\mathbf{s}_j))' \\ (Z_2(\mathbf{s}_i), Z_2(\mathbf{s}_j))' \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_c \\ \sigma_1\sigma_2\rho_c & \sigma_2^2 \end{pmatrix} \otimes \begin{pmatrix} 1 & \rho^h \\ \rho^h & 1 \end{pmatrix}$. *If attention is restricted to collocated designs only, then the optimal design with respect to prediction for* $\mathbf{Z}(\mathbf{s})$ *is identical to the optimal design with respect to prediction for* $Z(\mathbf{s})$, *where* $Z(\mathbf{s}) \in \Re^1$, *which has a constant mean structure and a covariance function* $\text{cov}(h) = \text{cov}(Z(\mathbf{s}_i), Z(\mathbf{s}_j)) = \sigma^2\rho^h$.

Proof: For the univariate case, suppose that we observe variable values at $\mathbf{s}_1, \ldots, \mathbf{s}_n$, and we wish to predict at point $\mathbf{s}_0$. Then the univariate prediction criterion, i.e., the kriging

error variance, can be written as

$$\begin{aligned}
\sigma_k^2(\mathbf{s_0}) &= \sigma^2 - (\sigma^2\rho^{h_{10}} \dots \sigma^2\rho^{h_{n0}}) \begin{pmatrix} \sigma^2 & \cdots & \sigma^2\rho^{h_{1n}} \\ \vdots & \ddots & \vdots \\ \sigma^2\rho^{h_{1n}} & \cdots & \sigma^2 \end{pmatrix}^{-1} \begin{pmatrix} \sigma^2\rho^{h_{10}} \\ \vdots \\ \sigma^2\rho^{h_{n0}} \end{pmatrix} \\
&\quad + \left[1 - (1 \dots 1) \begin{pmatrix} \sigma^2 & \cdots & \sigma^2\rho^{h_{1n}} \\ \vdots & \ddots & \vdots \\ \sigma^2\rho^{h_{1n}} & \cdots & \sigma^2 \end{pmatrix}^{-1} \begin{pmatrix} \sigma^2\rho^{h_{10}} \\ \vdots \\ \sigma^2\rho^{h_{n0}} \end{pmatrix} \right]' \\
&\quad \begin{pmatrix} (1 \dots 1) \begin{pmatrix} \sigma^2 & \cdots & \sigma^2\rho^{h_{1n}} \\ \vdots & \ddots & \vdots \\ \sigma^2\rho^{h_{1n}} & \cdots & \sigma^2 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \end{pmatrix}^{-1} \\
&\quad \left[1 - (1 \dots 1) \begin{pmatrix} \sigma^2 & \cdots & \sigma^2\rho^{h_{1n}} \\ \vdots & \ddots & \vdots \\ \sigma^2\rho^{h_{1n}} & \cdots & \sigma^2 \end{pmatrix}^{-1} \begin{pmatrix} \sigma^2\rho^{h_{10}} \\ \vdots \\ \sigma^2\rho^{h_{n0}} \end{pmatrix} \right] \\
&= \sigma^2 \left(1 - \mathbf{C}_u' \mathbf{\Sigma}_u^{-1} \mathbf{C}_u + [X_u(\mathbf{s_0}) - \mathbf{X}_u' \mathbf{\Sigma}_u^{-1} \mathbf{C}_u]' (\mathbf{X}_u' \mathbf{\Sigma}_u^{-1} \mathbf{X}_u)^{-1} [X_u(\mathbf{s_0}) - \mathbf{X}_u' \mathbf{\Sigma}_u^{-1} \mathbf{C}_u] \right),
\end{aligned}$$

where $\mathbf{C}_u = \begin{pmatrix} \rho^{h_{10}} \\ \vdots \\ \rho^{h_{n0}} \end{pmatrix}$, $\mathbf{\Sigma}_u = \begin{pmatrix} 1 & \cdots & \rho^{h_{1n}} \\ \vdots & \ddots & \vdots \\ \rho^{h_{1n}} & \cdots & 1 \end{pmatrix}$, $X_u(\mathbf{s_0}) = 1$ and $\mathbf{X}_u = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$. For the bivariate case, suppose we observe both of the variables at $\mathbf{s}_1, \dots, \mathbf{s}_n$, and we wish to predict at point $\mathbf{s}_0$. Let $\mathbf{A} = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_c \\ \sigma_1\sigma_2\rho_c & \sigma_2^2 \end{pmatrix}$. It can be shown that

$$\mathbf{C} = \begin{pmatrix} \sigma_1^2\rho^{h_{10}} & \sigma_1\sigma_2\rho_c\rho^{h_{10}} \\ \vdots & \vdots \\ \sigma_1^2\rho^{h_{n0}} & \sigma_1\sigma_2\rho_c\rho^{h_{n0}} \\ \sigma_1\sigma_2\rho_c\rho^{h_{10}} & \sigma_2^2\rho^{h_{10}} \\ \vdots & \vdots \\ \sigma_1\sigma_2\rho_c\rho^{h_{n0}} & \sigma_2^2\rho^{h_{n0}} \end{pmatrix} = \mathbf{A} \otimes \mathbf{C}_u;$$

$$\mathbf{\Sigma}_0 = \mathbf{A};$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \cdots & \sigma_1^2\rho^{h_{1n}} & \sigma_1\sigma_2\rho_c & \cdots & \sigma_1\sigma_2\rho_c\rho^{h_{1n}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sigma_1^2\rho^{h_{1n}} & \cdots & \sigma_1^2 & \sigma_1\sigma_2\rho_c\rho^{h_{1n}} & \cdots & \sigma_1\sigma_2\rho_c \\ \sigma_1\sigma_2\rho_c & \cdots & \sigma_1\sigma_2\rho_c\rho^{h_{1n}} & \sigma_2^2 & \cdots & \sigma_2^2\rho^{h_{1n}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sigma_1\sigma_2\rho_c\rho^{h_{1n}} & \cdots & \sigma_1\sigma_2\rho_c & \sigma_2^2\rho^{h_{1n}} & \cdots & \sigma_2^2 \end{pmatrix} = \mathbf{A} \otimes \boldsymbol{\Sigma}_u;$$

$$\mathbf{X}(\mathbf{s}_0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{I}_2;$$

$$\mathbf{X} = \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix} = \mathbf{I}_2 \otimes \mathbf{X}_u;$$

$$\mathbf{X}_u(\mathbf{s}_0) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Based on the above equations, the bivariate prediction error variance is

$$\begin{aligned} & \mathbf{M}(\mathbf{s}_0) \\ = \; & \mathbf{A} - (\mathbf{A} \otimes \mathbf{C}_u)' (\mathbf{A} \otimes \boldsymbol{\Sigma}_u)^{-1} (\mathbf{A} \otimes \mathbf{C}_u) + \Big(\mathbf{I}_2 - (\mathbf{I}_2 \otimes \mathbf{X}_u)' (\mathbf{A} \otimes \boldsymbol{\Sigma}_u)^{-1} (\mathbf{A} \otimes \mathbf{C}_u)\Big)' \\ & \Big((\mathbf{I}_2 \otimes \mathbf{X}_u)' (\mathbf{A} \otimes \boldsymbol{\Sigma}_u)^{-1} (\mathbf{I}_2 \otimes \mathbf{X}_u)\Big)^{-1} \Big(\mathbf{I}_2 - (\mathbf{I}_2 \otimes \mathbf{X}_u)' (\mathbf{A} \otimes \boldsymbol{\Sigma}_u)^{-1} (\mathbf{A} \otimes \mathbf{C}_u)\Big) \\ = \; & \mathbf{A} \otimes \Big[1 - \mathbf{C}_u' \boldsymbol{\Sigma}_u^{-1} \mathbf{C}_u + [\mathbf{X}_u(\mathbf{s}_0) - \mathbf{X}_u' \boldsymbol{\Sigma}_u^{-1} \mathbf{C}_u]' (\mathbf{X}_u' \boldsymbol{\Sigma}_u^{-1} \mathbf{X}_u)^{-1} [\mathbf{X}_u(\mathbf{s}_0) - \mathbf{X}_u' \boldsymbol{\Sigma}_u^{-1} \mathbf{C}_u]\Big] \\ = \; & \mathbf{A} \otimes (\sigma_K^2(\mathbf{s}_0)/\sigma^2). \end{aligned}$$

Then it follows that

$$|\mathbf{M}(\mathbf{s}_0)| = \sigma_1^2\sigma_2^2(1-\rho_c^2) \cdot (\frac{\sigma_K^2(\mathbf{s}_0)}{\sigma^2})^2. \tag{5.1}$$

This means that if attention is restricted to collocated designs only, then the optimal bivariate collocated design for prediction is exactly the same as the univariate optimal design for prediction. □

5.2 Optimal collocated designs for covariance parameter estimation

Now we can prove similar results for optimal design for covariance parameter estimation. We have two different criteria in this case, a criterion for ML estimation and a criterion for REML estimation. In the following, we will show the equivalence of the univariate optimal design and bivariate optimal collocated design using each criterion. The same notations used in Theorem 5.1 are used here.

Theorem 5.2 *Suppose* $\mathbf{Z}(\mathbf{s})$ *and* $Z(\mathbf{s})$ *follow the same models as stated in Theorem 5.1. If attention is restricted to collocated designs only, then the bivariate collocated optimal design with respect to covariance parameter estimation is the same as the univariate optimal design with respect to covariance parameter estimation.*

Proof:

Maximum likelihood estimator

For the univariate case, let $\boldsymbol{\theta}_u = (\theta_1, \theta_2) = (\sigma^2, \rho)$, and we have

$$\frac{\partial(\sigma^2\boldsymbol{\Sigma}_u)}{\partial\theta_1} = \frac{\partial(\sigma^2\boldsymbol{\Sigma}_u)}{\partial\sigma^2} = \boldsymbol{\Sigma}_u$$

and

$$\frac{\partial(\sigma^2\boldsymbol{\Sigma}_u)}{\partial\theta_2} = \begin{pmatrix} 0 & h_{12}\sigma^2\rho^{h_{12}-1} & \dots & h_{1n}\sigma^2\rho^{h_{1n}-1} \\ \vdots & \vdots & \vdots & \vdots \\ h_{1n}\sigma^2\rho^{h_{1n}-1} & h_{2n}\sigma^2\rho^{h_{2n}-1} & \dots & 0 \end{pmatrix} = \sigma^2\frac{\partial\boldsymbol{\Sigma}_u}{\partial\rho}.$$

Since $Z(\mathbf{s})$ is a Gaussian process, if $\boldsymbol{\Sigma}$ is the covariance matrix of $Z(\mathbf{s})$, then the ij^{th} element of the Fisher information matrix associated with MLE is given by

$$\frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j}).$$

Hence, the Fisher information matrix can be written as follows:

$$\mathbf{I}_u^{ml} = \begin{pmatrix} \frac{1}{2}\mathrm{tr}((\sigma^2\boldsymbol{\Sigma}_u)^{-1}\boldsymbol{\Sigma}_u(\sigma^2\boldsymbol{\Sigma}_u)^{-1}\boldsymbol{\Sigma}_u) & \frac{1}{2}\mathrm{tr}((\sigma^2\boldsymbol{\Sigma}_u)^{-1}\boldsymbol{\Sigma}_u(\sigma^2\boldsymbol{\Sigma}_u)^{-1}\sigma^2\frac{\partial\boldsymbol{\Sigma}_u}{\partial\rho}) \\ \frac{1}{2}\mathrm{tr}((\sigma^2\boldsymbol{\Sigma}_u)^{-1}\boldsymbol{\Sigma}_u(\sigma^2\boldsymbol{\Sigma}_u)^{-1}\sigma^2\frac{\partial\boldsymbol{\Sigma}_u}{\partial\rho}) & \frac{1}{2}\mathrm{tr}((\sigma^2\boldsymbol{\Sigma}_u)^{-1}\sigma^2\frac{\partial\boldsymbol{\Sigma}_u}{\partial\rho}(\sigma^2\boldsymbol{\Sigma}_u)^{-1}\sigma^2\frac{\partial\boldsymbol{\Sigma}_u}{\partial\rho}) \end{pmatrix}.$$

Furthermore,

$$(\sigma^2\boldsymbol{\Sigma}_u)^{-1}\boldsymbol{\Sigma}_u(\sigma^2\boldsymbol{\Sigma}_u)^{-1}\boldsymbol{\Sigma}_u = \mathbf{I}_n/\sigma^4,$$

similarly, it can be shown that

$$(\sigma^2\boldsymbol{\Sigma}_u)^{-1}\boldsymbol{\Sigma}_u(\sigma^2\boldsymbol{\Sigma}_u)^{-1}\sigma^2\frac{\partial\boldsymbol{\Sigma}_u}{\partial\rho} = \frac{1}{\sigma^2}\boldsymbol{\Sigma}_u^{-1}\frac{\partial\boldsymbol{\Sigma}_u}{\partial\rho},$$

so the Fisher information matrix can be written as

$$\mathbf{I}_u^{ml} = \begin{pmatrix} \frac{n}{2\sigma^4} & \frac{1}{2\sigma^2}\mathrm{tr}(\mathbf{\Sigma}_u^{-1}\frac{\partial \mathbf{\Sigma}_u}{\partial \theta_2}) \\ \frac{1}{2\sigma^2}\mathrm{tr}(\mathbf{\Sigma}_u^{-1}\frac{\partial \mathbf{\Sigma}_u}{\partial \theta_2}) & \frac{1}{2}\mathrm{tr}(\mathbf{\Sigma}_u^{-1}\frac{\partial \mathbf{\Sigma}_u}{\partial \theta_2}\mathbf{\Sigma}_u^{-1}\frac{\partial \mathbf{\Sigma}_u}{\partial \theta_2}) \end{pmatrix}.$$

For the bivariate case, it can be shown that

$$\begin{aligned} \mathbf{I}^{ml} &= \begin{pmatrix} \frac{n(2-\rho_c^2)}{4\sigma_1^4(1-\rho_c^2)} & \frac{-n\rho_c^2}{4\sigma_1^2\sigma_2^2(1-\rho_c^2)} & \frac{1}{2\sigma_1^2}\mathrm{tr}(\mathbf{\Sigma}_u^{-1}\frac{\partial \mathbf{\Sigma}_u}{\partial \rho}) & \frac{-n\rho_c}{2\sigma_1^2(1-\rho_c^2)} \\ \frac{-n\rho_c^2}{4\sigma_1^2\sigma_2^2(1-\rho_c^2)} & \frac{n(2-\rho_c^2)}{4\sigma_2^4(1-\rho_c^2)} & \frac{1}{2\sigma_2^2}\mathrm{tr}(\mathbf{\Sigma}_u^{-1}\frac{\partial \mathbf{\Sigma}_u}{\partial \rho}) & \frac{-n\rho_c}{2\sigma_2^2(1-\rho_c^2)} \\ \frac{1}{2\sigma_1^2}\mathrm{tr}(\mathbf{\Sigma}_u^{-1}\frac{\partial \mathbf{\Sigma}_u}{\partial \rho}) & \frac{1}{2\sigma_2^2}\mathrm{tr}(\mathbf{\Sigma}_u^{-1}\frac{\partial \mathbf{\Sigma}_u}{\partial \rho}) & \mathrm{tr}(\mathbf{\Sigma}_u^{-1}\frac{\partial \mathbf{\Sigma}_u}{\partial \rho}\mathbf{\Sigma}_u^{-1}\frac{\partial \mathbf{\Sigma}_u}{\partial \rho}) & \frac{-\rho_c}{1-\rho_c^2}\mathrm{tr}(\mathbf{\Sigma}_u^{-1}\frac{\partial \mathbf{\Sigma}_u}{\partial \rho}) \\ \frac{-n\rho_c}{2\sigma_1^2(1-\rho_c^2)} & \frac{-n\rho_c}{2\sigma_2^2(1-\rho_c^2)} & \frac{-\rho_c}{1-\rho_c^2}\mathrm{tr}(\mathbf{\Sigma}_u^{-1}\frac{\partial \mathbf{\Sigma}_u}{\partial \rho}) & \frac{n(1+\rho_c^2)}{(1-\rho_c^2)^2} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{D} & \mathbf{E} \end{pmatrix}, \text{say,} \end{aligned}$$

where $\mathbf{B}$ is a 3×3 matrix at the upper left corner. By a well-known result on determinants of partitioned matrices,

$$\det(\mathbf{I}^{ml}) = \det(\mathbf{B})\det(\mathbf{E} - \mathbf{D}\mathbf{B}^{-1}\mathbf{C}).$$

Furthermore,

$$\begin{aligned} \mathbf{B} &= \begin{pmatrix} \frac{n(2-\rho_c^2)}{4\sigma_1^4(1-\rho_c^2)} & \frac{-n\rho_c^2}{4\sigma_1^2\sigma_2^2(1-\rho_c^2)} & \frac{1}{2\sigma_1^2}\mathrm{tr}(\mathbf{\Sigma}_u^{-1}\frac{\partial \mathbf{\Sigma}_u}{\partial \rho}) \\ \frac{-n\rho_c^2}{4\sigma_1^2\sigma_2^2(1-\rho_c^2)} & \frac{n(2-\rho_c^2)}{4\sigma_2^4(1-\rho_c^2)} & \frac{1}{2\sigma_2^2}\mathrm{tr}(\mathbf{\Sigma}_u^{-1}\frac{\partial \mathbf{\Sigma}_u}{\partial \rho}) \\ \frac{1}{2\sigma_1^2}\mathrm{tr}(\mathbf{\Sigma}_u^{-1}\frac{\partial \mathbf{\Sigma}_u}{\partial \rho}) & \frac{1}{2\sigma_2^2}\mathrm{tr}(\mathbf{\Sigma}_u^{-1}\frac{\partial \mathbf{\Sigma}_u}{\partial \rho}) & \mathrm{tr}(\mathbf{\Sigma}_u^{-1}\frac{\partial \mathbf{\Sigma}_u}{\partial \rho}\mathbf{\Sigma}_u^{-1}\frac{\partial \mathbf{\Sigma}_u}{\partial \rho}) \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{K} & \mathbf{L} \\ \mathbf{M} & \mathbf{N} \end{pmatrix}, \end{aligned}$$

where $\mathbf{K}$ is a 2×2 matrix at the upper left corner. By the well-known result on determinants mentioned above, we have

$$\det(\mathbf{B}) = \det(\mathbf{K})\det(\mathbf{N} - \mathbf{M}\mathbf{K}^{-1}\mathbf{L}).$$

It can be shown that

$$\det(\mathbf{K}) = \frac{n^2}{4\sigma_1^4\sigma_2^4(1-\rho_c^2)}$$

$$\mathbf{M}\mathbf{K}^{-1}\mathbf{L} = \frac{1}{n}\mathrm{tr}\left(\mathbf{\Sigma}_u^{-1}\frac{\partial \mathbf{\Sigma}_u}{\partial \rho}\right)\mathrm{tr}\left(\mathbf{\Sigma}_u^{-1}\frac{\partial \mathbf{\Sigma}_u}{\partial \rho}\right)$$

hence

$$\begin{aligned} \det(\mathbf{B}) &= \frac{n^2}{4\sigma_1^4\sigma_2^4(1-\rho_c^2)}\frac{1}{n}\left[n\mathrm{tr}\left(\mathbf{\Sigma}_u^{-1}\frac{\partial \mathbf{\Sigma}_u}{\partial \rho}\mathbf{\Sigma}_u^{-1}\frac{\partial \mathbf{\Sigma}_u}{\partial \rho}\right) - \mathrm{tr}\left(\mathbf{\Sigma}_u^{-1}\frac{\partial \mathbf{\Sigma}_u}{\partial \rho}\right)\mathrm{tr}\left(\mathbf{\Sigma}_u^{-1}\frac{\partial \mathbf{\Sigma}_u}{\partial \rho}\right)\right] \\ &= \frac{n^2}{4\sigma_1^4\sigma_2^4(1-\rho_c^2)}\frac{4\sigma^4}{n}\det(\mathbf{I}_u^{ml}) \\ &= \frac{n\sigma^4}{\sigma_1^4\sigma_2^4(1-\rho_c^2)}\det(\mathbf{I}_u^{ml}). \end{aligned}$$

It is not hard to show that

$$\mathbf{DB}^{-1}\mathbf{C} = \frac{n\rho_c^2}{(1-\rho_c^2)^2},$$

then it follows that

$$\begin{aligned}\det(\mathbf{I}^{ml}) &= \frac{n\sigma^4}{\sigma_1^4\sigma_2^4(1-\rho_c^2)}\det(\mathbf{I}_u^{ml})\left(\frac{n(1+\rho_c^2)}{(1-\rho_c^2)^2} - \frac{n\rho_c^2}{(1-\rho_c^2)^2}\right)\\ &= \frac{n^2\sigma^4}{\sigma_1^4\sigma_2^4(1-\rho_c^2)^3}\det(\mathbf{I}_u^{ml}).\end{aligned}$$

Hence, the optimal bivariate collocated design for covariance parameter estimation using MLE is the same as the optimal univariate design.

Restricted maximum likelihood estimator

For a Gaussian process, the ij^{th} element of the Fisher information matrix associated with REML is given by

$$\frac{1}{2}\mathrm{tr}(\mathbf{P}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}\mathbf{P}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j})$$

where $\mathbf{P} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}$. Again for the univariate case, the Fisher information matrix associated with REML estimation can be written as follows:

$$\mathbf{I}_u^{reml} = \begin{pmatrix} \frac{1}{2}\mathrm{tr}(\mathbf{P}_u\frac{\partial\boldsymbol{\Sigma}_u}{\partial\theta_1}\mathbf{P}_u\frac{\partial\boldsymbol{\Sigma}_u}{\partial\theta_1}) & \frac{1}{2}\mathrm{tr}(\mathbf{P}_u\frac{\partial\boldsymbol{\Sigma}_u}{\partial\theta_1}\mathbf{P}_u\frac{\partial\boldsymbol{\Sigma}_u}{\partial\theta_2}) \\ \frac{1}{2}\mathrm{tr}(\mathbf{P}_u\frac{\partial\boldsymbol{\Sigma}_u}{\partial\theta_2}\mathbf{P}_u\frac{\partial\boldsymbol{\Sigma}_u}{\partial\theta_1}) & \frac{1}{2}\mathrm{tr}(\mathbf{P}_u\frac{\partial\boldsymbol{\Sigma}_u}{\partial\theta_2}\mathbf{P}_u\frac{\partial\boldsymbol{\Sigma}_u}{\partial\theta_2}) \end{pmatrix}.$$

Furthermore,

$$\begin{aligned}&\mathbf{P}_u\frac{\partial(\sigma^2\boldsymbol{\Sigma}_u)}{\partial\theta_1}\mathbf{P}_u\frac{\partial(\sigma^2\boldsymbol{\Sigma}_u)}{\partial\theta_1}\\ =\ &\frac{1}{\sigma^2}(\boldsymbol{\Sigma}_u^{-1} - \boldsymbol{\Sigma}_u^{-1}\mathbf{X}_u(\mathbf{X}_u'\boldsymbol{\Sigma}_u^{-1}\mathbf{X}_u)^{-1}\mathbf{X}_u'\boldsymbol{\Sigma}_u^{-1})\boldsymbol{\Sigma}_u\frac{1}{\sigma^2}(\boldsymbol{\Sigma}_u^{-1} - \boldsymbol{\Sigma}_u^{-1}\mathbf{X}_u(\mathbf{X}_u'\boldsymbol{\Sigma}_u^{-1}\mathbf{X}_u)^{-1}\mathbf{X}_u'\boldsymbol{\Sigma}_u^{-1})\boldsymbol{\Sigma}_u\\ =\ &(\mathbf{I}_n - \boldsymbol{\Sigma}_u^{-1}\mathbf{X}_u(\mathbf{X}_u'\boldsymbol{\Sigma}_u^{-1}\mathbf{X}_u)^{-1}\mathbf{X}_u')/\sigma^4\\ =\ &\mathbf{R}/\sigma^4,\end{aligned}$$

where $\mathbf{R} = \mathbf{I}_n - \boldsymbol{\Sigma}_u^{-1}\mathbf{X}_u(\mathbf{X}_u'\boldsymbol{\Sigma}_u^{-1}\mathbf{X}_u)^{-1}\mathbf{X}_u'$ and $\mathrm{tr}(\mathbf{R}) = \mathrm{tr}(\mathbf{I}_n) - \mathrm{tr}(\boldsymbol{\Sigma}_u^{-1}\mathbf{X}_u(\mathbf{X}_u'\boldsymbol{\Sigma}_u^{-1}\mathbf{X}_u)^{-1}\mathbf{X}_u') = \mathrm{tr}(\mathbf{I}_n) - \mathrm{tr}((\mathbf{X}_u'\boldsymbol{\Sigma}_u^{-1}\mathbf{X}_u)^{-1}\mathbf{X}_u'\boldsymbol{\Sigma}_u^{-1}\mathbf{X}_u) = n-1$. Similarly, it can be shown that

$$\mathbf{P}_u\frac{\partial(\sigma^2\boldsymbol{\Sigma}_u)}{\partial\theta_1}\mathbf{P}_u\frac{\partial(\sigma^2\boldsymbol{\Sigma}_u)}{\partial\theta_2} = \frac{1}{\sigma^2}\mathbf{R}\boldsymbol{\Sigma}_u^{-1}\frac{\partial\boldsymbol{\Sigma}_u}{\partial\rho}$$

and

$$\mathbf{P}_u\frac{\partial(\sigma^2\boldsymbol{\Sigma}_u)}{\partial\theta_2}\mathbf{P}_u\frac{\partial(\sigma^2\boldsymbol{\Sigma}_u)}{\partial\theta_2} = \mathbf{R}\boldsymbol{\Sigma}_u^{-1}\frac{\partial\boldsymbol{\Sigma}_u}{\partial\rho}\mathbf{R}\boldsymbol{\Sigma}_u^{-1}\frac{\partial\boldsymbol{\Sigma}_u}{\partial\rho}.$$

Let $\mathbf{G} = \mathbf{R}\boldsymbol{\Sigma}_u^{-1}\frac{\partial\boldsymbol{\Sigma}_u}{\partial\rho}$ and $\mathbf{H} = \mathbf{R}\boldsymbol{\Sigma}_u^{-1}\frac{\partial\boldsymbol{\Sigma}_u}{\partial\rho}\mathbf{R}\boldsymbol{\Sigma}_u^{-1}\frac{\partial\boldsymbol{\Sigma}_u}{\partial\rho}$, then the Fisher information matrix may be written as

$$\mathbf{I}_u^{reml} = \begin{pmatrix} \frac{n-1}{2\sigma^4} & \frac{1}{2\sigma^2}\mathrm{tr}(\mathbf{G}) \\ \frac{1}{2\sigma^2}\mathrm{tr}(\mathbf{G}) & \frac{1}{2}\mathrm{tr}(\mathbf{H}) \end{pmatrix}.$$

For the bivariate case, it can be shown that

$$\begin{aligned}\mathbf{I}^{reml} &= \begin{pmatrix} \frac{(n-1)(2-\rho_c^2)}{4\sigma_1^4(1-\rho_c^2)} & \frac{-(n-1)\rho_c^2}{4\sigma_1^2\sigma_2^2(1-\rho_c^2)} & \frac{1}{2\sigma_1^2}\mathrm{tr}(\mathbf{G}) & \frac{-(n-1)\rho_c}{2\sigma_1^2(1-\rho_c^2)} \\ \frac{-(n-1)\rho_c^2}{4\sigma_1^2\sigma_2^2(1-\rho_c^2)} & \frac{(n-1)(2-\rho_c^2)}{4\sigma_2^4(1-\rho_c^2)} & \frac{1}{2\sigma_2^2}\mathrm{tr}(\mathbf{G}) & \frac{-(n-1)\rho_c}{2\sigma_2^2(1-\rho_c^2)} \\ \frac{1}{2\sigma_1^2}\mathrm{tr}(\mathbf{G}) & \frac{1}{2\sigma_2^2}\mathrm{tr}(\mathbf{G}) & \mathrm{tr}(\mathbf{H}) & \frac{-\rho_c}{1-\rho_c^2}\mathrm{tr}(\mathbf{G}) \\ \frac{-(n-1)\rho_c}{2\sigma_1^2(1-\rho_c^2)} & \frac{-(n-1)\rho_c}{2\sigma_2^2(1-\rho_c^2)} & \frac{-\rho_c}{1-\rho_c^2}\mathrm{tr}(\mathbf{G}) & \frac{(n-1)(1+\rho_c^2)}{(1-\rho_c^2)^2} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{D} & \mathbf{E} \end{pmatrix}, \text{say,}\end{aligned}$$

where $\mathbf{B}$ is a 3×3 matrix at the upper left corner. By a well-known result on determinants of partitioned matrices,

$$\det(\mathbf{I}^{reml}) = \det(\mathbf{B})\det(\mathbf{E} - \mathbf{D}\mathbf{B}^{-1}\mathbf{C}).$$

Similar as the maximum likelihood estimator case, we have,

$$\begin{aligned}\mathbf{B} &= \begin{pmatrix} \frac{(n-1)(2-\rho_c^2)}{4\sigma_1^4(1-\rho_c^2)} & \frac{-(n-1)\rho_c^2}{4\sigma_1^2\sigma_2^2(1-\rho_c^2)} & \frac{1}{2\sigma_1^2}\mathrm{tr}(\mathbf{G}) \\ \frac{-(n-1)\rho_c^2}{4\sigma_1^2\sigma_2^2(1-\rho_c^2)} & \frac{(n-1)(2-\rho_c^2)}{4\sigma_2^4(1-\rho_c^2)} & \frac{1}{2\sigma_2^2}\mathrm{tr}(\mathbf{G}) \\ \frac{1}{2\sigma_1^2}\mathrm{tr}(\mathbf{G}) & \frac{1}{2\sigma_2^2}\mathrm{tr}(\mathbf{G}) & \mathrm{tr}(\mathbf{H}) \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{K} & \mathbf{L} \\ \mathbf{M} & \mathbf{N} \end{pmatrix},\end{aligned}$$

where $\mathbf{K}$ is a 2×2 matrix at the upper left corner. By the well-known result on determinants mentioned above, we have

$$\det(\mathbf{B}) = \det(\mathbf{K})\det(\mathbf{N} - \mathbf{M}\mathbf{K}^{-1}\mathbf{L}).$$

It can be shown that

$$\begin{aligned}\det(\mathbf{K}) &= \frac{(n-1)^2}{4\sigma_1^4\sigma_2^4(1-\rho_c^2)} \\ \mathbf{M}\mathbf{K}^{-1}\mathbf{L} &= \frac{1}{n-1}\mathrm{tr}\,(\mathbf{G})\,\mathrm{tr}\,(\mathbf{G})\end{aligned}$$

hence

$$\det(\mathbf{B}) = \frac{(n-1)\sigma^4}{\sigma_1^4\sigma_2^4(1-\rho_c^2)}\det(\mathbf{I}_u^{reml}).$$

It is not hard to show that

$$\mathbf{D}\mathbf{B}^{-1}\mathbf{C} = \frac{(n-1)\rho_c^2}{(1-\rho_c^2)^2},$$

therefore

$$\det(\mathbf{I}^{reml}) = \frac{(n-1)^2\sigma^4}{\sigma_1^4\sigma_2^4(1-\rho_c^2)^3}\det(\mathbf{I}_u^{reml}).$$

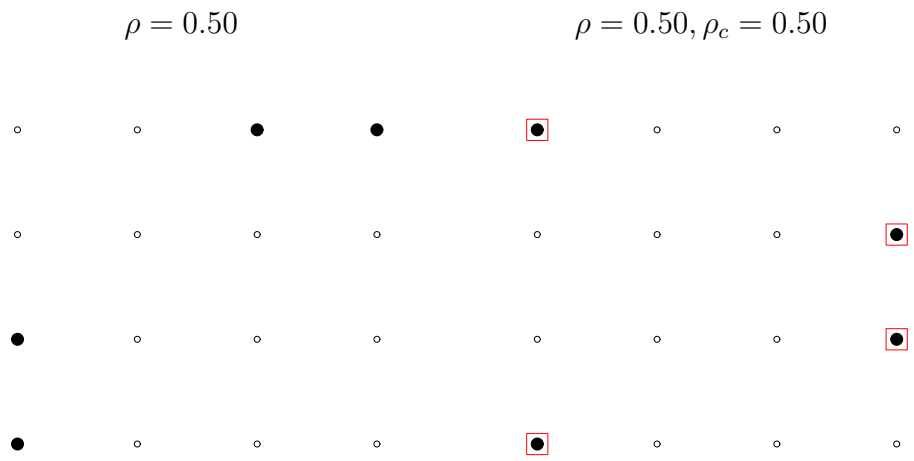

Figure 5.1: Comparison of the optimal univariate design with the optimal bivariate collocated design for empirical prediction

Then it follows that the optimal bivariate collocated design for covariance parameter estimation using REML is the same as the optimal univariate design.

Hence, we conclude that if attention is restricted to collocated designs only, then the optimal bivariate collocated design for covariance parameter estimation is exactly the same as the univariate optimal design. □

5.3 Optimal collocated designs for empirical prediction

However, similar results do not hold for empirical prediction when restricted to collocated designs only. In fact, numerical studies showed that the optimal bivariate collocated design under model (2.8) is not necessarily the optimal univariate design for empirical prediction when the covariance function is given by $\text{cov}(h) = \sigma^2\rho^h$. Figure 5.1 displays the optimal design for the univariate case (left) when $\rho = 0.50$ and the optimal bivariate collocated design (right) when $\rho = 0.50, \rho_c = 0.50$. Obviously, these two designs are not identical.

CHAPTER 6
EFFICIENCY OF COLLOCATED DESIGN

As mentioned before, collocation is especially useful in terms of saving the cost of data collection. Our studies have shown that optimal designs with respect to prediction, covariance parameter estimation and empirical prediction do have some collocated points. Now we will investigate how much efficiency in the optimized criterion we would lose if we were to seek the optimal designs within the class of completely collocated designs. Adding the collocation constraint can also speed up the simulation program. Table 6.1 through Table 6.6 show the relative efficiency loss by comparing the optimized criterion when constrained to collocated designs to the optimized criterion when there is no such constraint.

As can be seen from the tables, seeking the optimal design within the collocated design will result in some efficiency loss, especially for prediction (Table 6.1) and empirical prediction (Table (6.3)). In the toy example, the largest loss is 20%. There is no efficiency loss in the optimal design with respect to covariance parameter estimation (Table 6.2), as we already obtain completely collocated designs when we do optimization without the constraint.

However, adding the collocation constraint to the larger design problem where SAA is used, we may gain, rather than lose, efficiency. Table 6.4 shows that the efficiency loss for prediction is within 5%. But we actually get improvement for covariance estimation (Table 6.5) and empirical prediction (Table 6.6). Especially for covariance parameter estimation, we get great improvement by adding the collocation constraint to the optimal design, which is consistent with the toy example results where we have complete collocation in all the optimal designs.

These results suggest that we should start with a collocated design and always restrict to the collocated designs for large design problems using SAA. The efficiency loss is not large by doing that, and we could gain great efficiency for some cases.

ρ	0.20	0.50	0.80
$\rho_c = 0.50$	0%	5%	20%
ρ_c	0.20	0.50	0.80
$\rho = 0.50$	12%	5%	0%

Table 6.1: Efficiency loss when restricting to collocated designs for prediction using $\min\{\max_{i \in S} |\mathbf{M}(\mathbf{s}_i, \boldsymbol{\theta})|\}$

ρ	0.20	0.50	0.80
$\rho_c = 0.50$	0%	0%	0%
ρ_c	0.20	0.50	0.80
$\rho = 0.50$	0%	0%	0%

Table 6.2: Efficiency loss when restricting to collocated designs for covariance estimation using $\min\{1/|\mathbf{I}_{REML}(\boldsymbol{\theta})|\}$

ρ	0.20	0.50	0.80
$\rho_c = 0.50$	0%	7%	11%
ρ_c	0.20	0.50	0.80
$\rho = 0.50$	1%	7%	18%

Table 6.3: Efficiency loss when restricting to collocated designs for empirical prediction using $\min\{\max_{i \in S} |\mathbf{M}(\mathbf{s}_i, \hat{\boldsymbol{\theta}})|\}$

ρ	0.20	0.50	0.80
$\rho_c = 0.50$	2%	3%	2%
ρ_c	0.20	0.50	0.80
$\rho = 0.50$	1%	3%	5%

Table 6.4: Efficiency loss when restricting to collocated designs for prediction (SAA) using $\min\{\max_{i \in S} |\mathbf{M}(\mathbf{s}_i, \boldsymbol{\theta})|\}$

ρ	0.20	0.50	0.80
$\rho_c = 0.50$	-83%	-73%	-58%
ρ_c	0.20	0.50	0.80
$\rho = 0.50$	-69%	-73%	-60%

Table 6.5: Efficiency loss when restricting to collocated designs for covariance estimation (SAA) using $\min\{1/|\mathbf{I}_{REML}(\boldsymbol{\theta})|\}$

ρ	0.20	0.50	0.80
$\rho_c = 0.50$	-9%	-4%	0%
ρ_c	0.20	0.50	0.80
$\rho = 0.50$	-3%	-4%	-4%

Table 6.6: Efficiency loss when restricting to collocated design for empirical prediction (SAA) using $\min\{\max_{i \in S} |\mathbf{M}(\mathbf{s}_i, \hat{\boldsymbol{\theta}})|\}$

CHAPTER 7
CONCLUSIONS AND FURTHER STUDY

In Chapter 2 and 3, we studied optimal multivariate spatial design with respect to three objectives: prediction, covariance parameter estimation and empirical prediction. A multivariate optimal design criterion was developed for each objective. We considered two types of spatial domains in this thesis. The first type is a planar region, for which Euclidean distance is the metric used. The second type of domain is a stream or river network, for which stream distance is the metric. For the second type of domain, moving average constructions which incorporate both stream distance and flow direction were used to construct valid covariance models. We investigated the characteristics of designs through two types of numerical studies for both the planar case and stream network case: a small design problem and a large design problem using a simulated annealing algorithm. In these studies, we assumed a constant mean and an exponential covariance model. In order to see whether the design obtained was efficient, we compared the criterion value either to the median criterion value of all the designs or to the median criterion value of 10,000 randomly chosen designs. We applied our methodology to a real stream network from Alaska with regard to two different study objectives. The resulting optimal design was given in Section 3.5.

In terms of the spatial configuration of the design sites, we found that the optimal design for prediction tends to have widely spread out points. For precise prediction, it is desirable that for any unsampled point, there are design points that are close to it such that values of the variable at the sampled and unsampled points have large correlation. For covariance parameter estimation, there are distinct clusters in the optimal designs, reflecting the need for some small lags for good estimation of covariance parameters. We also found that the optimal design for empirical prediction shares some characteristics of those for prediction and covariance parameter estimation, that is, it has some small clusters but also some points broadly scattered over the study region. This is sensible,

since the design criterion we used for empirical prediction is a combination of the other two criteria.

Since we are dealing with multivariate design problems, it is of interest to study the positions of design points for the first variable relative to those for the second variable. In particular, we are interested in the degree of collocation in the optimal design because the higher the degree of collocation, the more practically feasible the design may be. In the planar case, the degree of collocation decreases as the spatial correlation increases, and increases as the cross-correlation increases, in the optimal design for prediction. We observed a high degree of collocation in the optimal designs for covariance parameter estimation. We also found that the degree of collocation in the optimal designs for empirical prediction lies between that for prediction and that for covariance parameter estimation in all the studies.

By comparing the criterion value either to the median criterion value of all the designs or to the median criteria value from 10,000 randomly chosen designs, we found that there are substantial improvements in the design criteria in most cases, which indicates that the designs we obtained tend to be efficient.

In Chapter 4, we added a shift parameter to the covariance function in order to allow asymmetry in the covariance function. We found that adding this shift parameter could reduce the degree of collocation in the optimal design. In Chapter 5, we studied the relationship between multivariate optimal design and univariate optimal design if we constrain the design to be complete collocated. Under this constraint, we found that the multivariate optimal designs for prediction and covariance parameter estimation coincide with the optimal univariate designs. However, this does not hold for empirical prediction. In Chapter 6, we further investigated how much efficiency we would lose if the design were constrained to be completely collocated. We found that especially for large design problems, we should restrict the design to be completely collocated.

In all of the case studies presented in this thesis, we assumed an exponential covariance function and a constant mean structure. For the toy example in the plane, however, we also studied optimal designs for a Matérn covariance function and a linear mean structure.

The results of those investigations were qualitatively similar to those with an exponential covariance function and a constant mean, hence they were not shown.

As a validation of the simulated annealing algorithm as an appropriate algorithm in this optimal design problem, we also tried the simulated annealing algorithm for the toy example in plane and on stream networks. We found that the algorithm picked out the optimal design or a design that is close to the optimal design in terms of the design criterion.

In this thesis, we assume that the two variables are of equal importance. However, in reality, sometimes one of the variables may be of more interest or importance than the other. To deal with this, we can consider putting different weights on the two variables when developing design criteria. For example, for bivariate prediction, suppose the prediction error variance matrix associated with BLUP is $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, then we can consider deriving the criterion based on a weighted prediction error variance matrix $\begin{pmatrix} wa & \sqrt{w(1-w)}b \\ \sqrt{w(1-w)}c & (1-w)d \end{pmatrix}$, where w in this matrix is the weight. We will discuss this issue further in our future work.

REFERENCES

Barnes, R. J. (1989). Sample design for geologic site characterization. *Geostatistics 2*, 809–822.

Bras, R. L. and I. Rodriguez-Iturbe (1976). Network design for the estimation of areal mean of rainfall events. *Water Resources Research 12*, 1185–1195.

Cressie, N., C. A. Gotway, and M. O. Grondona (1990). Spatial prediction from networks. *Chemometrics and Intelligent Laboratory Systems 7*, 251–271.

Harville, D. A. and D. R. Jeske (1992). Mean squared error of estimation or prediction under a general linear model. *Journal of the American Statistical Association 87*, 724–731.

Kackar, R. N. and D. A. Harville (1984). Approximations for standard errors of estimators of fixed and random efects in mixed linear models. *Journal of the American Statistical Association 79*, 853–862.

Lark, R. (2002). Optimized spatial sampling of soil for estimation of the variogram by maximum likelihood. *Geoderma 105*, 49–80.

Mardia, K. V. and R. J. Marshall (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika 71*, 135–146.

McBratney, A. B., R. Webster, and T. M. Burgess (1981). The design of optimal sampling schemes for local estimation and mapping of regionalize variables, i-theory and method. *Computer and Geosciences 7*, 331–334.

McCullagh, P. and J. A. Nelder (1989). *Generalized linear models (Second edition)*. Chapman & Hall Ltd.

Müller, W. G. and D. L. Zimmerman (1999). Optimal designs for variogram estimation. *Environmetrics 10*, 23–37.

Ver Hoef, J. M. and N. Cressie (1993). Multivarible spatial prediction. *Mathematical Geology 25*(2).

Ver Hoef, J. M. and E. E. Peterson (2008). A moving average approach for spatial statistical models of stream networks. *Manuscript*.

Ver Hoef, J. M., E. E. Peterson, and D. Theobald (2005). Spatial statistical models that use flow and stream distance. *Manuscript*.

Sain, S. R. and N. Cressie (2007). A spatial model for multivariate lattice data. *Journal of Econometrics 140*, 226–259.

Yfantis, E. A., G. T. Flatman, and J. V. Behar (1987). Efficiency of kriging estimation for square, triangular and hexagonal grids. *Mathematical Geology 19*, 183–205.

Zhu, Z. and M. L. Stein (2005). Spatial sampling design for parameter estimation of the covariance function. *Journal of Statistical Planning and Inference 134*, 583–603.

Zhu, Z. and M. L. Stein (2006). Spatial sampling design for prediction with estimated parameters. *Journal of Agriculture, Biology, and Environmental Statistics 11* (1), 24–44.

Zimmerman, D. L. (2006). Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. *Environmetrics 17*, 635–652.

Zimmerman, D. L. and N. Cressie (1992). Mean squared prediction error in the spatial linear model with estimated covariance parameters. *Annals of the Institute for Statistical Mathematics 44*, 27–43.