DISSERTATION

SMALL AREA ESTIMATION

Submitted by

Jin-Mann Lin

Department of Statistics

In partial fulfillment of the requirements

for the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring, 2003

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

COLORADO STATE UNIVERSITY

December 3, 2002

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY JIN-MANN LIN ENTITLED " SMALL AREA ESTIMATION " BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR PHILOSOPHY .

Committee on Graduate Work

Co-adviser

Adviser

Department Head

ABSTRACT OF DISSERTATION

SMALL AREA ESTIMATION

Small area estimation techniques with survey data are now widely being investigated in an increasingly crowded field. In forestry, on-the-ground inventories could collect very intensive grid inform· ·on, but this is costly. Therefore, utilizing the low-cost auxiliary information from remote sensing sources such as Landsat Thematic Mapper (TM) or topographical (Topo) information is preferred. Since the data is spatially dependent, it is crucial to develop procedures that combine the existing small area methods with spatial models. In the thesis, the following methods are proposed using transformed and untransformed data: spatial multivariate (MV) distributions on transformed data, spatial zero-inflated exponential (SZIE) models, spatial zero-inflated gamma (SZIG) models, spatial zero-inflated Poisson (SZIP) models and spatial zero-inflated negative binomial (SZINB) models. The spatial zero-inflated models are designed to accommodate the excessive number of zeros among the data. Further, the ancillary data is incorporated into simple spatial multivariate models, called spatial multivariate regression models. The comparison of the result from spatial multivariate regression models with general linear models (GLMs) and the Most Similar Neighbor (MSN) procedure was provided. Making predictions for non-sampled locations is also a subject of this study. In this thesis, a method for predicting an individual plot response at a non-sampled site on the 0.85-mile grid is determined, based on untransformed data and based on transformed data. Also, predictions for individual plot locations

and several simulations designed to yield realizations similar to the available data were investigated to see the reliability of mean squared prediction errors proposed.

One focus of this thesis is on modeling spatial dependence of the data for Siuslaw National Forest. The plot data used in this study is 1.7-mile lattice data; however, there is subplot data as close as 40.8 meters. Basically, there is little spatial dependence between plot data, but sufficient spatial dependence at the subplot level to be useful in predictions.

For spatial models without dealing with numerous zeros in our data, simple spatial multivariate model are considered for transformed data. Moment and maximum pseudo-likelihood (MPL) estimations are used. To obtain MPL estimators, the normality assumption is made. Such models can only make predictions for plots on the 1.7-mile grid. Best linear unbiased predictions (BLUP) are designed to make predictions for plots on shorter distance grids. BLUPs based on untransformed and transformed data for non-sampled sites are given. Sveral back-transformed predictors are compared to BLUPs based on untransformed data. Such predictions for non-sampled sites on 0.85-mile grid points were explored using distance-based correlation functions. The mean squared prediction errors (m.s.p.e.s) of the predictions are also provided.

To deal with zeros and non-zeros separately, spatial zero-inflated (SZI) models are proposed. Moment estimators for the directional mean parameters are given. Several predictions rules based on spatial zero-inflated models are also provided. Such spatial zero-inflated models cannot make predictions for non-sampled plots on the 0.85-mile grid. The performance of different models in this study have been compared. Spatial multivariate regression models incorporating auxiliary information did not show improvement based on reduced mean-squared errors.

For those prediction rules which do not have a simple closed form for their mean squared prediction errors, mean-squared errors are compared via simulations.

Jin-Mann Lin
Department of Statistics
Colorado State University
Fort Collins, Colorado 80523
Spring, 2003

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation and deep gratitude to my advisors, Duane C. Boes and Hans Schreuder. I greatly appreciate their enthusiasm, patience, encouragement and endless support during the completion of this work. Their dedication to research and their determination to complete each task to the fullest has been truly inspiring to me; their insights, ideas and guidance were crucial to accomplish this dissertation.

Many thanks are due to Professors Richard A. Davis, Paul Mielke and Robin Reich for their comments and assistance. I especially appreciate the generous help by Professor Robin Reich with Arc/Info. Additionally, I thank the assistance of the Beauro of Land Management in providing the Landsat TM data for the Siuslaw National Forest.

To all my friends who have listened to my incessant babblings about things in which you really had no interest, but knew were important to me, I thank you. You have made the good times my memories and the bad times disappear. Especially, I appreciate the generous help by Mike William for loaning me his super computer. I reserve special thanks to my parents for their love, enormous encouragement and limitless support.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1 Background

In recent years there has been an explosion in the amount of data collected in forest surveys. Data collected in these surveys can be used to estimate parameters for the areas, or strata, which were planned for in the survey design. With additional information from related variables estimation for small areas, or substrata, which otherwise would yield either no estimate at all or estimates with very large variances, can be attempted. Small area estimation techniques with survey data are now widely being investigated in the forestry literature to predict conditions at unobserved locations. On-the-ground inventories could collect very intensive grid information, but this is costly. Therefore, an alternative approach is to produce a map by interpolation from the sample data to other grid locations or ideally to construct a map for an entire area. The survey estimates for small areas are likely to have large prediction errors. Since there often is low-cost auxiliary information from remote sensing sources such as the Landsat Thematic Mapper (TM) or geographical information, we can utilize this information to improve estimates by classifying the similarity between sampled locations and locations of interest in our investigation. The small areas of study are 1-hectare plots (primary sampling units) in the Siuslaw National Forest. The key points of this thesis are first to find an "appropriate" prediction model to fit the data without considering the possibility of many zeros; second, to obtain reliable estimated errors of individual plot predictions; third, to handle zero outcomes and non-zero outcomes separately and to find an "appropriate" prediction model to fit the data with the nature of

many zeros; and to assess the improvement in mean-squared-error (MSE) due to considering large number of zeros in the data.

1.2 Study Objectives

The overall objectives of this study are

(1) To predict total number of live trees (Lt), total number of dead standing trees (M) and total basal area (Tba) on non-sampled 1-ha plot on a 0.85-mile grid in the Siuslaw National Forest based on observations on a 1.7-mile grid of sampled plots. Those non-sampled plots are located between the 1.7-mile grid plots measured in the Forest.

(2) To predict the seedlings variable such as total number of live small trees or seedlings (Lt_s), total number of dead small trees or seedlings (M_s), and total basal area of seedlings (Tba_s) using the model modified to separate zero outcomes and non-zero outcomes.

(3) To determine reliable approaches for error estimation in the predictions in (1) and (2) based on the untransformed data (original observations) and transformed data.

(4) Recommend whether such methods should be used by Forest Inventory and Analysis (FIA) and the National Forest Service (NFS).

(5) Recommend follow up research and data collection techniques as needed.

1.3 Thesis Outline

This study is organized into the following chapters. Chapter 2 presents a literature review of small area statistics, most similar neighbor, model selection, sample reuse, zero-inflated models and fuzzy coordinates. Chapter 3 describes

the data preparation and data summarization for ground data, satellite data and photo-interpreted data. None of the univariate and bivariate descriptive tools can capture the spatial feature for the data which belong to some location in space. In this chapter, data posting was used to describe the spatial feature and also sample correlations among the data were calculated. Such routine data analyses show the potential for spatial information in the plot data. Since distance-based correlations are intuitively plausible, we checked if there is a stronger spatial dependence for short distance lag, i.e. spatial dependence among subplot data.

Chapter 4 develops the methods for spatial models without auxiliary data. In this chapter we start with simple spatial models which regress the forest variables of interest on themselves with spatial dependence and which follow a multivariate distribution. The Siuslaw N.F. data is spatially a rectangular lattice with holes due to plots on the road, being located in non-forest areas, or because of inaccessibility, so we consider rectangular lattices with and without holes in this chapter. We discuss two parameter estimation methods: moment and maximum likelihood estimations (normal assumption is needed for the later estimation). In general, simulation studies for each model are examined as well as real data.

Chapter 5 investigates predictions of response variables at non-sampled plots on a 0.85-mile grid based on the observations at sampled plots, i.e. fill-in predictions or interpolations. Linear combinations of the observations are used, i.e. linear prediction, and minimization of mean-squared error (MSE) leads to the classical best linear prediction (BLP). Since our data are not normally distributed, transformations to normality are investigated. The 3-parameter lognormal transformation works well for all of mortality (M), number of live trees (Lt), and total basal area (Tba) data. We fit the distance-related autocorrelation function to the untransformed data (the Y scale) and the transformed data (the Z scale). For correlation function fitting, we consider the direct fitting to the Y scale and

indirect fitting to the Z scale. Then the predictions for non-sampled plots can be obtained by the predictions based on the Y scale or back-transforming the predictions based on the Z scale. The predictor is no longer a linear prediction after the back-transformations on the Z scale. In the last subsection, we conduct several simulation studies to assess the performance of prediction equations based on plot and subplot based data. Cross-validation is used for the performance of prediction equations based on the Y scale and the Z scale.

A detailed description of the methods for spatial zero-inflated data are presented in Chapter 6. Transformations to normality did not change the property of numerous zero outcomes in the data. Instead of doing transformation on them, we handle the zero and non-zero values separately. In this chapter, we propose spatial zero-inflated models to deal with the large number of zeros in the data. For count data, we study spatial zero-inflated Poisson (SZIP) models and spatial zero-inflated negative binomial (SZINB) models which are mixture models with one component degenerate at zero and the other components having a Poisson or Negative Binomial shifted to start at one. A Poisson distribution possesses the property of equal mean and variance. Through mixing several Poisson components, a SZIP model is more dispersed than a standard Poisson distribution. This chapter also investigates alternative models, spatial zero-inflated negative binomial (SZINB) models with an additional parameter relative to SZIP models. For continuous data, we introduce spatial zero-inflated exponential (SZIE) models to capture the phenomenon of numerous zeros in the data and alternative models, spatial zero-inflated gamma (SZIG) models. We will use moment estimation for each kind of models.

Chapter 7 adds the auxiliary data to the models which were introduced in Chapter 4. This chapter includes a comparison of the results for spatial regression models, with the ordinary linear regression models and most similar neighbor

procedures. In Chapter 8, prediction is studied using the spatial model with auxiliary data.

Finally, Chapter 9 summarizes the results and offers suggestions for future research including spatial cross validation, location perturbation models, Bayesian inference, and spatio-temporal models.

2. REVIEW OF LITERATURE

Improved estimation for small areas, also called small domains, can be done in three ways. First, small area estimation, second, additional sampling with regression estimation (sampling covariate information on non-sampled locations), and third additional sampling in the area (measuring a finer grid, such as sampling on a 0.85 mile grid rather than the current 1.7 mile grid). The last two usually involve considerable additional expense so we focus on small area estimation.

2.1 Small Area Statistics

Consider a large area, divided into several small areas. Domains are sub-populations of interest in a survey and small domains are also called small areas. The socio-demographic areas are often classified by gender, age, ethnicity, income class, etc. Small areas can be areas such as regions, states, counties, or forests or even primary sampling units (PSU) in forestry. Small areas may be sampling strata, but often are not. Small areas are small so the sample size in them may be small or even zero.

All small area estimation techniques use model-dependent estimators (Kalton 1987). All can be considered special cases of either linear or non-linear regression. For example Dalenius (1987) noted that synthetic estimation is just a special case of regression estimation. Dalenius believes that the mean square errors for individual area estimates is useful in assessing the reliability of predictions but not the average of those over a series of areas since that can hide some awful truths for certain areas. Reliable prediction errors for individual plots are needed. To

examine this point closely, we use several techniques to estimate the prediction errors for each plot and the prediction errors for the Siuslaw Forest. For small area estimation, the use of auxiliary information is likely to be severely restricted by the available data. If the auxiliary variables are only weakly correlated with the variable of interest, the estimates for all small areas will be close to one another and may not reflect the true variability in the population parameter estimates so the quality of the estimates are likely to be highly suspect. Also, there is a concern that the definitions of auxiliary variables used for the small area estimation and in the overall survey are not always consistent.

Kalton (1987) advocated that a cautious approach be used in generating small area estimates especially when published by government statistical agencies. Such estimates need to be clearly distinguished from conventional sample-based estimates. This is why the term "synthetic" estimates was coined.

Malec et al. (1997) examine small area inferences for binary variables in the National Health Interview Survey (NHIS). Although the NHIS sample size is very large, estimates were needed for 3000x72 subpopulations. Using cross-validation they showed that hierarchical Bayes estimates provided satisfactory inferences for parameters relative to empirical Bayes and especially randomization-based estimates. They note that synthetic estimates $\hat{\Theta}$, for subpopulations within each of the 51 geographical areas, are not variable enough to be plausible. Empirical Bayes methods would work well for larger sample sizes.

Ghosh and Rao (1994) is one in an extensive series of review articles on small area estimation and a very good one. Of the classes of estimators they discuss, we consider the following attractive ones for estimation of the above parameters for a small area:

a. Direct Estimation

Direct estimators use data only from the study units in the small area and time period of interest. Direct estimation is design unbiased or asymptotically design unbiased (ADU). It includes standard survey estimators like Horvitz-Thompson, Hajek, and ratio estimators. Direct estimation is not reliable if the sample size is extremely small.

b. Modified Direct Estimation

Modified direct estimators use data from outside the small area and/or time period of interest and implicitly use some statistical model. This estimation is model-assisted and remains design unbiased or ADU. Also, it could include survey regression estimators with slope coefficients estimated for auxiliary variables from all data. If sample size is not too small and good auxiliary information is available, then modified direct estimation may be reliable.

c. Indirect Estimation

Indirect estimators use data from outside the small area and/or time period of interest and explicitly use some statistical models to "borrow strength" across time or space. Generally, indirect estimation is not design unbiased or ADU. Indirect estimation may be OK if good auxiliary information is available and the model is correctly specified.

d. Synthetic estimation (indirect estimation approach)

A synthetic estimate for each small area is calculated by weighting and summing the estimates from each class of the small area. Such estimation techniques assume that small areas have the same characteristics as a large area for which an unbiased estimate is available, and use this estimate to derive estimates for the small areas. Also, auxiliary information correlated to the variable is studied. The higher the correlation, the better the estimate. As an example, we partition the population into large subpopulations g and assume that auxiliary information in the form of totals, X_{ig}, small area i, subpopulation g, is available. A

synthetic estimator of the small area total for area i is: $\hat{Y}_i = \sum_g (X_{ig}/X_{.g})\hat{Y}_{.g}$, where $\hat{Y}_{.g}$'s are direct estimates. If the model assumption holds, synthetic estimation borrows strength from the larger area across time. Synthetic estimation is sometimes synonymous with "model-based prediction". Unlike in regression estimation, we don't have to clearly specify models. Such methods have been used traditionally because of simplicity, applicability to general sampling designs, and potential of improved accuracy in estimation using information from similar small areas (Ghosh and Rao 1994). Synthetic estimation is satisfactory if good auxiliary information is available and the model is correctly specified.

e. Composite Estimation

This technique attempts to balance the potential bias of a synthetic estimator against the instability of a direct estimator, by weighting them. A direct estimator $\hat{Y}_{1i}$ and an indirect estimator $\hat{Y}_{2i}$ are combined to yield the composite estimator of total ($Y_i = \sum_g Y_{ig}$) of small area i :

$$\hat{Y}_i^C = w_i \hat{Y}_{1i} + (1 - w_i)\hat{Y}_{2i},$$

where w_i is a suitably chosen weight ($0 \leq w_i \leq 1$). If weights can be obtained objectively, this approach has the virtues of simplicity, applicability to general sampling designs, protection against estimation bias, and potential for improved accuracy. The choices for w_i can be some *ad hoc* weighting, James-Stein shrinkage rule, or w_i derived from small area models. The problem is that weights are usually subjective.

2.2 K-nearest Neighbor Methods

In forestry, considerable work is going on now in small area estimation, with much of the impetus provided by research in Finland. Multiple imputation methods (including regression models) and k-nearest neighbor techniques have been

proposed for continuous variables. In k-nearest neighbor techniques, field sample information is propagated to the entire population using a similarity function. In multiple imputations for each unit without sample data, a series of l predictions are made using randomly selected data with an underlying model and database. Then the data sets are analyzed separately and pooled into a final result, usually an average of the results.

Tomppo (1991) and Tokola et al. (1996) first classified pixels with field data to update categorical variables. The method relies on spectral distance using a Euclidean distance function between pixels on satellite imagery from pixel p for which the prediction is to be made to each pixel for which ground truth is available. Then the estimate for ground variable m for pixel p is:

$$\hat{m}_p = \sum_{j=1}^{k} w(j,p)\ m(j,p)$$

where the $w(j,p)$ are distance-related weights and $m(j,p)$ are the values of m in the k closest spectrally related pixels. Then classifications are imputed as the mode or nearest class (Franco-Lopez 1999).

Tomppo et al. (1999) modified and tested the Finnish satellite image-aided National Forest Inventory to develop a preharvesting forest inventory method to assess the volumes of potential log products in New Zealand Pinus radiata D. Don plantations. Data from 188 ground plots for a 1000 hectare block of radiata pine was combined with known stand boundaries from maps and data from a Landsat TM image to predict for each pixel and stand the pruned and unpruned sawlog, pulp and total standing volumes using a k-nearest neighbor algorithm, which calculates the estimates by weighting the values of k nearest neighbors with a Euclidean distance function. Some stand characteristics based on stand histories were also used as covariates as well as spectral characteristics from the TM. All estimates appeared to have low bias (2-3 %) but with high mean square errors (up

to 54 % for pixel-level pruned volume predictions). They indicate that the high mean square errors will be much less for larger areas according to earlier studies. Note however that the stands and plots are probably homogeneous, presumably they did not deal with plots falling in different conditions.

Franco-Lopez (1999) reviews methods for projecting and propagating forest plot and stand information. As he notes, considerable effort has gone into combining forest monitoring information, remote sensing, and geographic information systems (GIS) to develop maps for forest variables such as cover type, stand density and timber volume in the Nordic countries with emphasis on the k-nearest neighbor technique. The cross-validation estimator of overall accuracy (OA) applying the nearest neighbor only for cover type using weighting parameters and 200 bootstrap samples, the cross-validation estimator was 0.4745 and the bootstrap $.632^{+}$ estimator of OA was 0.5211 where $OA = 1 - Err$ and the mean of errors, $Err = \sum_{i=1}^{n}(y_i - \hat{y}_i)/n$, with y_i and $\hat{y}_i$, the actual and predicted dichotomous responses, i.e. does y belong or not to class i. These results may seem poor, but the OAs for cross-validation and bootstrap $.632^{+}$ did not improve much as the number of neighbors increased up to 15. Thus they are fairly representative of the type of results obtained by other methods in this region. This kind of bootstrapping methods is problematic. The k-NN methods that he used predicts the future observations using neighboring observations with proportional distance weights. But the bootstrapping approaches used in his study resampled the i.i.d observations without preserving the spatial dependence among the neighboring observations.

There are two ways to compute the distances among neighbors, Euclidean and Mahalanobis distances. Moeur and Stage (1995) proposed a different approach, most similar neighbor (MSN) inference, by bringing in information about the indicator attributes to define the similarity function. When the complete infor-

mation about a landscape cannot be collected, data are usually collected through multiphase sampling. They focus on two-phase sampling data. The first-phase attributes can be observed economically from maps or by remote sensing of the entire area characterized by indicator attributes, e.g. topographic characteristics, crown coverage by vegetation type, and canopy height. The second-phase sample provides more detailed information about important design attributes–say, a detailed stand inventory, census of bird species, or estimate of sediment production. Moeur and Stage diagramed the partioned data matrix as

$$\begin{pmatrix} \mathbf{Y}_{(nxr)} & \mathbf{Z}_{(nxq)} & \\ & & \mathbf{X}_{(N\mathbf{x}p)} \\ \hat{\mathbf{Y}}_{(N-n)\mathbf{x}r} & \hat{\mathbf{Z}}_{(N-n)\mathbf{x}q} & \end{pmatrix}$$

where $\mathbf{Y}_{n\mathbf{x}r}$ and $\mathbf{Z}_{n\mathbf{x}q}$ are the observed and derived design attributes, $\hat{\mathbf{Y}}_{(N-n)\mathbf{x}r}$ and $\hat{\mathbf{Z}}_{(N-n)\mathbf{x}q}$ are the design attributes to be estimated, and $\mathbf{X}_{N\mathbf{x}p}$ is the observed indicator attributes. The upper matrix parcels correspond to second-phase samples for which both $\mathbf{X}$ and $\mathbf{Y}$ are known; the lower matrix correspond to first-phase attributes, those for which only $\mathbf{X}$ is known. Also, Moeur and Stage defined the derived design attributes $\mathbf{Z}$ of dimension (N x q), as the elements of a one-to-one mapping of $\mathbf{Y}$, a subset of them, and/or functions of them particularly relevant to the response of the models – e.g., if $\mathbf{Y}$ are the basal areas for r species, then $\mathbf{Z}$ might be the basal area of the major or minor species. Similarity is defined using the relationships of $\mathbf{Z}$ to $\mathbf{X}$ and the similarity function is developed using canonical correlation analysis. There are six steps in the MSN analysis. Once the indicator and design attribute data are set up, software developed by Moeur and Stage can be used to perform the canonical correlation analysis, calculate the set of most similar neighbors, and compute summary statistics for the MSN estimates. The most similar neighbor technique is basically a stratified procedure where each non-sampled plot is classified into the same stratum as 'similar' sampled units. This inference procedure falls into the kNN procedure, considering

$k = 1$. A disadvantage of MSN is that the imputed values for unmeasured plots are the values of the most similar neighbors, so the predictions were limited to the range of the observations in the original sample.

Moeur and Hershey (1998) applied the following procedures for interpolating forest inventory data across the landscape: (1) geostatistical simulation (GS), (2) Most Similar Neighbor sampling inference (MSN), and (3) combined MSN-GS approach. They investigated the performance of the MSN predictions using FIA observed sample values and their predictions obtained by choosing the second most similar neighbor to the FIA data, the most similar being the plot itself.

The study most similar to this one is described in Moisen and Edwards (1999). They used generalized linear regression models to predict forest type and timber volume in the northern Utah mountains using predictor variables that include elevation, aspect, slope, geographic coordinates, and vegetation cover types derived from Advanced Very High Resolution Radiometer (AVHRR) and TM satellite data. The relative precision of estimates of area by forest type and mean cubic foot volume was assessed using six models including the standard double sampling for stratification used by the FIA unit. Model predictions were also compared by assessing various map accuracy criteria. The models used were standard double sampling for stratification with 6 strata, stratified random sampling with 9 AVHRR strata, stratified random sampling with 29 TM strata, generalized regression models called Topo which uses elevation, aspect, slope, and UTM coordinates as predictor variables, Atopo using AVHRR as well as the Topo predictor variables, and Ttopo, using TM and the Topo predictor variables. For area and mean cubic foot volume estimation, the use of photo interpretation (PI) or TM based data for stratification were only slightly more precise than the others, while models using topography and spatial coordinates alone were competitive. For mapping accuracy, the model including TM-based vegetation was the best one while to-

pography and spatial coordinate models provided substantial information at less cost. The field plots at the time of data collection were rotated into the condition of the center of the plot so that the issue of mixed condition plots did not arise. The authors note that topography and geographic position are highly correlated with certain ecological processes so that the distribution of vegetation and precipitation is primarily a function of elevation, latitude, and storm patterns from the west and the Gulf of Mexico with local effects due to slope exposure and/or aspect.

Kangas (1996) assessed the applicability of several small area estimation techniques in some forestry applications. He stressed that improved estimation can be done by utilizing additional information from areas adjacent to the one of interest and that better estimates can be made with improved covariate information. At the same time bias is injected using such information from adjacent areas and the bias is also affected by the independent variables used. The more reliable the model is used, the smaller the bias will be. However, a model can at best capture the main features of the variability encountered and the soundness of the inference depends on its validity.

2.3 Zero-Inflated Models

Real data can display overdispersion through an excess of zeros and transformations to normality can not change the property of a large number of zeros, so zero-inflated models are commonly used to deal with such kind of data.

The idea of adjusting the probability of zero outcomes for Poisson distribution was developed by Johnson and Kotz (1969, pp. 204-206) and termed the Poisson with zeros (PWZ) distribution. Other names for these distributions are "inflated" (Singh, 1966; Pandey, 1965) and "pseudo-contagious" Poisson distribution (Cohen, 1960). Winkelmann (2000) gave a good introduction for inflated

models, including zero-inflated, one-inflated, and zero-and-one inflated models. McLachlan and Peel (2000) discussed mixture models for discrete and continuous data. In their book, they labeled a mixture model with a degenerate and other non-degenerate components as a nonstandard mixture model.

If Y_i follows a standard "zero-inflated" Poisson (ZIP) model, then Y_i can be expressed as

$$Y_i = D_0 (1 - B_i) + P_i' B_i \equiv P_i' B_i$$

where $D_0 = 0$, the B_i's are independent Bernoulli random variables with mean p, $0 < p < 1$, and the P_i''s are Poisson with mean parameter λ.

It's unlikely to get a closed form for the maximum likelihood estimators (MLE) of λ and p. In Fong and Yip (1993), an EM-algorithm is proposed to estimate the parameters for a mixture model of two discrete distribution components including Binomial, Negative Binomial and Poisson.

Lambert (1992) proposed that p and λ depending on covariates forms the basis of the zero-inflated Poisson (ZIP) regression framework. In ZIP regression, the responses $\mathbf{Y} = (Y_1, \ldots, Y_N)'$ are independent and

$$Y_i = \begin{cases} 0 & \text{with probability } 1 - p_i \\ \text{Poisson}(\lambda_i) & \text{with probability } p_i \end{cases}$$

and the parameters $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)'$ and $\boldsymbol{p} = (p_1, \ldots, p_N)'$ satisfy

$$\begin{aligned} log(\boldsymbol{\lambda}) &= \mathbf{X}\boldsymbol{\beta} \\ logit(\boldsymbol{p}) &= log\left(\boldsymbol{p}/(1 - \boldsymbol{p})\right) = G\boldsymbol{\gamma} \end{aligned}$$

for covariate matrices $\mathbf{X}$ and G.

2.4 Fuzzy Coordinates

Since location in the field is the main relational feature of spatial models, accurate information on location in any inventory effort is critical. It is possible

that we would like to observe total basal area (Tba) at plot i, while in fact we observed Tba at somewhere else. This is known as misregistration error. The misregistration error could happen in both ground data and satellite data. Considering a four direction perturbation, an observed value at plot i, in fact, could be the value at one of four locations east, west, south, or north of plot i.

Observations on this process are made under possible perturbations in the locations. These perturbations can be a stochastic process, say $\{S_{(s_1,s_2)}\}$. The experimenter hoped to observe $\{Y_{(s_1,s_2)}\}$ intending to make inferences regarding some model for the process $\{Y_{(s_1,s_2)}\}$, but observed the process at location $\{S_{(s_1,s_2)}\}$ instead of (s_1, s_2). The objective is to study the loss in information in observing the location perturbed process rather than the Y-process. For example, a four direction perturbation $\{S_{(s_1,s_2)}\}$ process can be is defined by

$$\begin{aligned}
\{S_{(s_1,s_2)}\} &= (s_1, s_2) \text{ with probability } (w.p.)\ p_0, \\
&= (s_1 - 1, s_2)\ w.p.\ p_1, \\
&= (s_1 + 1, s_2)\ w.p.\ p_2, \\
&= (s_1, s_2 - 1)\ w.p.\ p_3, \\
&= (s_1, s_2 + 1)\ w.p.\ p_4,
\end{aligned}$$

where $0 \leq p_i$'s ≤ 1 and $\sum p_i$=1.

For a one-dimensional process, Lin (1996) found perturbation models $\{S_t\}$ satisfying some conditions such as t-mean which means that the expectation of $\{S_t\}$ is t, stationary increments for which the distribution of $S_{t+h} - S_t$ only depends on increments h, $\forall t$, etc, where t can be time or depth. Analogously, we can find perturbation models $\{S_{(s_1,s_2)}\}$ possessing some properties called (s_1, s_2)-mean and stationary increments in two-dimensional index.

Brimicombe (1998) introduced a fuzzy coordinates system for uncertainty of location in space and time. He utilized fuzzy numbers to construct a fuzzy coordi-

nate system for use in two, three and four dimensions which may be incorporated within a traditional GIS data structure. Coordinates of easting (E) and northing (N) could thus be represented in the form:

$$E^{(-e,+e)},\ N^{(-n,+n)}$$

where positive and negative signs are taken as implicit. In a two-dimensional triangular fuzzy coordinate system, the easting and northing coordinates represent a two-dimensional space and a fuzzy coordinate in a three-dimensional concept. The third dimension concerns the level of membership $\mu_E(x)$, $\mu_N(x)$ of E, N in ordinary linguistic terms of confidence, belief, certainty or plausibility. Also, Brimicombe introduced another term *expectation*, which is similar to the (s_1, s_2)-mean in the location perturbation scheme. Uncertainty exists both in attributes and coordinates. Here we are concerned with uncertainty in the coordinates, referred to as "location uncertainty."

2.5 Model Selection and Sample Reuse

Predictions and the assessment of prediction errors are often the primary goal in many area of statistics. A good model ought to fit the data well. The more variables/parameters are added to the model, the better the apparent fit ought to be. Model selection aims to balance the increase in fit against the increase in model complexity. In univariate or multiple regression model, the adjusted R-squared is the most commonly used criterion of model selection. R^2 always increases whenever a variable is added to the model, and therefore it will always recommend additional complexity regardless of the relative contribution to model fit. There are other model selection techniques such as Akaike's FPE (Akaike, 1969), Mallows's C_p (Mallows, 1973), Akaike Information Criterion (Akaike, 1973, 1974), one based on the Kullback-Leibler discrepancy, Bayesian Information Criterion

(BIC, Akaike, 1978), the Schwarz Information Criterion (SIC, Schwarz, 1978), etc. Hurvich and Tsai (1989) developed an improved small-sample unbiased estimator of the Kullback-Leibler discrepancy, AICC, which is one of the best model selection criteria in Time Series. McQuarrie and Tsai (1998) pointed out that no single model selection criterion will always be better than another; certain criteria perform best for specific model types.

As mentioned in the previous chapter, one of our objectives is to get a reliable estimated prediction error for individual plots. When the data are independent, the classical jackknife (Quenouille, 1949a, 1956; Tukey, 1958) and bootstrap (Efron, 1979, 1982) for estimating an error estimate can be used. But it is crucial that the assumption of independence among the data holds. For spatial data, generally the data are spatially correlated. Blind application of the classical jackknife and bootstrap techniques can be misleading. There are several methods of bootstrapping dependent data. Hall (1985) provided two principal approaches to bootstrapping dependent data, based on 'fixed' and 'moving' tiles (blocks), called block bootstrap methods. In the classical bootstrap method, we bootstrap the i.i.d. observations, but for dependent data we bootstrap the independent spatial pattern instead. To avoid the difficulties with the block bootstrap method for dependent data, Garcia-Soidan and Hall (1997) suggested an alternative sampling window approach. In this setting the spatial pattern is not resampled, but instead is examined repeatedly within a relatively small sampling window, which is moved to all possible locations within the data set. An estimate of a quantity of interest (such as distribution or a variance) may be obtained by averaging over locations of the window. Hall and Jing (1996) have demonstrated this approach for time series data and showed it is also true for certain spatial processes. Lahiri et al. (1999) utilized spatial subsampling to predict the spatial cumulative distribution function. If one makes parametric or semiparametric assumptions about large

and small scale variation in the data, it is possible to resample the i.i.d. components (Freednab and Peters, 1984; Solow, 1985). This is called a semiparametric bootstrap method.

Although grid sampling was used in the Siuslaw National Forest, it is a lattice with holes due to possible plot locations on the road, in the river, or not in the Forest. No matter what block size we choose in the block bootstrapping, we won't get congruent subareas (which each subarea has the same size of the sampled points). To avoid this difficulty, we utilize a semiparametric bootstrap method in the study and we compare its performance to a simple bootstrap method. Cross-validation (CV) is an alternative model selection technique based on data resampling we used in this study.

3. DATA PREPARATION AND DATA SUMMARIZATION

3.1 Study Area

The study area is the Siuslaw National Forest. As shown in Figure 3.1, the Forest is located in the western part of the state of Oregon and extends from Coos Bay to Tillamook. It consists of 8 counties, and comprises about 0.63 million acres (0.25 million hectares) administered by the United States Department of Agriculture Forest Service. The Siuslaw Forest runs north-south with the length of 135 miles (217 km) and the width of 27 miles (44 km). Elevation ranges from sea level to 4049 feet (1234.14 m). Established in 1908, the Forest takes its name from a Yakonan Indian word meaning "far away waters". In the Forest, there are 30 natural lakes and 1200 miles of anadromous (meaning ascending from the sea or running upward for breeding referring to the life history of salmon and steelhead) streams. Major forest types are Big Leaf Maple, Douglas-fir, Red Alder, Sitka spruce, Western Hemlock, and Western Red Cedar.

3.2 Field Data

A grid sampling strategy was adopted for broad-scale inventory and monitoring of the forests. The grid consisted initially of a regular square spacing with 3.4 miles (5.47 km) between grid points; but for land management planning, Forest plan implementation, and monitoring changes in land and vegetation for all NFS lands excluding wilderness, a more intensive grid of 1.7-mile (2.74 km) was used ultimately. If the 1.7-mile grid for data acquisition did not meet all the vegetation or modeling needs of the forest planning effort, a 0.85-mile grid of 1-ha

plots was installed. This kind of grid sampling method is often implemented to explore mineral and fossil fuel deposits in the mining field. In the core drilling, the larger grid sampling was adopted in small area D, and then the observations were obtained at locations $D_n = \{s_1, \ldots, s_n\}$ where s_i's denote the locational indices of the i-th sample point. While we use the finer grid sampling in the same area D, more and more observations might be sampled in the same area. In this way, the finer grid sampling we use, the more observations we obtain. The idea is to infill the more extending sample points between existing points. This kind of grid sampling is called the infill sampling technique in the drilling problem. In forestry, the same kind of sampling strategy has been used, but it hasn't been termed "infill sampling".

The field data for the Current Vegetation Survey (CVS) in the Siuslaw Forest was collected at two 5-year cycle periods, called the first occasion data and the second occasion data. The first occasion data was collected between 1993 to 1997. The installations for 1-ha plots in the Siuslaw National Forest are listed in Table 3.1(a). During this period there were numerous changes to field procedures and plot design such as the 1/20-acre subplot changed to 1/24-acre subplot for remeasurement at the second occasion. A remeasurement panel consists of approximately 25% of the first occasion sample units. Originally, a different panel was to be remeasured every 2 years on each Forest resulting in the entire Forest being totally remeasured over an 8-year period. Beginning with the 1999 projects, however, the remeasurement cycle was adjusted to a panel every 3 years with the total coverage completed every 12 years. The second occasion data hasn't been completed, so in this thesis we only study the first occasion CVS data set of the Siuslaw National Forest. The CVS data for the Siuslaw Forest can be downloaded at http://www.fs.fed.us/r6/survey/data_tables.htm. The primary sample unit (PSU) in the sampling system is a 1-hectare (1-ha) circular plot as shown

in Figure 3.2. The 1-ha plots were established on a 3.4-mile and a 1.7-mile offset grid for the Siuslaw Forest. To establish the PSU and to locate positions within the PSU, stake position 1 representing the center of the PSU is installed first. Stake positions 2 though 5 are at different cardinal directions, 133.9 feet (40.8 m) away from stake position 1. If stake position 1 could not be installed, one of the remaining installable stake positions was used as the PSU reference. When none of the stake positions are on National Forest land or if all stake positions are on National Forest lands but are not accessible by foot, a PSU is not established. A stake position with corresponding subplot represents one-fifth of the PSU area. Each 1/5-ha area contains 6 subplots (4 concentric fixed areas and 2 line samples), as diagramed in Figure 3.2. The stake position is the center of each circular subplot and the starting point of each sample line. Each subplot is installed in the following order, 1/100 acre, 1/20 acre (2nd occasion, remeasured on 1/24 acre) , 1/5.3 acre, sample line, and 1/2 acre (one-fifth of the 1-ha plot, the corresponding areas to stake position numbers, as illustrated in Figure 3.3). Information was collected on 1-ha plots subsampled by 5 subplots of different radii depending on the size of trees. The concentric subplot sizes and associated diameter ranges are as follows:

1. 0.004 hectare (1/100 acre) for trees 1.0″ to 4.9″ diameter breast height (DBH) which is tree diameter outside bark at breast height of 3.4 inches, (diameter at a height of 1.3 m (4′3″) on a tree steam and seedlings which are trees 6″ tall and less than 1″ DBH.

2. 0.02 hectare (1/20 acre) (2nd occasion, remeasured on 1/24 acre) for trees 3.0″ to 12.9″ DBH, excluding trees 3.0″ to 4.9″ DBH tallied on the 0.004 ha.

3. 0.076 hectare (1/5.3 acre) for trees 13.0″ to 47.9″

4. 0.2 hectare (0.5 acre) for trees $>$ 48″ DBH

In the Siuslaw Forest, there are 86 sample plots on a 3.4-mile grid and 242 additional sample plots on a 1.7-mile grid (Figure 3.4). The resulting sampling intensity was approximately 0.0013 plots/ha. For the 3.4-mile grid, there are 5 plots that were not installed since they were not considered forest as shown by condition indicated by the following particular plant association codes, two plots with GR82 (Hummocks beachgrass, sand dune geology), one with NRL912 (steep, moist rock land with minimal vegetation potential), one with NSN0 (open sand of any dunal character, no vegetation), and one with NSG0 (sand dunes with scattered grass). Hence the available plots are 81. For the 1.7-mile grid plots, 3 additional plots were omitted; one was not installed due to the plant association, with both NSN0 and NSG8 (Coastal sand dune, rolling, partial beachgrass stability), one was installed but had no trees or indicator species and plant association (NSG8), and one was installed but had no trees tallied (NSN0). There is more detailed information about installations for some particular plots summarized in Table 3.1(b).

In this study, we consider only the fully installed stake position areas and the plots with at least 3 stake position areas having trees tallied. Thus Plots 2067024 and 2089032 are excluded in the study due to having only 2 stake positions tallied. Hence, the number of studied 1-ha plots was reduced to 313. Plot 2125072 which is marked by $\odot$ in Figure 3.4 was installed and tallied but not contiguous to the rest of the Forest. Instead of throwing it out immediately, we studied the data with and without it. Figure 3.5 displays the lattice locations of all 313 plots on the 1.7-mile grid.

With the 3.4-mile grid or with the 1.7-mile grid additional plots, the resulting intensities are respectively 0.00034 and 0.00095 plots/ha. At each sample plot, forest variables such as total number of dead standing trees (mortality (M)) excluding very small trees and seedlings tallied on the 0.004 ha., total number of live

trees (Lt) with diameter at breast height (DBH) $\geq 3''$ excluding very small trees and seedlings tallied on the 0.004 ha., total basal area (Tba) (excluding very small trees and seedlings tallied on the 0.004 ha.) which is the sum of cross-sectional areas of all live trees having a diameter breast height (DBH) of 1 inch or larger measured in m^2/ha. Also, we are interested in the mortality of seedlings, total number of live seedlings, and total basal area of seedlings which are for those very small trees and seedlings tallied only on the 0.004 ha., referred to as M_s, Lt_s, and Tba_s. These are the variables we selected for our study. The variable Tba, linearly corrected with tree volume, is easy to measure and has been one of the primary variables of interest for a long time as is number of trees per hectare. Tba is calculated from measurements of the diameter at breast height (DBH) of all live trees at breast height ($DBH^2 \times \pi/4$). The number of live trees per hectare is of ecological interest and is also needed to estimate volume/ha. Mortality, the number of dead standing trees in the 1-ha PSU, is of considerable interest to ecologists and foresters as one indicator of forest health.

The actual subplot measurements were used to generate estimates for the 1-ha PSU, using the following formula:

$$y = \frac{\sum_{j=1}^{n_4} y_{4j}}{0.2\, n_4} + \frac{\sum_{j=1}^{n_3} y_{3j}}{0.076\, n_3} + \frac{\sum_{j=1}^{n_2} y_{2j}}{0.02\, n_2}$$

and

$$y_s = \frac{\sum_{j=1}^{n_1} y_{1j}}{0.004\, n_1}$$

where y is the estimate of the attribute excluding those on the 0.004-ha. circular subplot, y_s is the corresponding estimate for very small trees and seedlings on the 0.004-ha. circular subplot, y_{ij} is the j-th tree attribute on subplot of size i, and n_i is the total number of trees on subplot of size i, as illustrated on p.22, $i = 1, \ldots, 4$. This calculation is based on the inclusion probabilities (also called

the expansion factor in forest literature) of the actual areas sampled relative to a hectare plot area.

The plot attributes were combined with information of the remote sensing analysis and ancillary data described below.

3.3 Landsat TM Data

The satellite imagery from the Landsat 5 Thematic Mapper (TM) data was provided by the US Department of the Interior Bureau of Land Management (BLM). TM data has 7 spectral bands encompassing the blue, green, red (R), near-infrared (NIR), thermal infrared (TIR) and mid-infrared (MIR) regions of the electromagnetic spectrum. The resolutions for each spectral band data are described in Table 3.2.

Reflectance is the ratio of the amount of light leaving the target to the amount of light striking the target, whereas radiance is the variable directly measured by remote sensing instruments. Basically, radiance is how much light the instrument sees from the object being observed expressed in watts/steradiance/square. But reflectance has no units. The number coming from the bands are light reflectance values in different wavelengths (in micrometers). The larger the number, the higher the reflectance. Because the resolution of TM spectral band 6 is coarse, we only used TM bands 1-5 and 7. We used the primary portion of Landsat in path 47 row 29 and a small portion falling within row 30 in path 46 (August 1995). A resampling procedure was used to reduce the pixels to 25 m x 25 m from 30x30 resolution. To ensure compatibility between images and with the ground data, each image was rectified and geo-referenced to the Universal Transverse Mercator (UTM) system using the parameter: Spheroid Clarke 1866, datum NAD27 (values determined by Clark in 1866 described the spheroid commonly used with the reference datum for North America; this is often referred to as the North American

Datum 1927) and Zone 10. For the UTM system, the earth is divided into sixty zones, each spanning six degrees of longitude. Each zone has its own central median and spans 3 degrees west and 3 degrees east. This system is a specialized application of the Transverse Mercator projection. The limits of each zone are 84° N and 80° S. Using the UTM system, the coordinates of the plot location are described by (UTMe, UTMn) in meters where e and n indicate easting and northing respectively.

3.4 Image Enhancement and Ancillary Data

There are several common ways to enhance the TM data such as filtering, Tasseled Cap transformations (Crist and Cone, 1984) and ratioing transformations.

The Tasseled Cap transformation of Landsat data combined the original six TM bands (1-5 and 7) into 3 linear combinations labeled: measures of vegetation (brightness), soil (greenness), and the interrelationship of soil and canopy moisture (wetness). The "greenness" of a reference green surface can be used to determine threshold values above which a given surface could be considered "green". The empirically derived coefficients are

$$\begin{aligned}
\text{Brightness} &= 0.3037\,\text{Band}_1 + 0.2793\,\text{Band}_2 + 0.4743\,\text{Band}_3 + 0.5585\,\text{Band}_4 \\
&\quad +0.5082\,\text{Band}_5 + 0.1863\,\text{Band}_7 \\
\text{Greenness} &= -0.2848\,\text{Band}_1 - 0.2435\,\text{Band}_2 - 0.5436\,\text{Band}_3 + 0.7243\,\text{Band}_4 \\
&\quad +0.0840\,\text{Band}_5 - 0.1800\,\text{Band}_7 \\
\text{Wetness} &= 0.1509\,\text{Band}_1 + 0.1973\,\text{Band}_2 + 0.3279\,\text{Band}_3 + 0.3406\,\text{Band}_4 \\
&\quad -0.7112\,\text{Band}_5 - 0.4572\,\text{Band}_7.
\end{aligned}$$

The commonly used ratioing transformations of 6 Bands data are the simple vegetation index (VI), which is calculated by NIR/R, the normalized difference

vegetation index (NDVI) which is similar transformation to the VI but with the normalization and is calculated from the reflected solar radiation in the near-infrared (NIR) and red (R) wavelength bands via the algorithm: NDVI=(NIR-R)/(NIR+R), varying between -1 and +1, and the ratioing of TM band 7 with TM bands 5, 4, and 3.

The combination of bands has proven quite useful for general vegetation analysis. $Band_3$ is strongly absorbed by active vegetation, whereas $Band_4$ is strongly reflected. $Band_5$ is sensitive to moisture in vegetation and soil. $Band_7$ is also sensitive to vegetation moisture content and as a result on sites of higher basal area there should be a corresponding decrease in the TM band 7 response. There are 3 ratios adopted in this study; $Ratio_1$: $Band_4/Band_3$ i.e. VI (NIR/R), $Ratio_2$: $Band_5/Band_4$, and $Ratio_3$: $Band_5/Band_7$. Vegetation is usually shown in red by remote sensing people. The reason is the fact that the human eye (0.6 - 0.7 μm) perceives the longest visible wavelengths to be red and the shortest visible wavelengths to be blue. Therefore, we are not seeing real colors. $Ratio_1$ utilizes the NIR and red (R) region of the spectrum which represents active vegetation. $Ratio_2$ and $Ratio_3$ accentuate the difference between the change in vegetation moisture content in band 7 and the relationship between infrared bands 5 and 4.

A 1-ha plot is covered by 13 pixels of size 25m x 25m. In the window filtering, we consider the mean of the 13 grid pixels, the median of the 13 grid pixels, or the majority of the 13 grid pixels. We compare the different enhancements against the original Band reading at the center pixel. We integrated other "spatial data layers" such as climate, Digital Elevation Models (DEMs) based on topographic measures, geology, soil and vegetation data, but did not use all of them in this study because the information is not available for some non-sampled plots.

A DEM is a digital representations of the shape of the earth's surface. The Bureau of Land Management (BLM) has a DEM that covers all of Oregon. For the

interagency vegetation mapping project (IVMP), slope, aspect (direction of slope), and elevation were generated in ERDAS Imagine format and ArcInfo grid format with spatial resolution 25m x 25m. There are many possible transformations of the aspect variable, but the commonly used ones are the cosine and the sine transformation. To avoid the problem of modeling a circular predictor variable, a transformation used by Roberts and Cooper (1989) was also applied to aspect, where

$$\text{Asp}_{RC} = 0.5\left\{cos\left(\frac{\pi}{180}[aspect - max(aspect)]\right) + 1\right\}. \tag{3.1}$$

The transformed variable ranges from zero to one and yields the highest values at the maximum of the aspect values. Similarly, we can transform aspect using the sine function instead of the cosine. The transformation function yields

$$\text{Asp}_{S} = 0.5\left\{sin\left(\frac{\pi}{180}[aspect - max(aspect)]\right) + 1\right\}. \tag{3.2}$$

In this study, values of the variables of interest at the interpolated points were mainly estimated from the 1.7-mile grid data. If the standard grid (1.7 mile square) for data acquisition does not meet all vegetation or modeling needs of the forest planning effort, a supplemental 0.85-mile grid was installed. But the predictions of 0.85-mile grid plots are also needed for other locations for management purposes.

3.5 Data Description

In this section we deal with univariate description first and then look at the relationships between pairs of variables. Finally, we incorporate the location of the data set and consider several ways of describing the spatial features of the data set.

3.5.1 Univariate description

A boxplot is a way to look at the overall shape of a set of data. The central box shows the data between the lower and upper quartiles, with the medians represented by a blank line inside the box. Brackets go out to the extremes of the data and very extreme values are portrayed by segment lines outside the brackets. The size of the box represents the spread about the central value (mean). Figure 3.6 portrays the boxplots for the field data of the the M, M_s, Lt, Lt_s, Tba, and Tba_s values. All variables except for M_s form a box; for Tba, there are only a few extreme values outside the brackets and the central box is half divided by the median blank line, i.e. the data have symmetry. An alternative presentation of data to the boxplot is the histogram. The histograms for M, M_s, Lt, Lt_s, Tba, and Tba_s are shown in Figure 3.7. The histograms for M and Lt are left skewed and for Tba the histogram appears to be symmetric and more normal. The histograms for M_s, Lt_s, and Tba_s have large mass at zero, so it might be dangerous to do further analyses ignoring the 0's among the data.

The descriptive statistics for the forest variables and the auxiliary variables are summarized in Table 3.3. The mean and median can give us some idea where the center of the distribution lies. The interquartiles (Q_1 and Q_3) and standard deviation (S.D.) are used to describe the variability of the data values. These descriptive statistics provide a valuable summary of the information contained in the histograms and boxplots. Except for Tba, the means and medians for the other variables differ considerably, as shown in the boxplots of M and Lt in Figure 3.7 and in the histograms of M and Lt in Figure 3.7

The descriptive statistics show the large number of zeros characteristic of histograms for M_s, Lt_s, and Tba_s; for example, the 3rd quartile for M_s is $Q_3 = 0$ which means at least 75% zeros, about 50% for Lt_s since Q_1=0, median=50.00

and mean=250.00, and the percentage of zeros for Tba_s is higher than 50% and lower than 75%.

The coefficients of variation (sample standard deviation/sample mean) for M, M_s, Lt, Lt_s, Tba, and Tba_s are 1.08, 3.23, 0.73, 1.79, 0.56, and 2.23, respectively. A large coefficient of variation can be produced by the presence of some unusually large sample values that may have a significant impact on the final estimates. The larger the values of coefficients of variation, the longer the tail of the histograms. The coefficient of variation for seedling variables M_s, Lt_s, and Tba_s are greater than 1, which are shown long tails in their histograms. The coefficient of variation for Tba is 0.56, which reflects the fact that the histogram does not have a long tail of large values (see Figure 3.7) and tends to normality.

Figure 3.8 portrays the boxplots and histograms of the topographic data. All boxplots of the topographic data form a nice central box. Aspect data has 6 extreme values and slope has one. The Landsat TM band data are boxplotted in Figure 3.9 and the histogram is shown in Figure 3.10. Band 4 has the most symmetrical boxplot of the band data. From Figure 3.9, we can see that there is no segment line outside the brackets, which means no extreme values for Band 4 data. After the Tasseled Cap transformation of Landsat TM band data, the 3 Cap data, brightness, greenness, and wetness are shown in Figure 3.11. Note that all three Tasseled Cap data form a box. For brightness and greenness, there are only a few very large values outside the brackets, but there are several large values for Wetness. The coefficients for $Band_4$ in the linear combinations of greenness and brightness (defined on p.26) are 0.5585 and 0.7243 respectively which are higher than that for wetness. Brightness and greenness are dominated by $Band_4$. Since the boxplot of $Band_4$ has no extreme values outside the bracket, there are only a few outside the bracket of the boxplots for brightness and greenness. As to ratioing transformations, the boxplots and histograms of the 3 ratio data are

shown in Figure 3.12. The Ratio_1=Band_4/ Band_3 forms a box with one extreme value; while for Ratio_2 and Ratio_3, there are more extreme values outside the brackets.

3.5.2 Bivariate description

The univariate tools discussed earlier are limited for describing the distribution of individual variables. The Siuslaw data set is a multivariate data set. If we analyze the data set one variable at a time, it is important to examine the relationships and dependence between variables. First, we describe the relationship among the plot-level forest (response) variables, M, M_s, Lt, Lt_s, Tba, and Tba_s. The scatter plots of paired variables, Lt versus M, Tba versus Lt, Lt_s versus M_s, and Tba_s versus Lt_s are shown in Figure 3.13(a). The scatter plots do not reveal any apparent linear relationship among the variables except for Tba_s vs. Lt_s if species is ignored. The R^2 between Lt_s and Tba_s is 0.8547, but the R^2 between Lt and Tba is only 0.0057. Obviously, there is strong correlation between seedlings variables, Lt_s and Tba_s. This implies that the more live trees the larger seedlings total basal area. The scatter plot of Tba vs. Lt forms a band ranging from 100 to 300 for Lt values and from 35 to 60 for Tba values. Sample plots are dominated by a few large trees. For those plots, although the number of live trees is small, the Tba value is large. For further investigation, we study the relationship between the Lt and Tba values for Douglas-fir (Pseudotsuga Menziesii Glauca). Douglas-fir trees grow fairly rapidly; the average height of the tree is 150 to 200 feet although some trees grow to heights of 300 feet. When rubbed, they smell of camphor. To examine the relationship of variables for Douglas-fir, the scatter plots of paired data for Douglas-fir are displayed in Figure 3.13(b). The data are concentrated in the scatter plot of Lt versus Tba. The scatter plots for Lt versus M, Tba versus Lt, and Lt_s versus M_s show no apparent linear trend. For Tba_s

versus Lt_s, there is a similar linear trend as when species are ignored. The R^2 of Tba_s with Lt_s is 0.8194. As mentioned earlier, Douglas-fir trees grow rapidly, so the basal area of very small trees and seedlings have a linear relationship with the number of live very small trees and seedlings. There are only 290 plots for Douglas-fir, 23 less than if species designation is ignored. Due to the small sample size, we do not analyze by species. As shown in Table 3.3, all variables for Douglas-fir have smaller sample deviations than if species designation is ignored.

Figures 3.14-3.17 portray the scatter plots of forest variables versus the auxiliary data such as topographic and TM data. Unfortunately, we couldn't find any apparent linear trends between the forest response variables and the auxiliary data. From Figure 3.14, the R^2's are very small. Only elevation is related to M and Tba with $R = 0.1694$ ($R^2 = 0.0287$) and $R = 0.1338$ ($R^2 = 0.0179$), respectively. The higher the elevation, the higher the mortality. As shown in Figure 3.15, the R^2's of M with bands 1-5, and 7 range from 0.0087 to 0.0290 and all are negatively correlated; of all R's the largest one, -0.1703, belongs to $Band_7$. Since $Band_7$ is sensitive to vegetation moisture content, it is reasonable to have a negative relationship with mortality. There are no apparent linear trends for M_s and bands; the R^2's range from 0 to 0.0030. For Lt, the R^2's are small except for $Band_4$. As mentioned in Section 3.4, $Band_4$ is strongly reflected by active vegetation. $Band_4$ is related to Lt by the correlation $R = 0.1086$ as expected. For $Band_1$, $Band_2$, $Band_3$, and $Band_4$ data are slightly related to Lt_s. Band values with Tba range from -0.2173 to -0.3161 in R-values; $Band_5$ is the strongest related to Tba. For Tbas, there is no apparent linear trend with bands. The Tasseled Cap transformations didn't do better than those original band data for any variables except for Lt and Tba. The correlation between brightness (Cap_1) and Tba is $R = -0.3324$ which is more correlated than those for the original bands. Of all ratioing transformed band data, $Ratio_3$ is correlated to M with the value of

$R = 0.1038$, Ratio_1 is related to Lt by $R = -0.1728$, Ratio_1 is related to Tba by $R = -0.2170$, and there are no apparent linear trends for the seedlings variables.

3.5.3 Spatial description

In describing the spatial characteristics of a population, one is often interested in spatial features of the data set such as the location of extreme values, the overall trend, and the degree of continuity. None of the description tools discussed earlier can provide much information on the spatial features of the variables of interest. We present some effective tools for displaying spatial data, the simplest being a data posting. But for large digital values, it's not feasible to post the variable values at the location. When this situation occurs, the symbol map posting for all locations is used instead. Data posting is useful in identifying erroneous values and revealing erroneous data locations. For example, the grid sampling system was adopted in our data collection; the distance between one-unit-apart horizontal and vertical plot locations should equal the grid distance. From this, we identified several errors in locations of plots as shown in Figure 3.4.

The location of the plots with mortality/ha is shown in Figure 3.18. We reclassify the M values by the 4 quartiles as follows: red 1=(0, 21], blue 2=(21, 50], purple 3=(50, 86], and green 4=($>$ 86). Note that there are 11 pairs of adjacent green 4's in the adjacent horizontal directions and 10 pairs of adjacent green 4's in the adjacent vertical directions in the southern area; 3 pairs of adjacent green 4's in the adjacent horizontal directions and 4 pairs of adjacent green 4's in the adjacent vertical directions in the northern area; while in the corner direction there are 16 pairs of adjacent green 4's in adjacent corner directions in the southern area, 5 pairs of adjacent green 4's in adjacent corner directions in the northern area. We can see that the M data seems to have some spatial dependence among the adjacent horizontal, vertical, or corner neighbors. From Figure 3.5 and 3.6

discussed earlier, we note that there are some extreme values shown outside the brackets of the boxplot and long tails in the histogram. One might want to know where the extreme values are and if they are unusual with respect to their neighbors. This can be achieved by the data posting on a location map. In the spatial data posting map, we also display where the large values are, depicted in Figure 3.19. Some of them are close to each other. The M_s data is depicted in Figure 3.20. We see that there are many zeros consistent with those shown in the interquartiles of M_s in Table 3.3. For non-zero values, some of them are close to each other, some are surrounded by zeros.

The location of the plots with number of live trees/ha (Lt) is shown in Figure 3.22. The Lt data is classified as follow: red 1=(0, 180], blue 2=(180, 312], purple 3=(312, 577], and green 4=($>$ 577). There are about the same number of green 4's in the adjacent horizontal direction as those in the adjacent vertical direction. Also, there are some green 4's in the adjacent corner neighbors. The large Lt values are depicted in Figure 3.23. None of them are adjacent to each other.

Figure 3.24 shows the data for the Lt_s data. Note that there are many zeros, as described earlier, about 50%. Non-zero values are adjacent to each other frequently. In chapters 4 and 5, we will examine this further. In Figure 3.25 we have highlighted the 5 largest values from Figure 3.24. Four values, 2562, 1800, 2000, and 1450 are located in the northern part of the Forest and adjacent to each other. The other one, 2375, is located in the southern part of the Forest.

The colors shown in Figure 3.26 indicate the same color more frequently for Tba in adjacent corner neighbors than in adjacent horizontal and vertical neighbors. Therefore, there might be stronger spatial dependence effects on corner neighbors than with horizontal and vertical neighbors. In Figure 3.27 the large Tba values 106 and 101 also indicate that stronger spatial dependence exists with corner neighbors than with horizontal and vertical neighbors.

A posting of the Tba_s data is displayed in Figure 3.28 showing many zeros. In Figure 3.29 we depict only the 6 largest Tba_s values of Figure 3.28. Four values, 3.3, 2.6, 3.8, and 2.1 are located as adjacent horizontal, vertical, and corner neighbors.

We incorporate the spatial dependence of our data and try to capture the phenomenon of many zeros in our modeling. Several spatial models are discussed in chapters 4, 6, and 7.

3.5.4 Sample correlations of plot-level variables

In the previous subsection, we noted the potential candidate for spatial dependence in the data by looking at the location of the plots with associated values for the variables. This subsection examines the spatial dependence for each variable by computing sample correlations.

For a 3x3 lattice depicted in Figure 3.30, define the neighborhood of location i in the following ways:

First order neighborhood The west, east (horizontal neighbors), south, or north (vertical neighbors) of location i are the first order neighbors marked by the hollow dots in Figure 3.30. These are the plots at the 1.7-mile distance from plot i.

Second order neighborhood The second order neighborhood includes the four neighbors of the first order plus 4 corner neighbors (southeast, northwest, southwest, and northeast), i.e. plots within the $1.7\sqrt{2}$-mile distance of plot i, marked by the hollow and solid dots in Figure 3.30.

The sample correlations for horizontal neighbors, vertical neighbors, and corner neighbors are summarized in Table 3.4. For M data, the sample correlations among one-unit-apart horizontal and vertical neighbors are larger than for one-unit-apart

corner neighbors (0.1361 and 0.1711 vs. 0.0454). For M_s data, the one-unit-apart corner neighbors have larger sample correlation than horizontal neighbors and vertical neighbors; but, the corner sample correlation is only 0.0727. The sample correlations in horizontal, vertical, corner directional neighbors for Lt data are 0.0947, 0.1149, and 0.0429, respectively. Those sample correlations are not significantly different from zero. For Tba, the sample correlations among horizontal and vertical neighbors are -0.0147 and 0.0322 respectively; the sample correlations among corner neighbors is 0.1489. As mentioned earlier, we note this tendency from Figure 3.26 stronger spatial dependence effects on corner neighbors than those on horizontal and vertical neighbors. There are positive sample correlations for Lt_s and Tba_s among horizontal, vertical, and corner neighbors. For Lt_s, the sample correlations for horizontal, vertical, and corner neighbors, are respectively 0.2300, 0.3022, 0.2101 and for Tba_s, they are 0.2243, 0.3176, and 0.1713, respectively. This is consistent with the results shown in Figure 3.24 and 3.28; for non-zero values, there are more pairs of close values with adjacent vertical neighbors than with horizontal and corner neighbors, and it is also true for zero values. The adjacent zeros make more contribution than adjacent non-zeros to the large sample correlations. If we divide the data into zeros and non-zeros, then the sample correlation drops abruptly.

The corner neighbors ($h = 1.7\sqrt{2}$ mile) in the second order neighborhood are farther apart than the one-unit horizontal and one-unit vertical neighbors ($h = 1.7$ mile). For variables except for M_s and Tba, the sample correlations for horizontal and vertical neighbors are larger than the sample correlation for corner neighbors. To see if there is any distance-related sample correlation in the data, we computed some sample correlations for "greater" lags, e.g. sample correlations for 2-unit horizontal and 2-unit vertical neighbors ($h = 3.4$ mile, marked by $\triangle$ in Figure 3.30), and sample correlation for 2-unit corner neighbors ($h = 3.4\sqrt{2}$, marked by

▲ in Figure 3.30). From Table 3.4, we note that most of the sample correlations for "greater" lags are much smaller than those for one-unit horizontal, vertical, and corner neighbors. But the sample correlation for two-unit vertical neighbors in the Lt data doesn't become much smaller as the lag-distance increases ($r_{2v} = 0.1151$ vs. $r_v = 0.1190$). This implies that the sample correlations in the vertical direction decay slowly as the lag-distance increases. It caught our attention that in the M_s data the sample correlation for two-unit vertical neighbors is larger than the one for one-unit vertical neighbors ($r_{2v} = 0.1505$ vs. r_v=0.0087). There are about 86% zeros in the M_s data. The greater the lag-distance is, the more zeros are included. This can lead to overestimating the parameters of spatial dependence. The scatter plot of the plot-level attributes for adjacent neighbors at different lag-distance was portrayed in Figure 3.31.

In Chapters 4, 6, and 7, we present several models incorporating the spatial features and dealing with the many zeros of the data for some variables.

3.5.5 Sample correlations of subplot-level variables

So far we only discussed the sample correlations for plot-level variables. This subsection investigates the sample correlations within the subplot-level variables in the five stake position areas, as diagramed in Figures 3.2 and 3.3. The minimum distance between subplot areas is 40.8 m which is much closer than that between plot locations (1.7 mile=2736 m). Stake positions are 40.8 m apart from the center of stake position 1, so the distance is close to the radius of a 1-ha plot, 56.42 m. Hence, the sample correlations between subplots are expected to be larger than those between plots.

Table 3.5 lists the sample correlations of subplot variables. We will do further testing in the later chapters on the significance of spatial dependence parameters. For M, the one-unit-apart (40.8 m) sample correlation in the vertical direction is

about 0.1 larger than the one in the horizontal direction, and the sample correlation between two-unit-apart (81.6 m) vertical neighbors is about 0.1 larger than the one between two-unit-apart (81.6 m) horizontal neighbors, too. The sample correlations between subplots do not decrease as the lag distance increases, but increases to 0.3835 at lag 81.6 m for horizontal neighbors and 0.4807 at lag 81.6 m for vertical neighbors. For M, the sample correlation did drop down as the lag-distance increases. The sample correlations for M_s at one-unit-apart (2736 m) plots in the horizontal and vertical directions are about 0.04 and 0.03 (shown in Table 3.4), respectively. But for the sample correlations between subplots, the 40.8-meter-apart horizontal sample correlation is 0.49 and the vertical one is 0.32; those are much larger than the sample correlations between plots. The sample correlations for M exhibit different behavior for plots than those for subplots; e.g. $r_v = 0.1711$ and $r_{2v} = -0.0104$ for plots as shown in Table 3.4; whereas $r'_v = 0.1711$ and $r'_{2v} = -0.0104$ for plots as shown in Table 3.5. The sample correlations of Lt between subplots in the different directions ranged from 0.43 to 0.57; the largest values are the one-unit-apart (40.8 m) sample correlations. The sample correlations between subplots for Lt decay slowly as the lag-distance increases. We note that at 81.6-meter-apart the sample correlations in the vertical and horizontal directions are still 0.43 and 0.55. For Tba on subplots, the sample correlations decay slowly as the lag-distance increases, too; the largest of those sample correlations is about 0.61 which is only 0.13 larger than the smallest. For Lt_s on subplots, the sample correlations are not only slightly larger than those for plot Lt_s and the 2-unit-apart sample correlation in the vertical direction is half as large as the 1-unit-apart sample correlations in the vertical direction and the sample correlations of subplot Lt_s drop much more slowly in the horizontal direction than those sample correlations of plot Lt_s did. The sample correlations for Tba_s on subplots are smaller than those for Tba_s on plots at one-unit-apart distance.

Both sample correlations for Tba_s on subplots and Tba_s on plots decrease as the lag-distance increases. The scatter plot of the subplot-level attributes for adjacent neighbors at different lag-distance is shown in Figure 3.32.

In general, the sample correlations of subplot variables are consistently larger than those of plot variables as expected.

3.6 Data Summary

Plot attributes of trees, M, Lt, and Tba do not have many zeros in the data, but plot attributes of seedlings do. For M_s, there are about 86% zeros. There are not many plots having dead seedlings. Of all plots having live seedlings, there are only 42% plots with very small trees or seedlings of DBH of 1" or larger. For those plot attributes of trees, Tba is the most bell-shaped and with only 1% zeros. The other tree attributes are not bell-shaped, so transformations are considered in Chapter 4. In fact, transformations to normality do not change the nature of many zeros in the data, so that a mixture model which addresses many zeros is studied in Chapter 6.

As mentioned in the previous section, sample correlations among subplots are much larger than those among plots. The one-unit-horizontal distance between subplots is 40.8 m which is much closer than the minimum distance between plots (2736 m). That implies that the distance-based correlation function is a potential candidate for our data to make predictions for non-sampled locations. Among M, Lt, and Tba, the smallest coefficient of variation belongs to Tba and so it should be easier to predict than M and Lt.

The reliability of prediction using spatial regression models depends on the availability of good auxiliary information. As we noted earlier, there is weak relationship between auxiliary data and forest data without considering spatial features. While considering the spatial feature of auxiliary data and forest data,

the correlation is much stronger. So estimation including area-specific auxiliary information in a regression model is worth considering.

In summary, considerable good quality data is available. There appears to be little correlation between plot information even on adjacent plots and with ancillary information. This might be due to the large lag-distance (1.7 mile=2736 m) we have for plot data since we note that the sample correlations of subplot based data are large. And this is not surprising since intuition tells us that the shorter lag-distance (40.8 m) should have a higher correlation than the larger lag-distance (2736 m). The ancillary information is not as good as we expected; there is only weak correlation between the ancillary data we used and the variables of interest. One can expect predictions to be satisfactory if good auxiliary information is available and the model is correctly specified. The reason we use ancillary data such as aspect, elevation, slope, and TM bands is because they are available for non-sampled plots and non-sampled subplots and for prediction purpose we only can use the information available for non-sampled plots.

In model building, we consider simple spatial models in Chapter 4, spatial zero-inflated models in Chapter 6, and spatial models with auxiliary data in Chapter 7. Chapters 5 and 8 respectively introduce the predictions based on plot information and the predictions based on plot information and auxiliary information.

Figure 3.1: Study Area: Siuslaw National Forest, Oregon

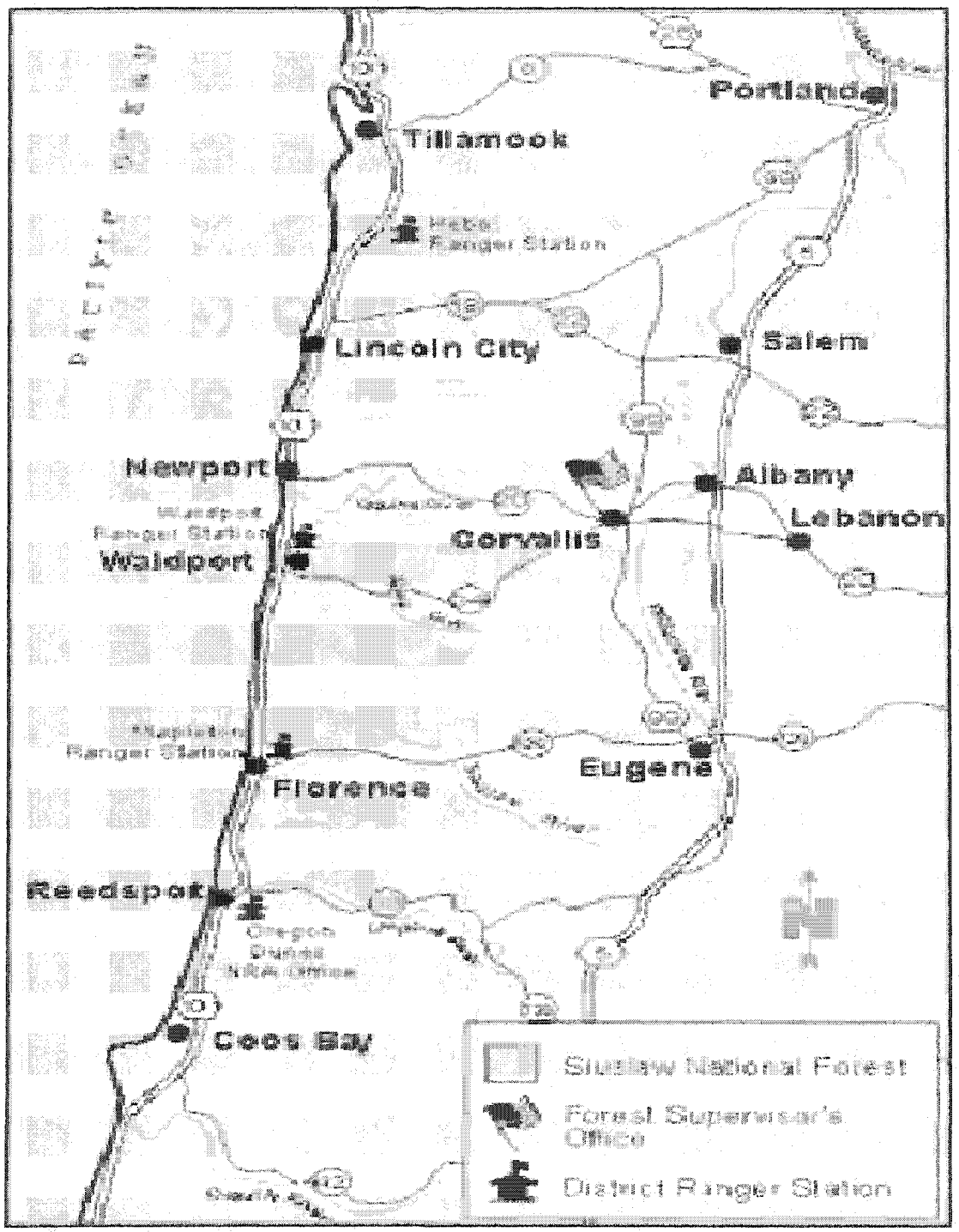


Table 3.1: Installations for 1-ha Plots in the Siuslaw N.F.

Installation Year	Number of Force Account Installations	Number of Contract Installations	Total Sample Units Installed
1993	2	0	2
1994	0	35	35
1995	0	148	148
1996	8	84	92
1997	0	43	43
Total	10	310	320

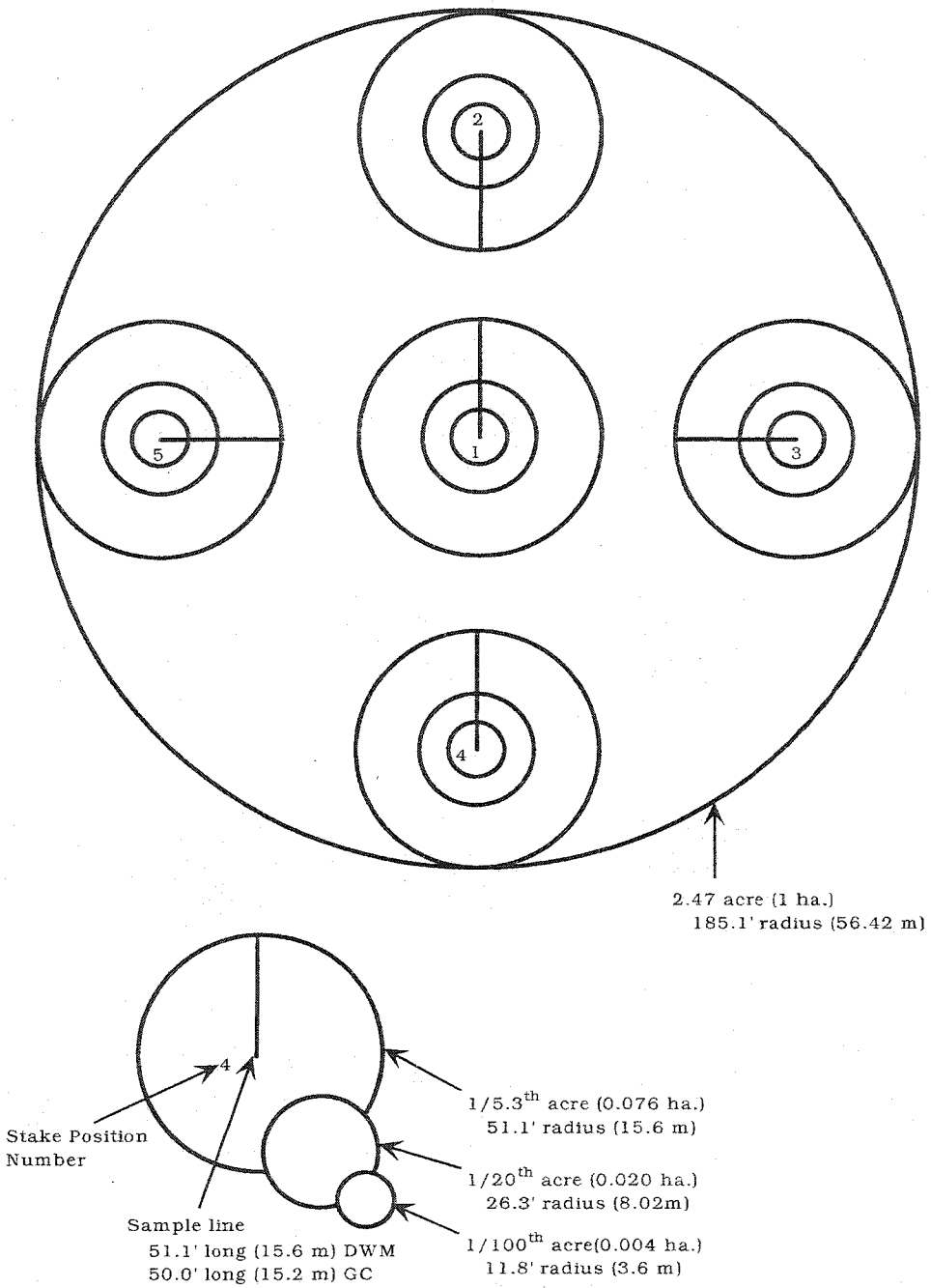
(a). Sample Units (1-ha Plots) Installed by Year and Method

Plot Number	Stake Positions				
	1	2	3	4	5
1092040	Full	Full	Full	Full	No (A)
2074046	Full	Full	Full	No (A)	Full
2076046	Partial (O)	No (O)	Partial (O)	Full	Partial (O)
2080030	No (A)	No (A)	No (A)	No (A)	Full
2082038	Full	Full	Full	Full	No (O)
2082046	Partial (O)	Full	Partial (O)	No (O)	Partial (O)
2084050	Full	Full	Full	Full	No (A)
2085032	No (A)	No (A)	Full	No (A)	No (A)
2103038	Full	No (O)	Full	Full	Full
2108050	Full	Full	Full	No (O)	Full
2125072	Partial	No	Full	Full	Full
2136050	Partial (O)	Partial (O)	Full	Partial (O)	No (O)
2139062	Full	Full	Full	Full	No (O)
2141056	Partial (O)	Full	Full	No (O)	Full

(b). All plots are not fully installed

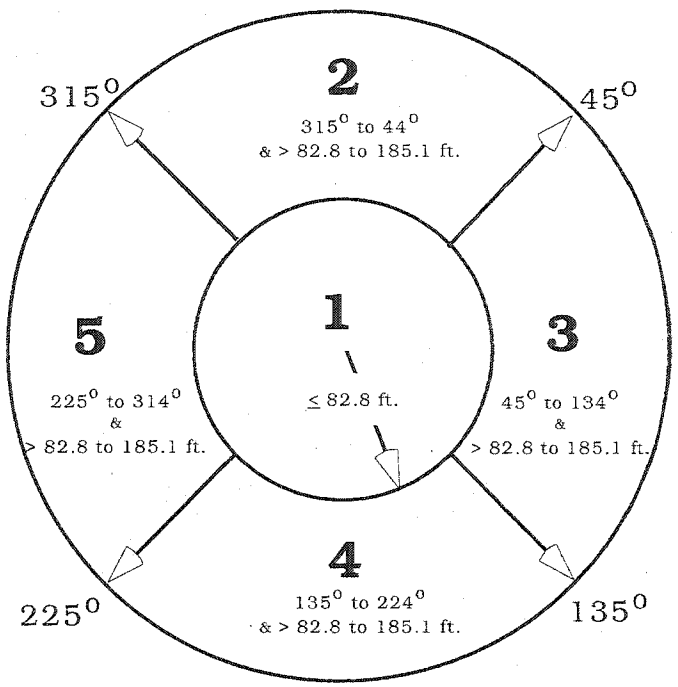
Note that "O" indicates that the stake position area was partially installed or was not installed due to ownership and "A" indicates the stake position area is not installed due to inaccessibility.

Figure 3.2: Diagram of Primary Sample Unit (PSU) Design[1]



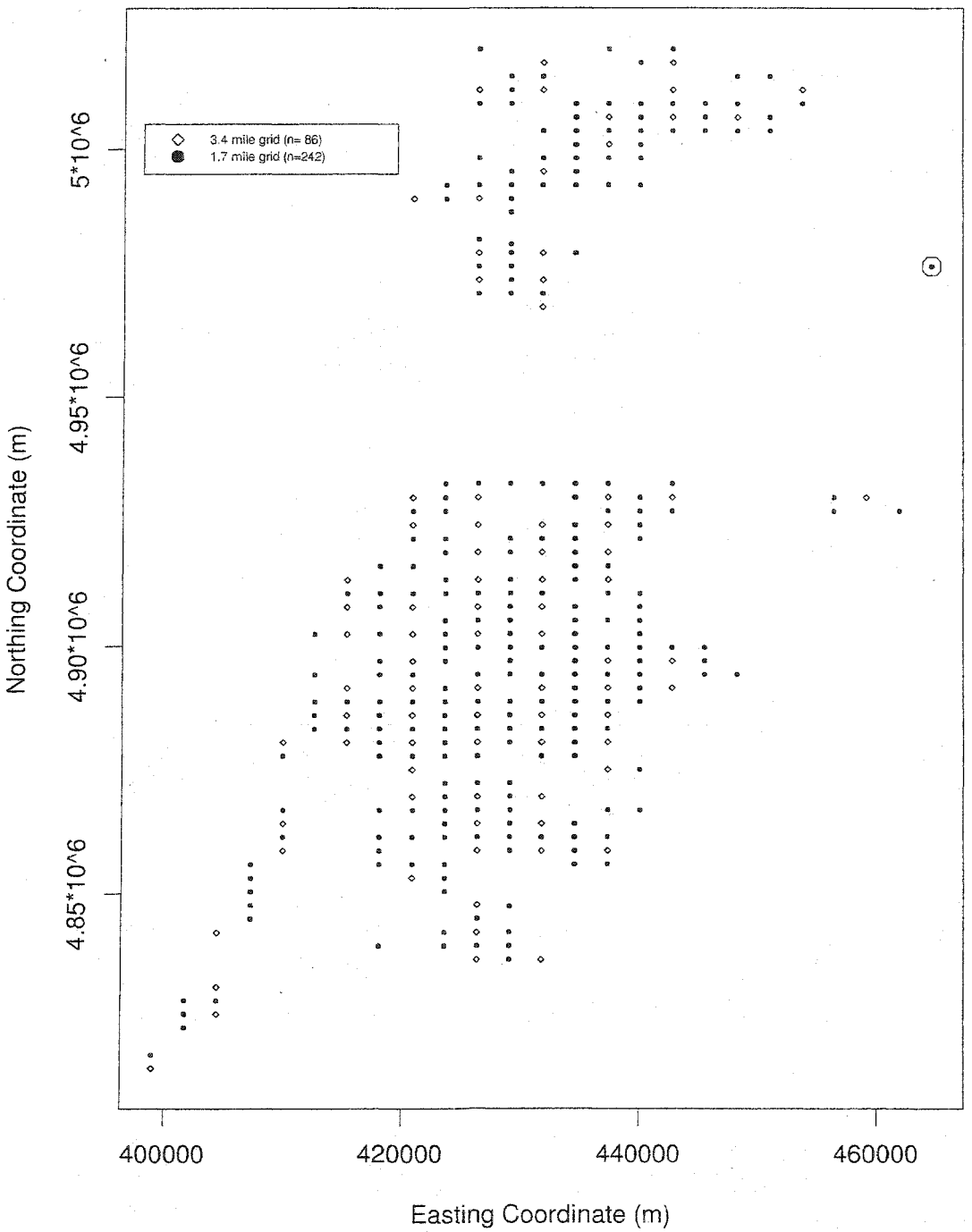
[1] 5 stake position areas (1-5) with concentric subplot sizes (0.004 ha., 0.02 ha., and 0.076 ha.), and sample lines

Figure 3.3: Stake Position Numbers, with Corresponding Areas and Boundary Limits[1]



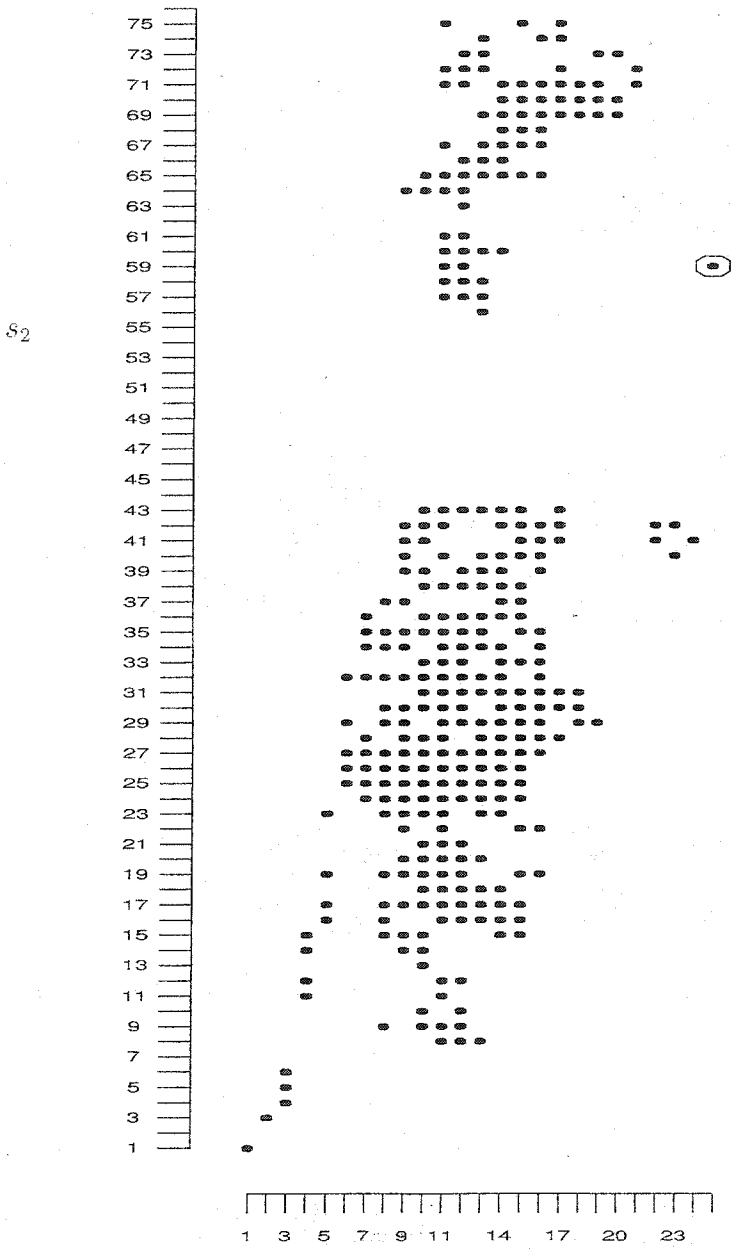
[1] All concentric subplots ((0.004 ha., 0.02 ha., and 0.076 ha.) are contained within one of the five areas corresponding to a stake position number (1-5).

Figure 3.4: Location of CVS plots (N=328) in the Siuslaw N. F.[1]



[1] ⊙ indicates the isolated plot which is not near the other plots in the Siuslaw National Forest.

Figure 3.5: Rescaled Integer-indexed Location of 1-ha plots in the Siuslaw N. F.[1]

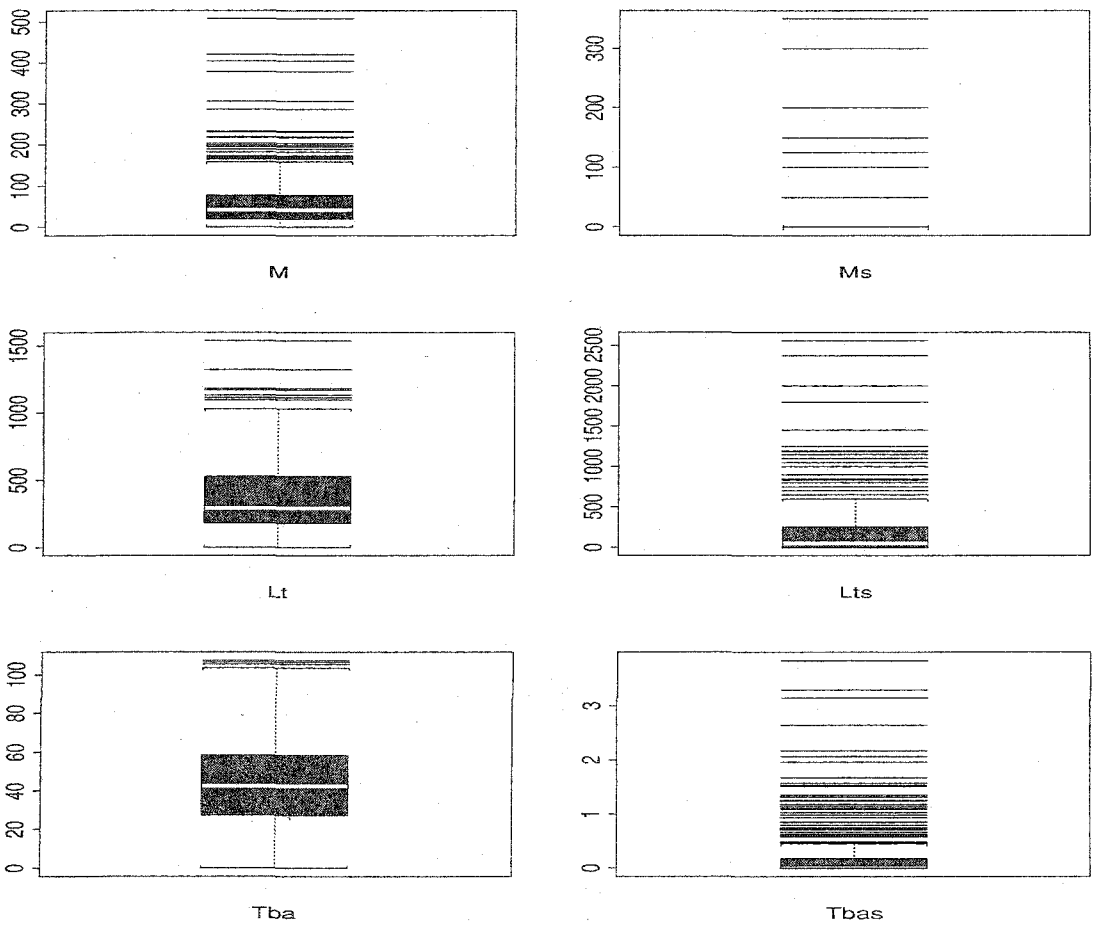


[1] Spatial locations were rescaled by 1.7 mile and the point furthest west and south is the starting point (1,1); the Siuslaw Forest is a 25x 75 rectangular lattice with holes.

Table 3.2: Thematic Mapper Spectral Bands (See Lillesand and Kiefer (2000))

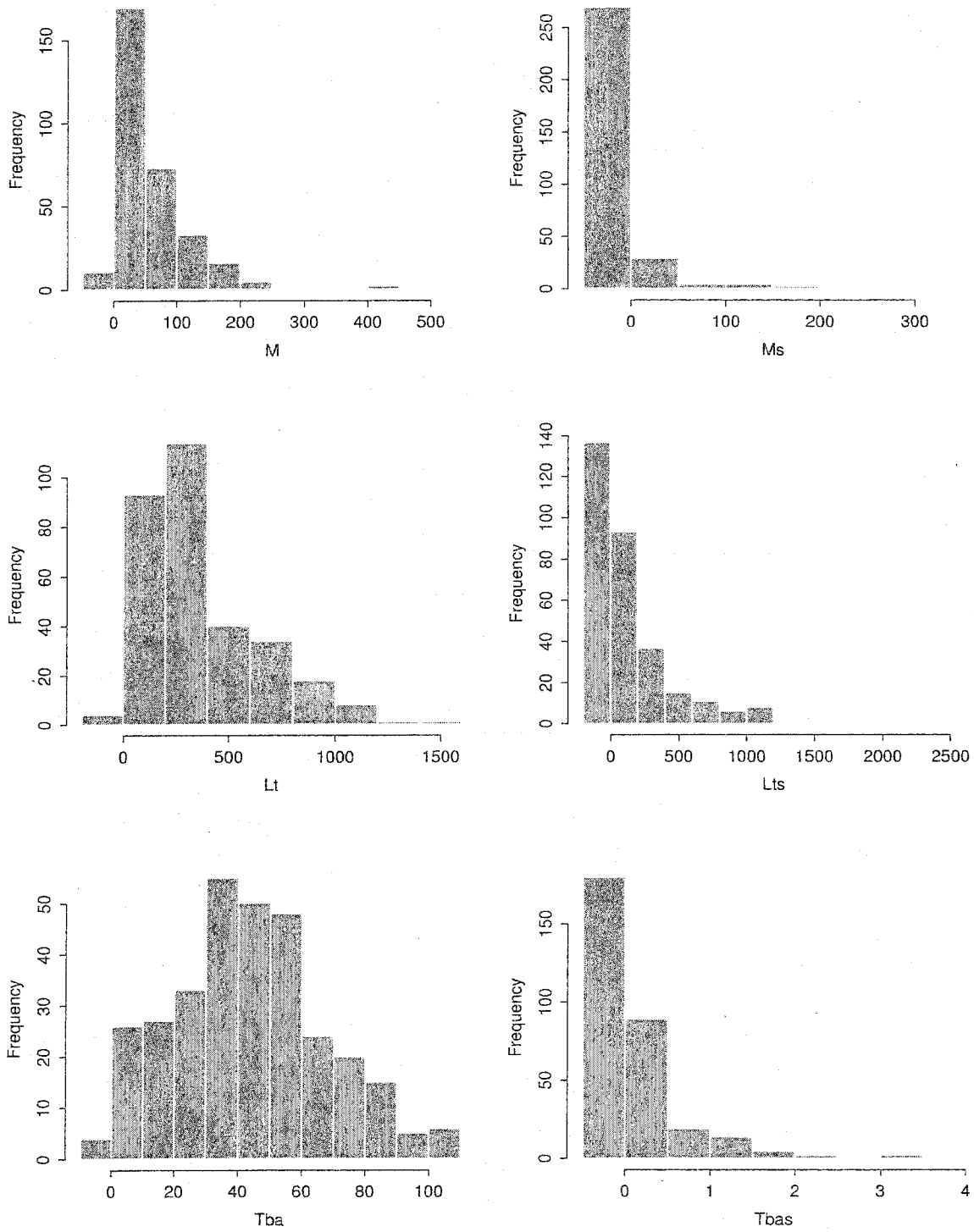
Band	Spectral Name	Spatial Resolution	Wavelength (μm)
1	blue	30 meters	0.45 - 0.52
2	green	30 meters	0.52 - 0.60
3	red	30 meters	0.63 - 0.69
4	near infrared (NIR)	30 meters	0.76 - 0.90
5	short mid-infrared (MIR)	30 meters	1.55 - 1.75
6	thermal infrared (TIR)	120 meters	10.40 - 12.50
7	long mid-infrared (MIR)	30 meters	2.08 - 2.35

Figure 3.6: Boxplots for M, M_s, Lt, Lt_s, Tba, and Tba_s [1]



[1] the y-axis shows data values

Figure 3.7: Histograms for M, M_s, Lt, Lt_s, Tba, and Tba_s [1]



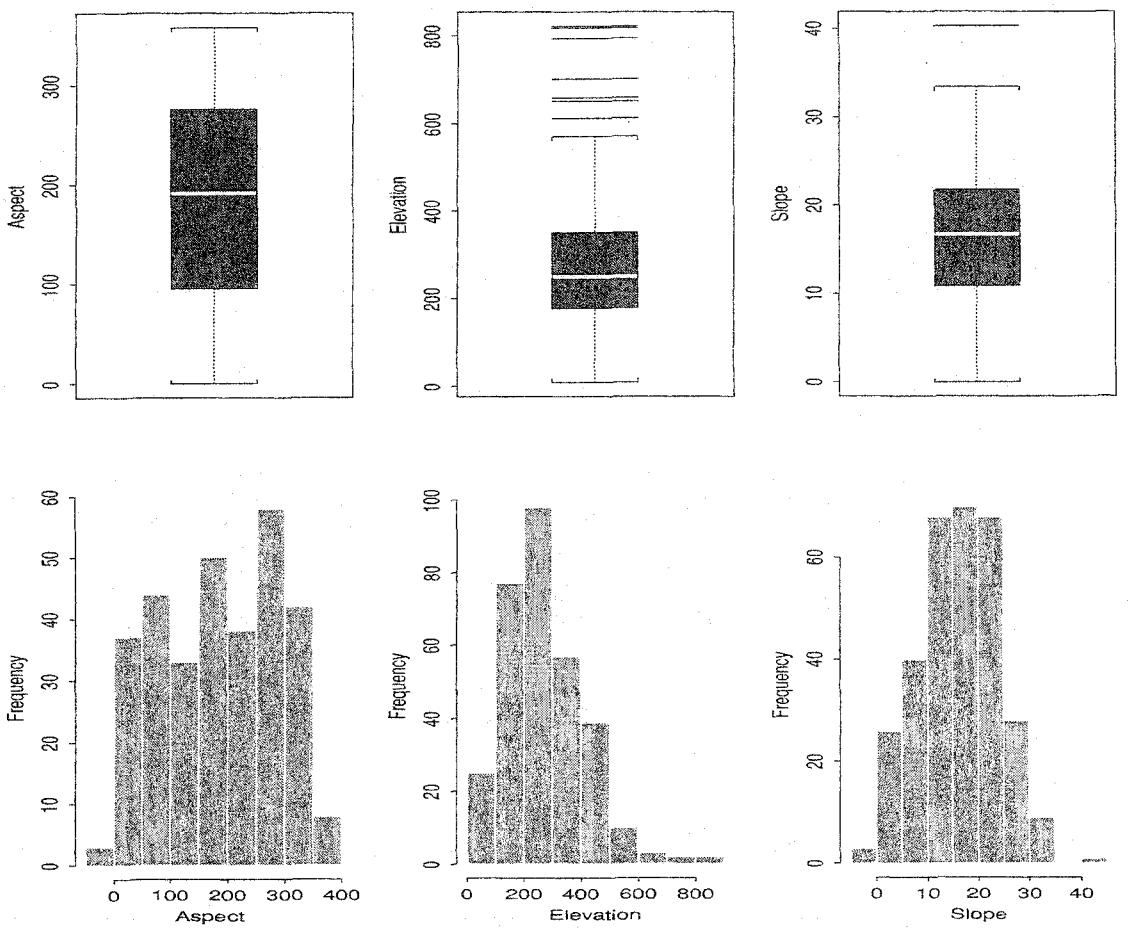
[1] the first bin is the counts of data values <= 0, i.e. the counts of zeros

Table 3.3: Descriptive statistics of the data

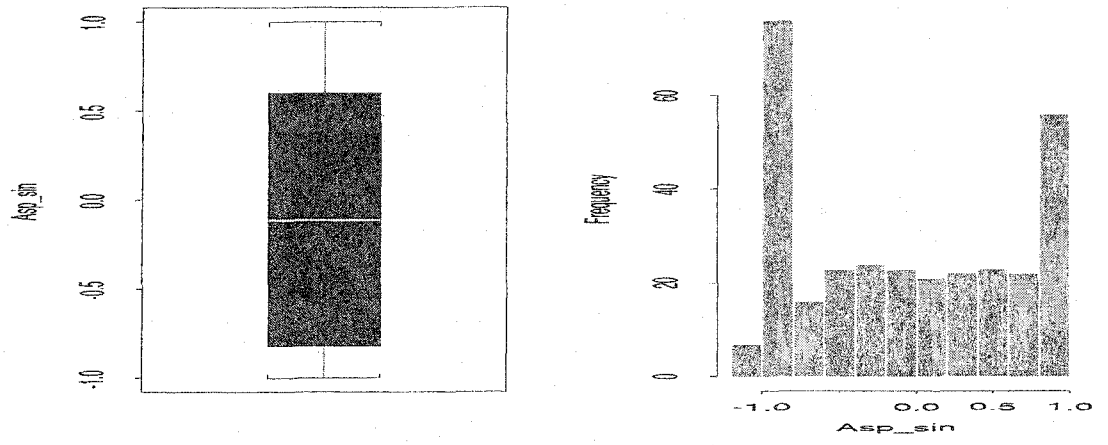
Variables	Min.	Q_1	Median	Mean	Q_3	Max.	S.D.
	N=313 plots						
Aspect (°)	0.00	95.08	191.30	186.20	276.00	358.21	105.73
Elev (m)	9.00	178.00	250.00	272.30	351.00	822.00	139.29
Slope(°)	0.00	10.88	16.74	16.20	21.80	40.33	7.69
Asp_{sin}	-1.00	-0.82	-0.12	-0.07	0.60	1.00	0.70
Asp_{cos}	-1.00	-0.71	0.00	0.02	0.74	1.00	0.71
Asp_{RC}	0.00	0.14	0.52	0.51	0.86	1.00	0.36
Asp_S	0.00	0.08	0.45	0.46	0.80	1.00	0.12
$Band_1$	55.00	61.00	62.00	63.06	65.00	108.00	4.41
$Band_2$	17.00	21.00	22.00	22.76	24.00	49.00	3.18
$Band_3$	12.00	16.00	17.00	17.84	19.00	63.00	4.09
$Band_4$	34.00	63.00	83.00	86.39	104.00	160.00	27.16
$Band_5$	21.00	32.00	41.50	45.05	54.00	111.00	16.53
$Band_7$	4.00	9.00	10.00	11.96	15.00	55.00	6.01
Cap_1	62.44	87.31	103.60	107.30	123.30	182.90	25.42
Cap_2	-25.00	16.14	28.78	31.00	42.13	83.23	18.80
Cap_3	-35.25	7.73	12.71	11.77	17.29	30.87	8.56
$Ratio_1$	1.14	3.93	4.79	4.86	5.80	8.42	1.24
$Ratio_2$	0.32	0.44	0.51	0.53	0.59	1.81	0.15
$Ratio_3$	1.04	3.46	3.89	3.97	4.38	7.80	0.84
	N=313 plots for all species						
M ($\#/ha$)	0.00	20.53	42.58	62.75	77.32	507.3	67.46
M_s ($\#/ha$)	0.00	0.00	0.00	12.54	0.00	350.00	40.49
Lt ($\#/ha$)	0.00	176.60	290.00	367.60	527.90	1538.42	268.93
Lt_s ($\#/ha$)	0.00	0.00	50.00	200.80	250.00	2562.50	359.72
Tba (m^2/ha)	0.00	27.37	42.46	43.50	58.41	107.29	24.34
Tba_s (m^2/ha)	0.00	0.00	0.00	0.23	0.18	3.84	0.52
	N=290 plots for Douglas-fir						
M	0.00	12.70	27.24	43.67	57.67	335.80	48.90
M_s	0.00	0.00	0.00	3.78	0.00	100.00	14.91
Lt	0.00	54.87	128.60	207.20	295.00	888.40	202.30
Lt_s	0.00	0.00	0.00	73.78	62.50	1000.00	158.68
Tba	0.00	14.72	28.55	32.90	45.63	103.40	22.95
Tba_s	0.00	0.00	0.00	0.09	0.00	1.97	0.27

Note: Tasseled Cap transformations of the 6 bands – Cap_1: brightness, Cap_2: greenness, Cap_3: wetness; Vegetation index – $Ratio_1$: $Band_4$ / $Band_3$. $Ratio_2$: $Band_5$ / $Band_4$, $Ratio_3$: $Band_5$ / $Band_7$.

Figure 3.8: Boxplots and histograms for the topographic data

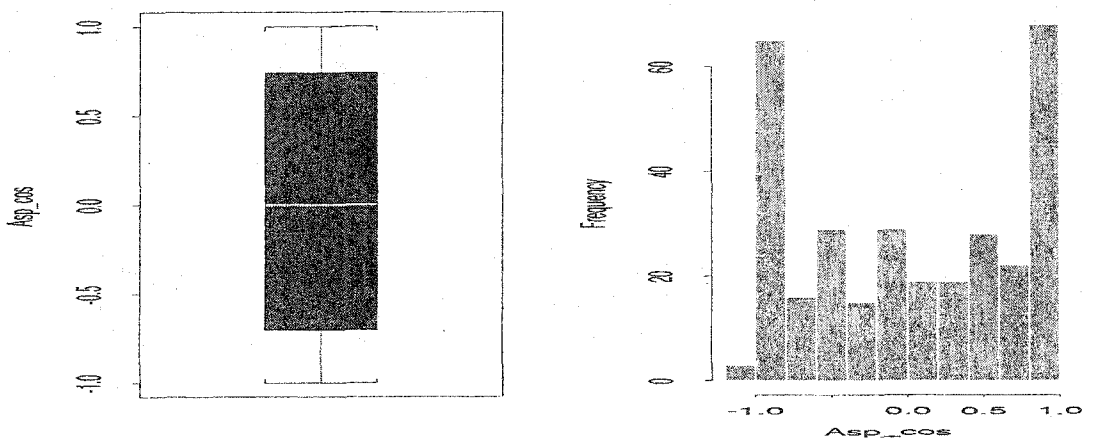


(a). the original aspect, elevation, and slope data

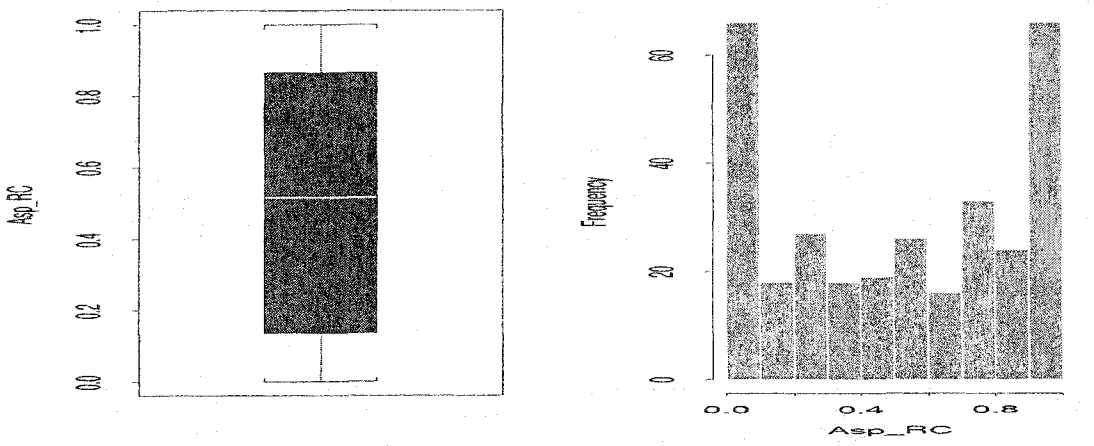


(b). sine-transformed aspect data

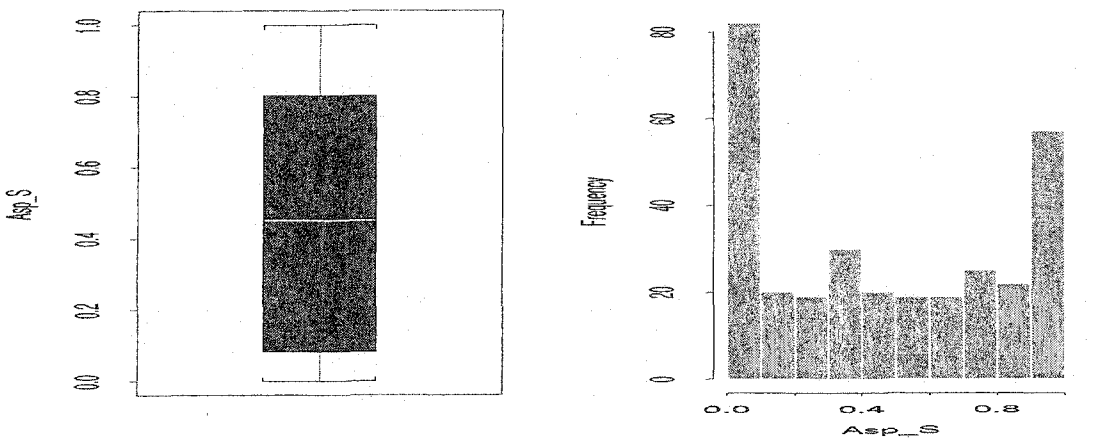
(Continued)



(c). cosine-transformed aspect data

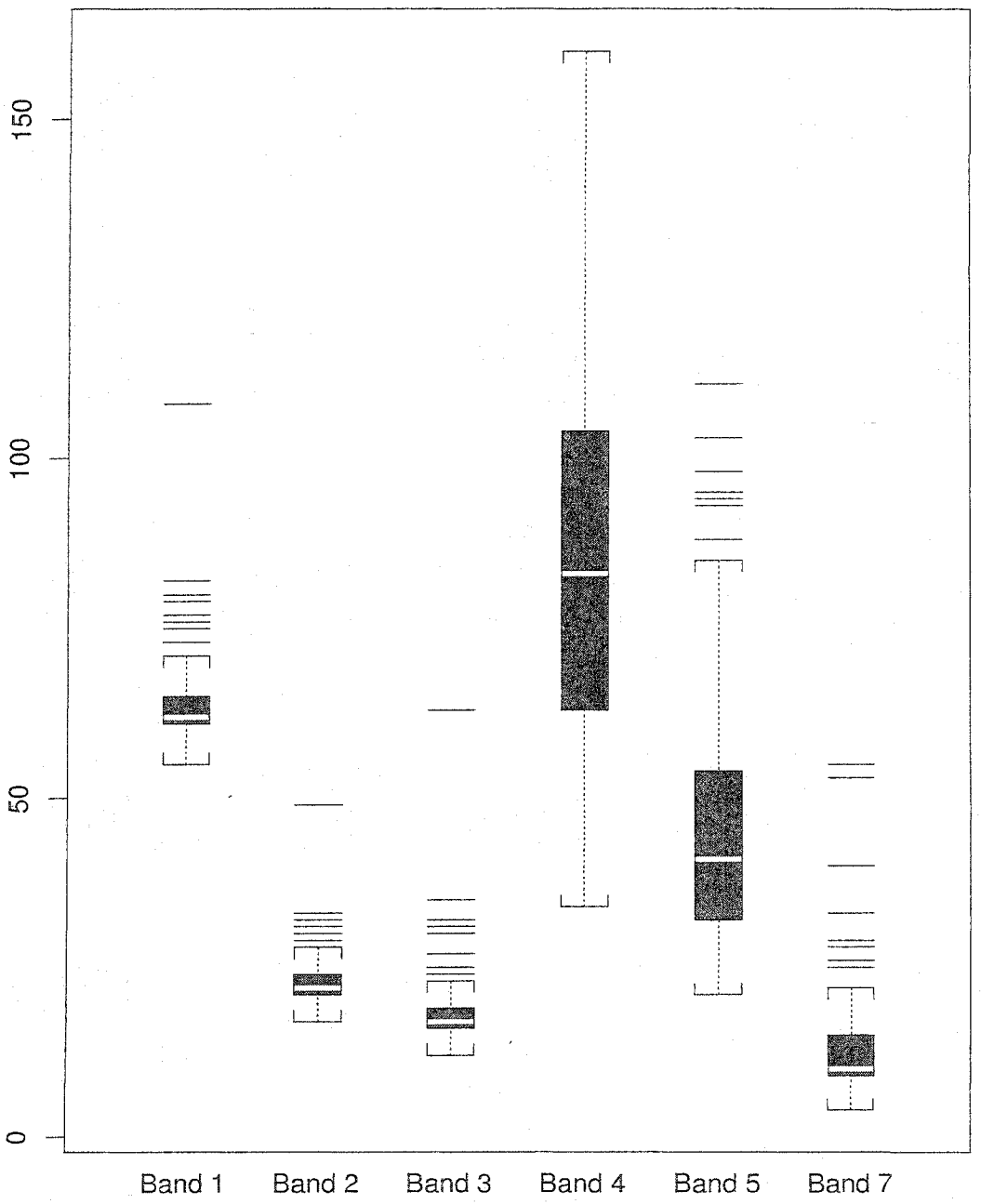


(d). Robert and Cooks cosine-transformed aspect data



(e). Robert and Cooks transformed aspect data, replacing cosine by sine

Figure 3.9: Boxplots for the 6 bands[1]



[1] the y-axis is the data values

Figure 3.10: Histograms for the 6 bands

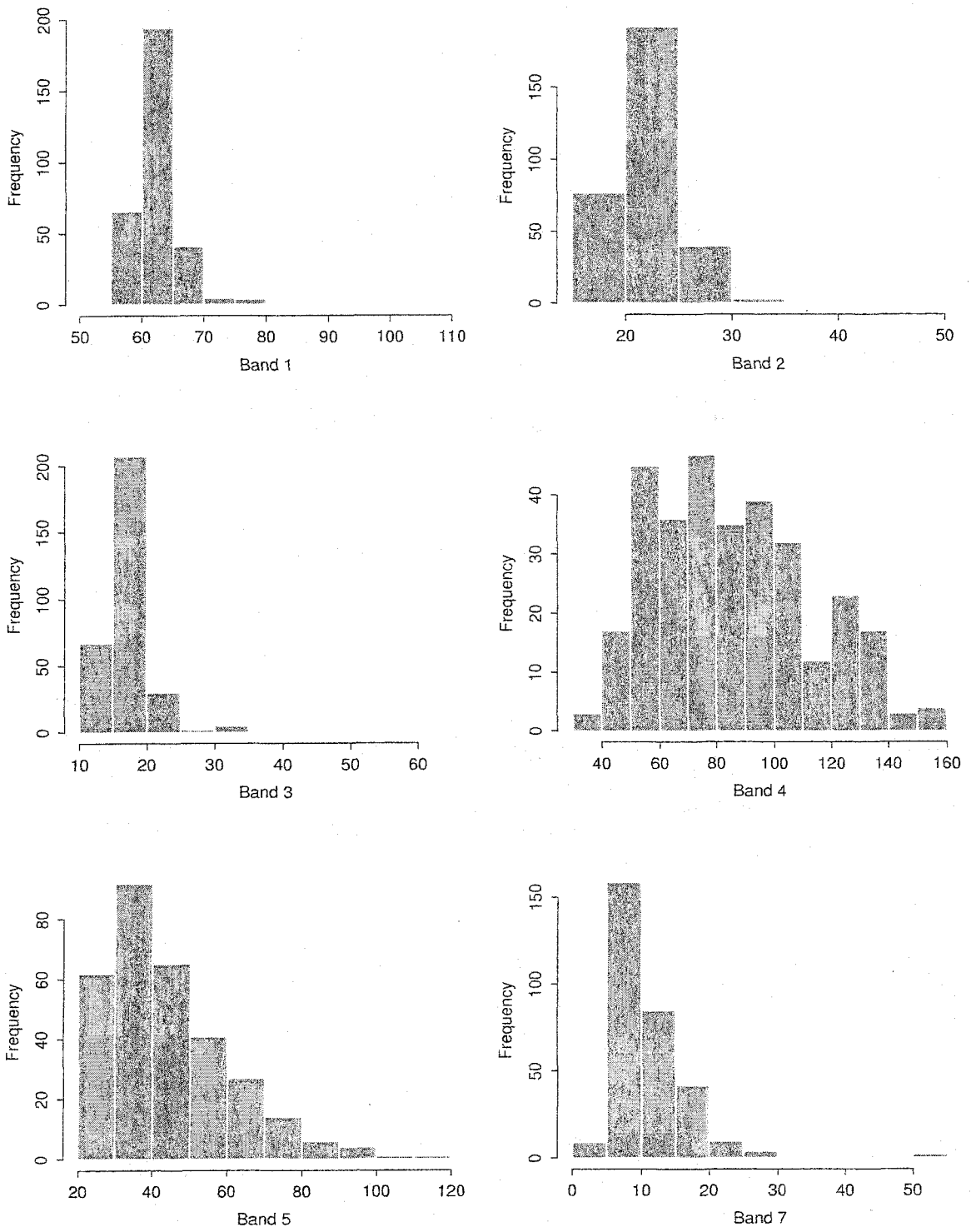


Figure 3.11: Boxplots and histograms for the Tasseled Cap transformation of the data from the 6 bands

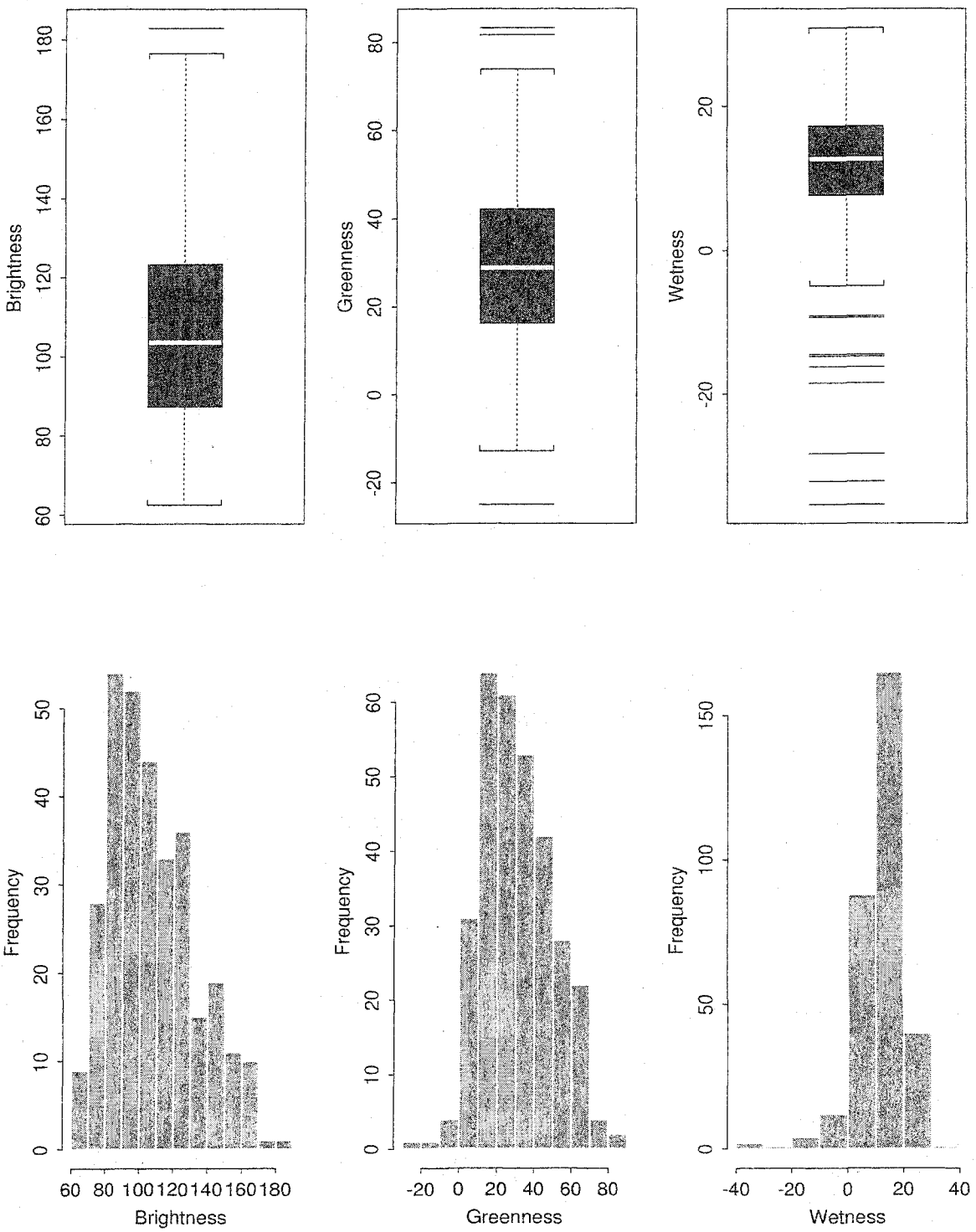


Figure 3.12: Boxplots and histograms for the ratio transformation of the data from the 6 bands

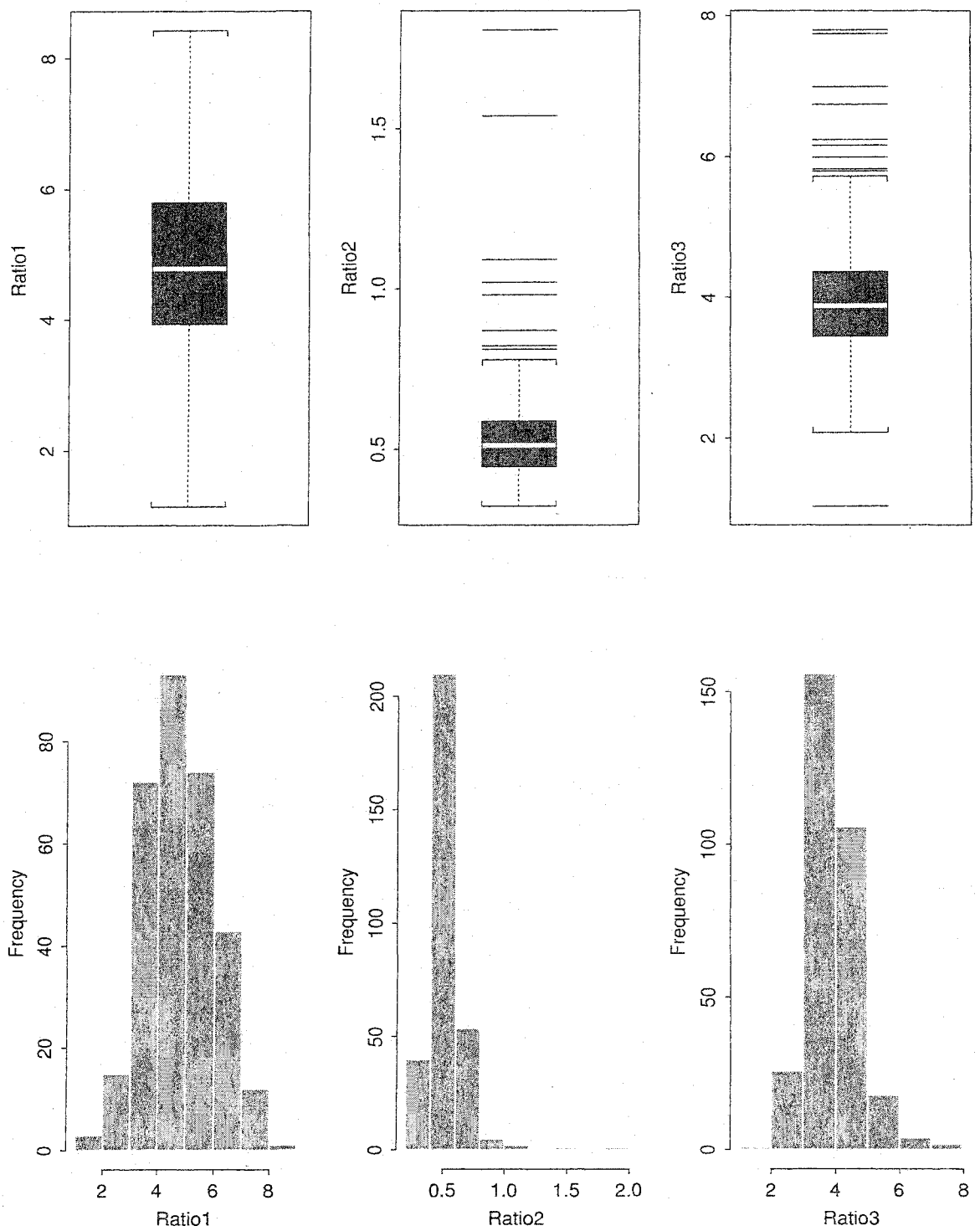
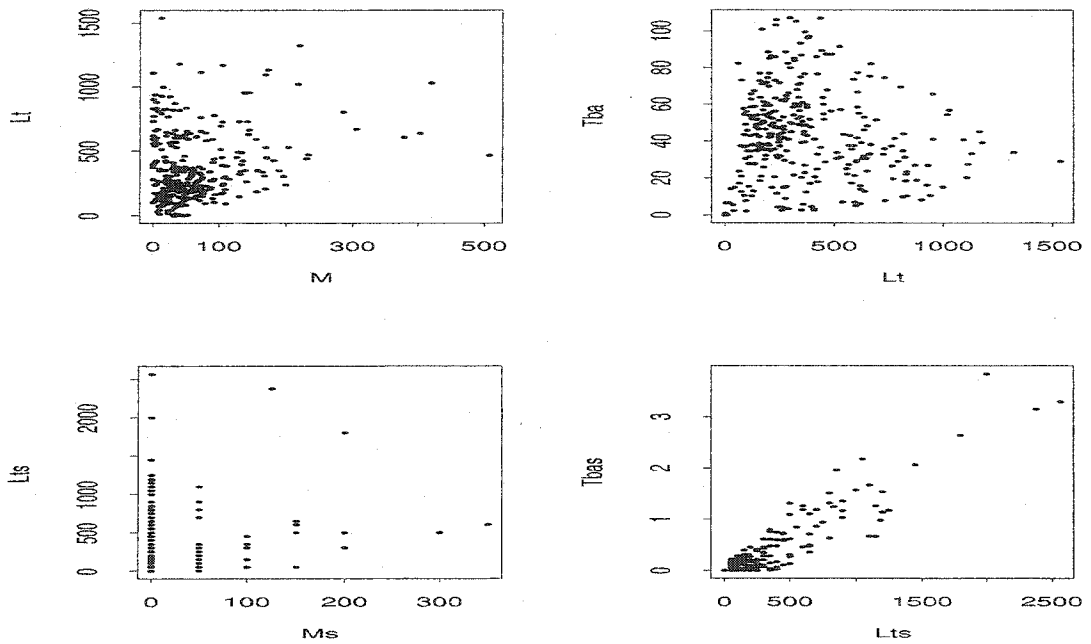
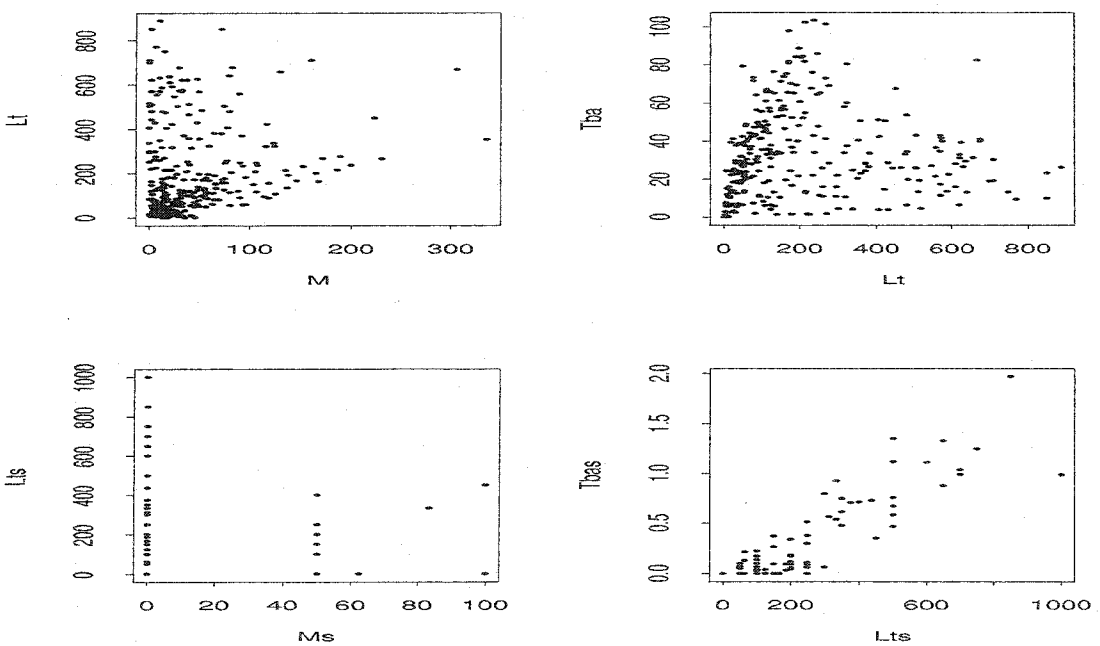


Figure 3.13: Scatter plots of M versus Lt, Lt versus Tba, M_s versus Lt_s, and Lt_s versus Tba_s.



(a) the forest variables ignoring species[1]



(b) the forest variables for Douglas-fir[2]

[1] r^2 for M with Lt is 0.0492, for Lt with Tba is 0.0057, for M_s with Lt_s is 0.0707, for Lt_s with Tba_s is 0.8547. [2] r^2 for M with Lt is 0.0214, for Lt with Tba is 0.0015, for M_s with Lt_s is 0.0069, and for Lt_s with Tba_s is 0.8191.

Figure 3.14: Scatter plots of all 6 variables versus the topographic data, R^2 in parenthesis

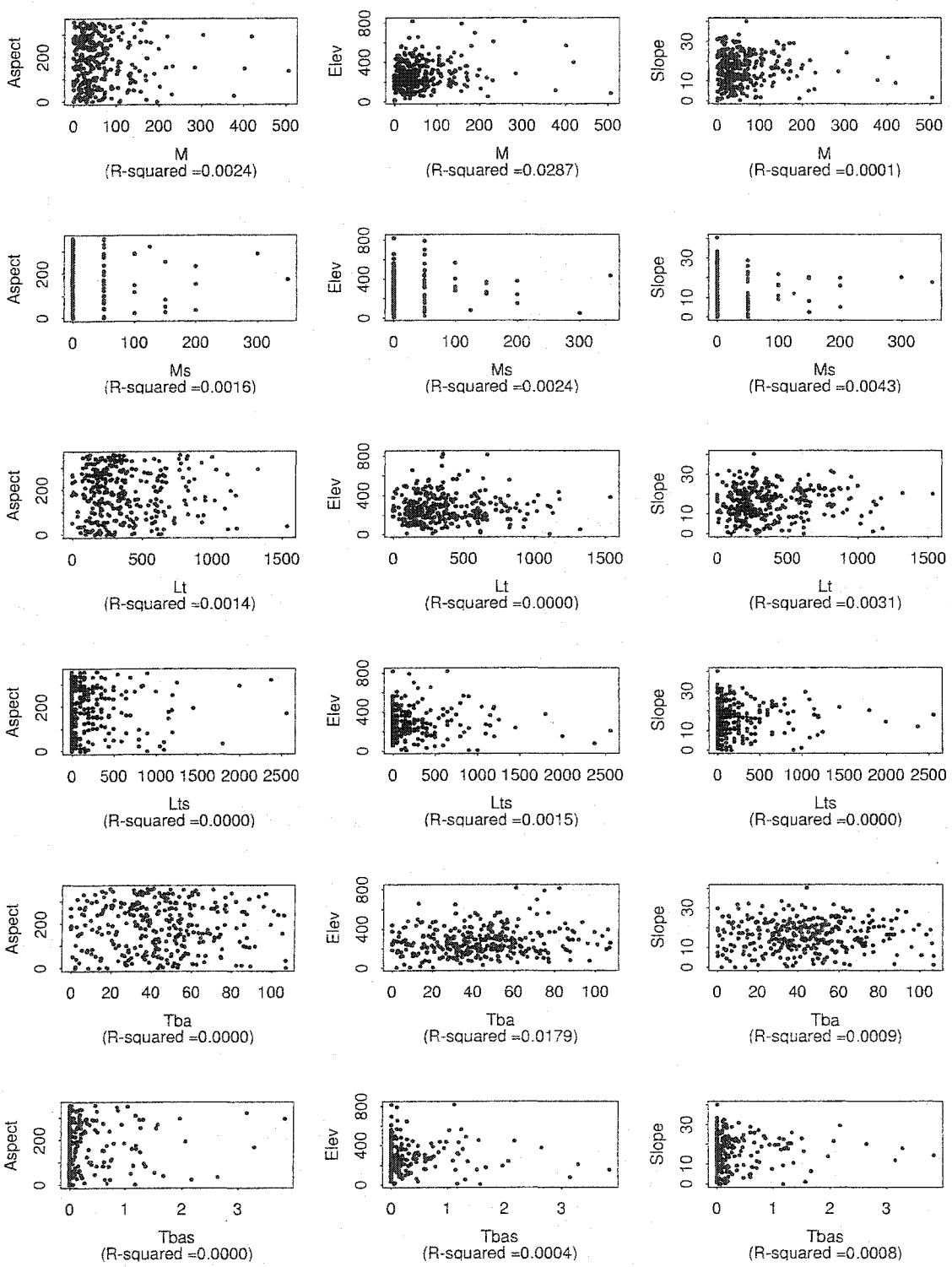


Figure 3.15: Scatter plots of all 6 variables versus all 6 band data, R^2 in parenthesis

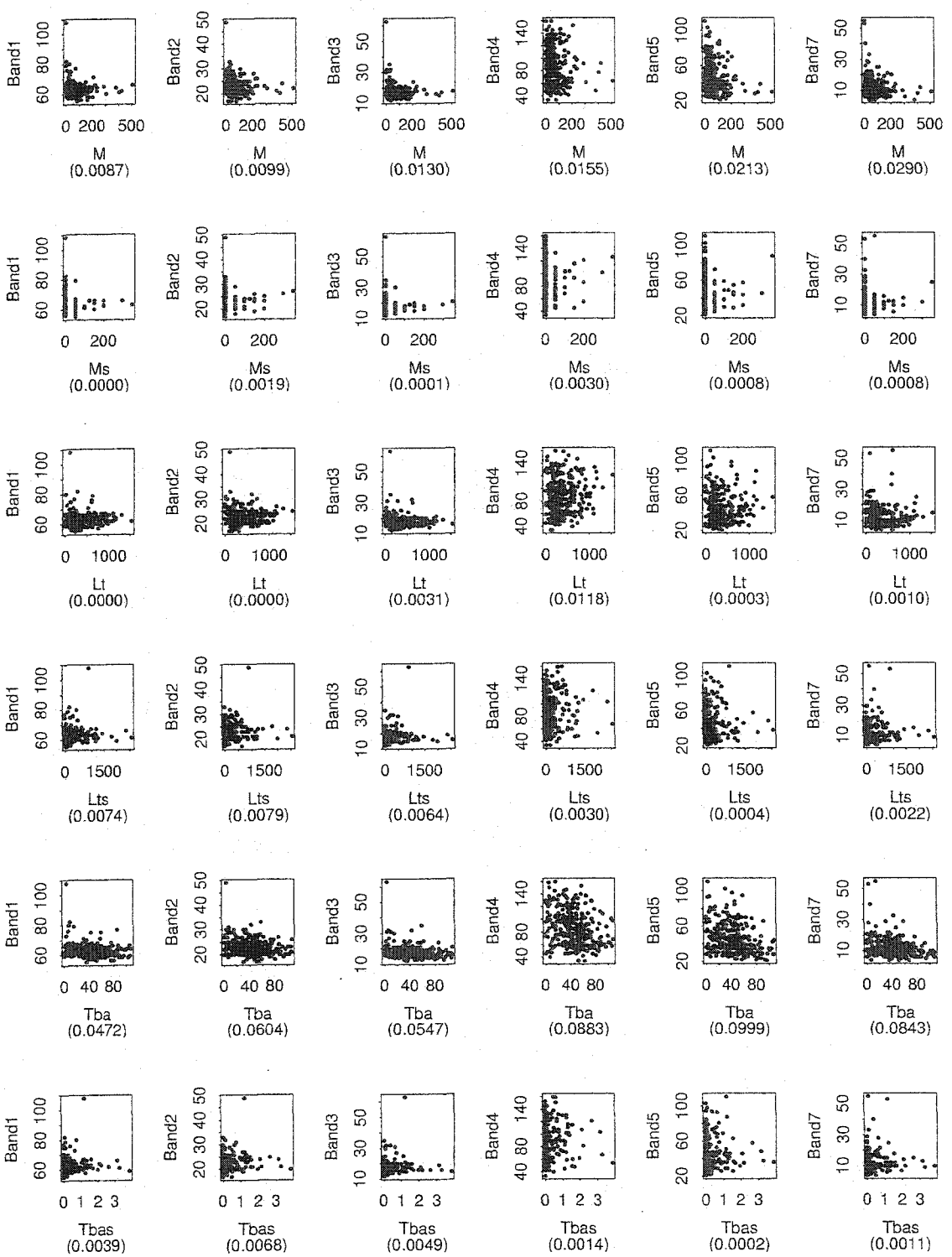


Figure 3.16: Scatter plots of all 6 variables versus 3 Tasseled Cap transformed band data, R^2 in parenthesis

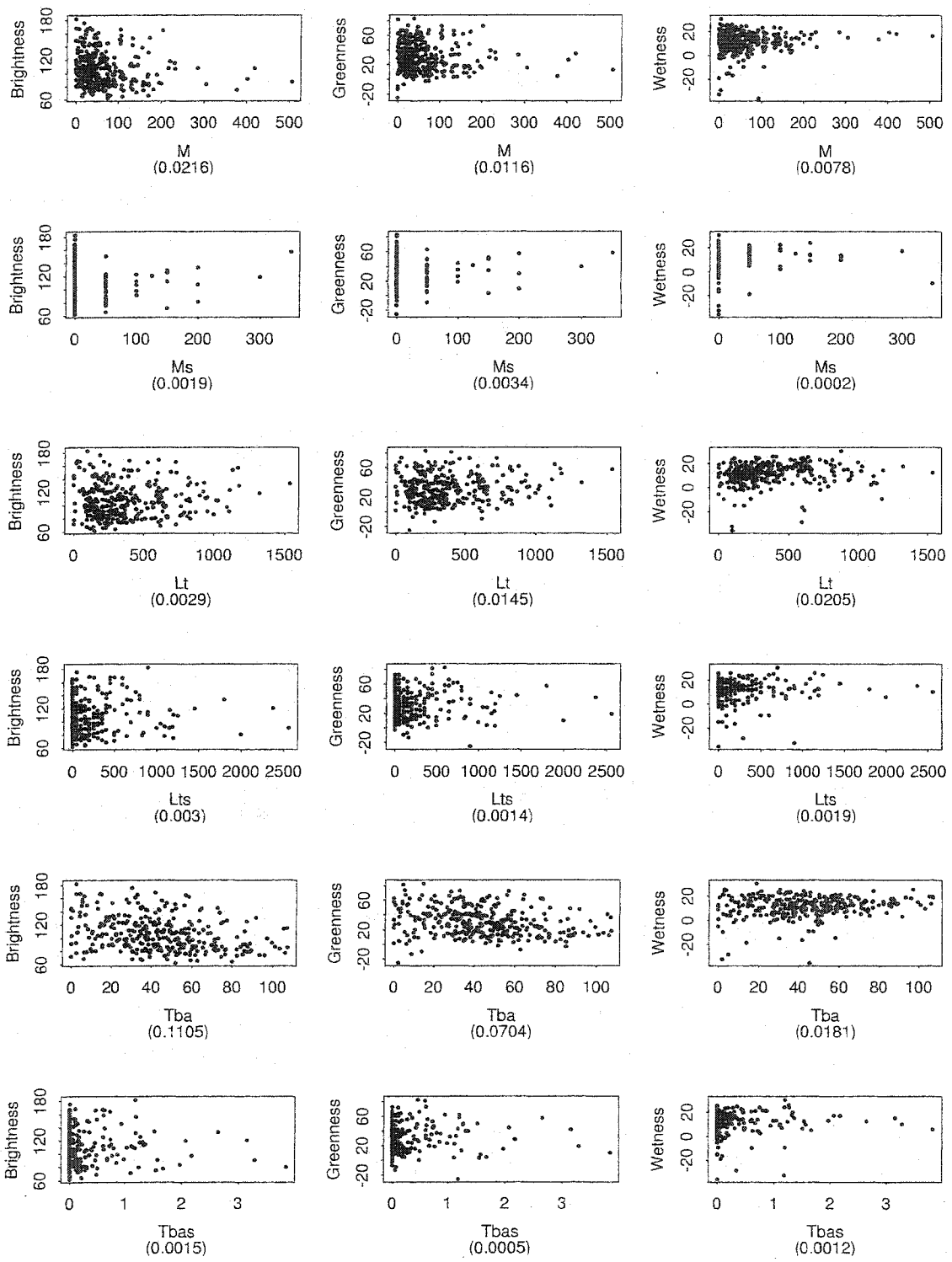


Figure 3.17: Scatter plots of all 6 variables versus 3 ratioing band data, R^2 in parenthesis

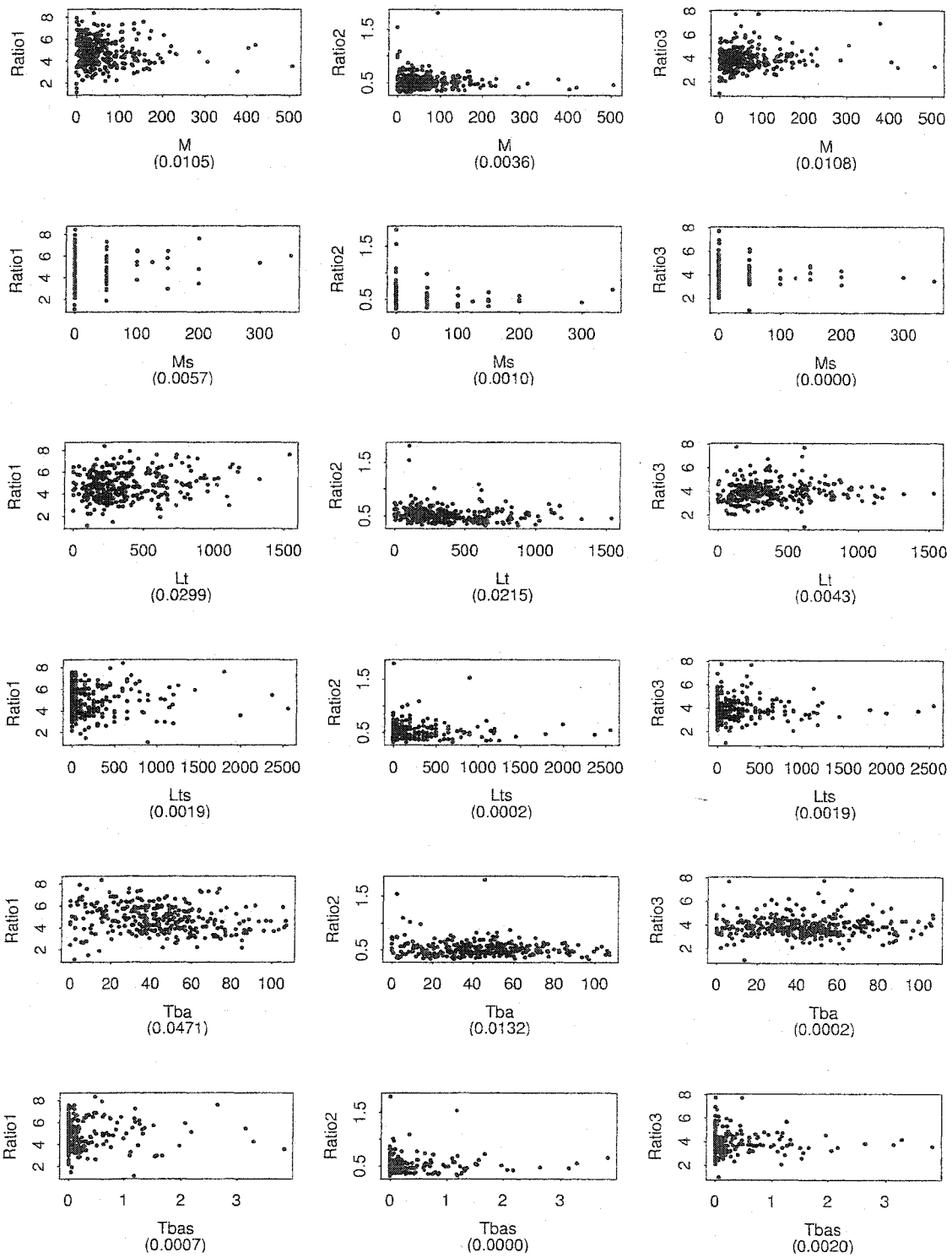


Figure 3.18: Mortality by plot (M) by location measured in trees/ha., classes 1=(0, 21], 2=(21, 50], 3=(50, 86], and 4=(> 86).

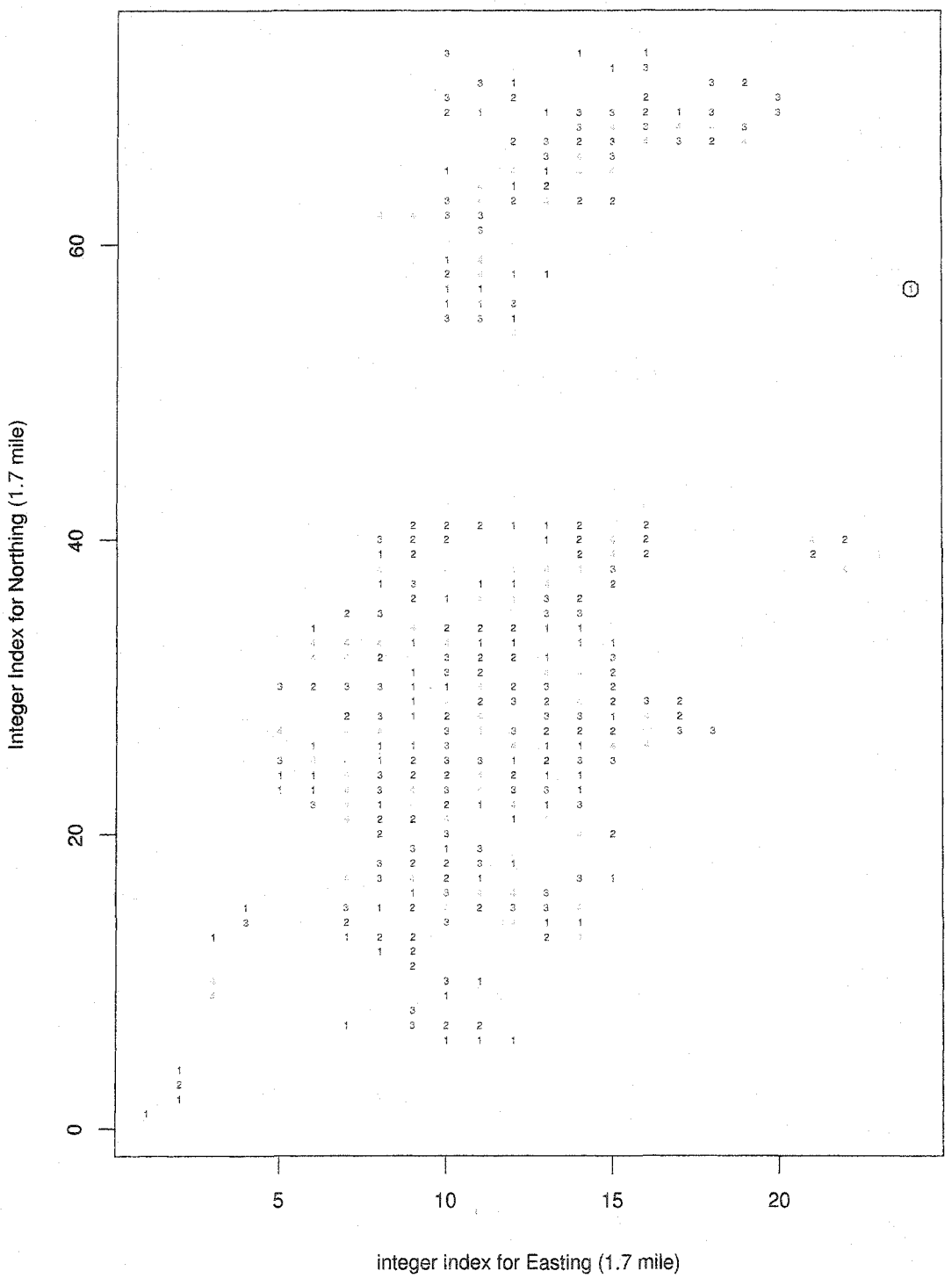


Figure 3.19: Locations for the large M (trees/ha.) values

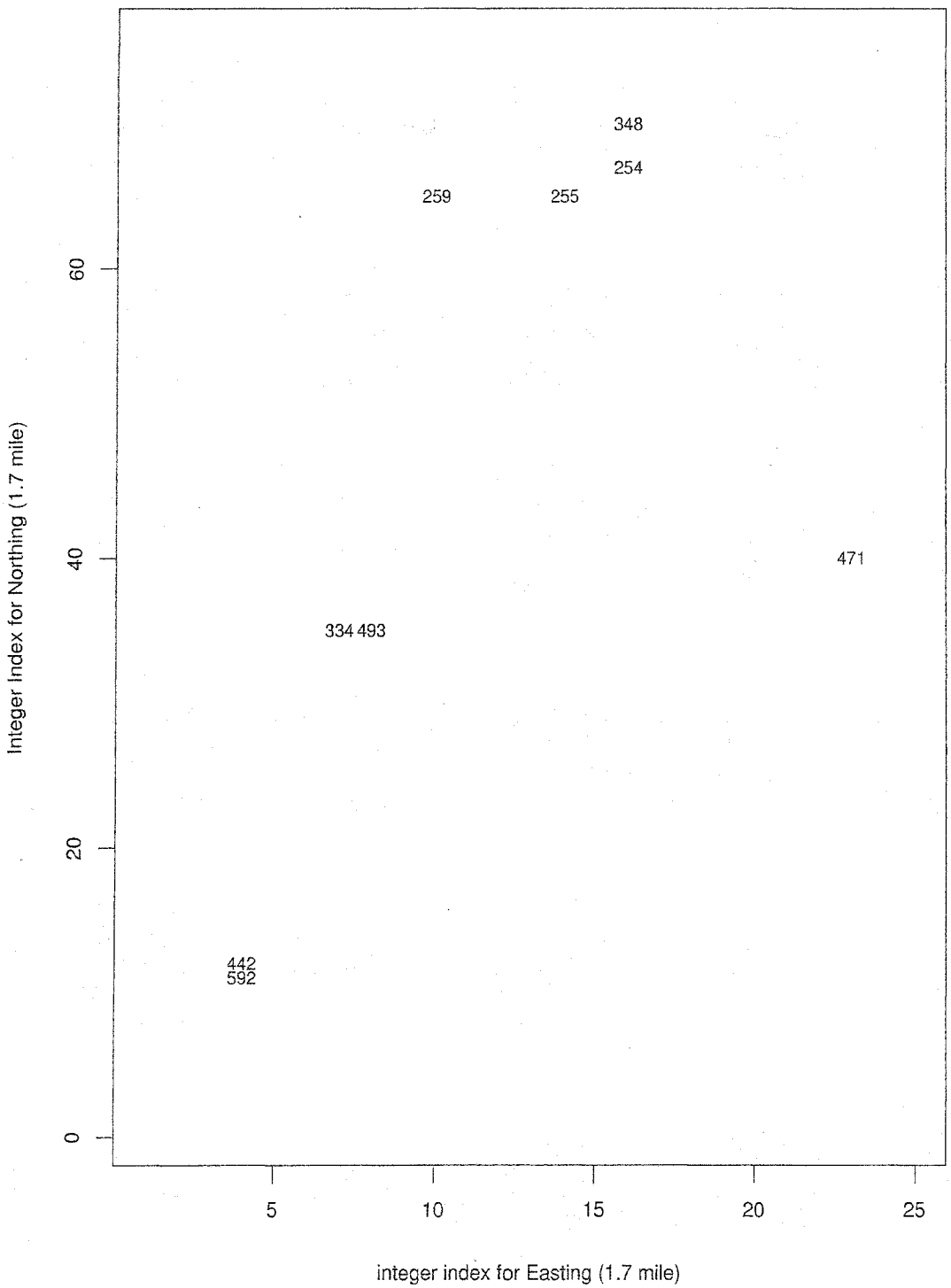


Figure 3.20: Mortality of seedlings by plot(M_s, trees/ha.) by location

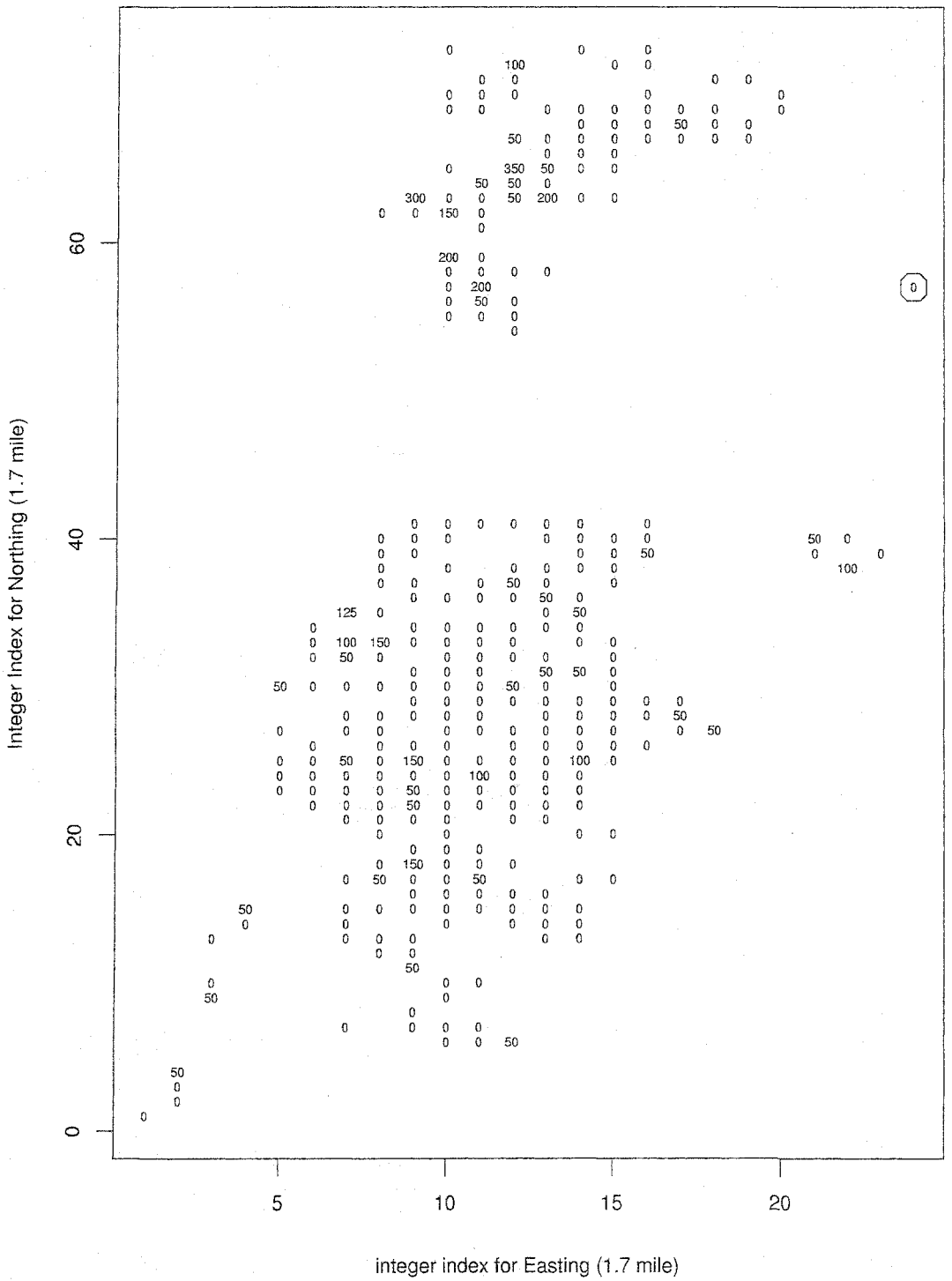


Figure 3.21: Locations for the large M_s (trees/ha.) values

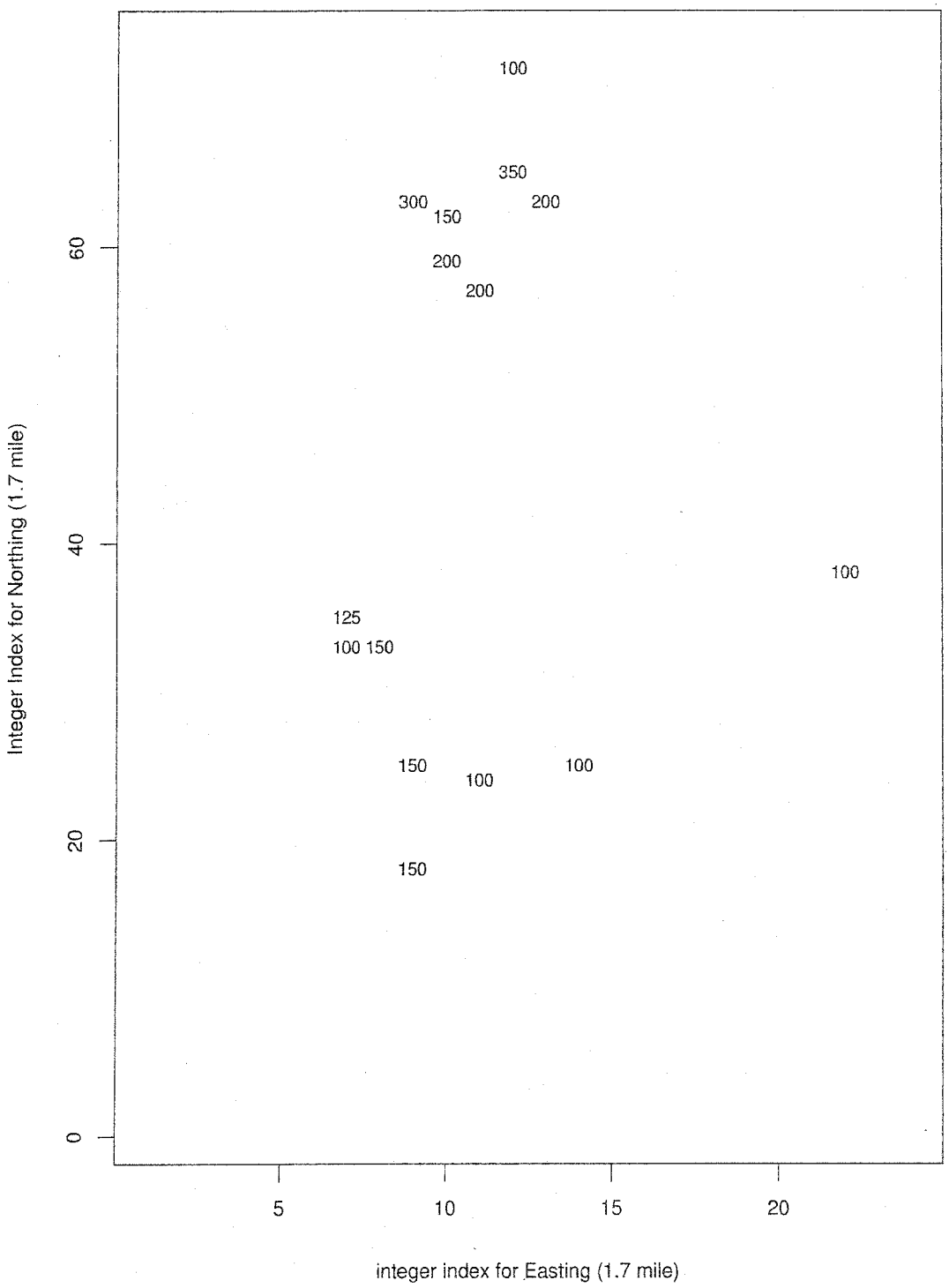


Figure 3.22: Number of live trees (Lt) per hectare by location, classes 1=(0, 180], 2=(180, 312], 3=(312, 577], and 4=(> 577).

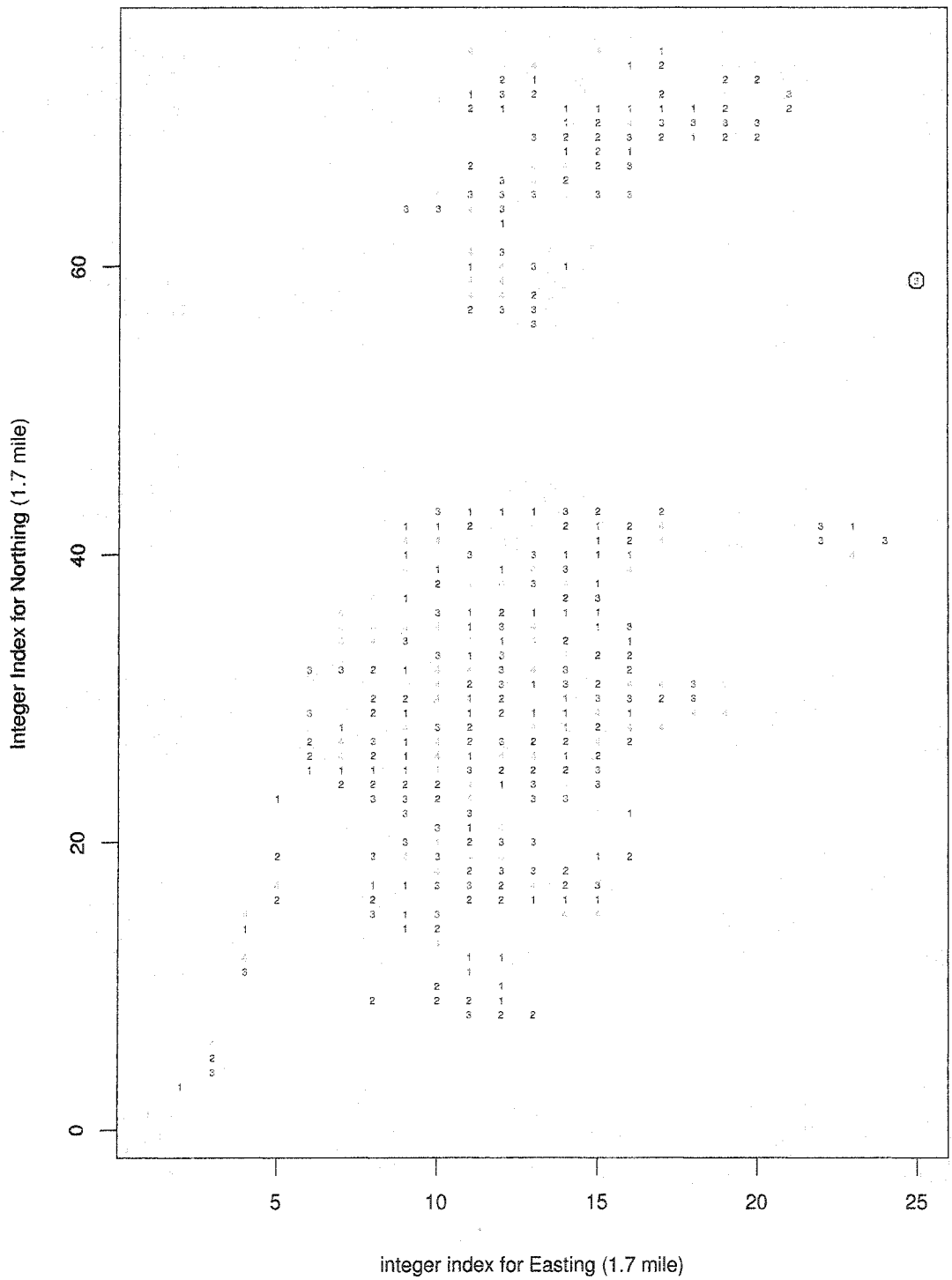


Figure 3.23: Locations for the large Lt (trees/ha.) values

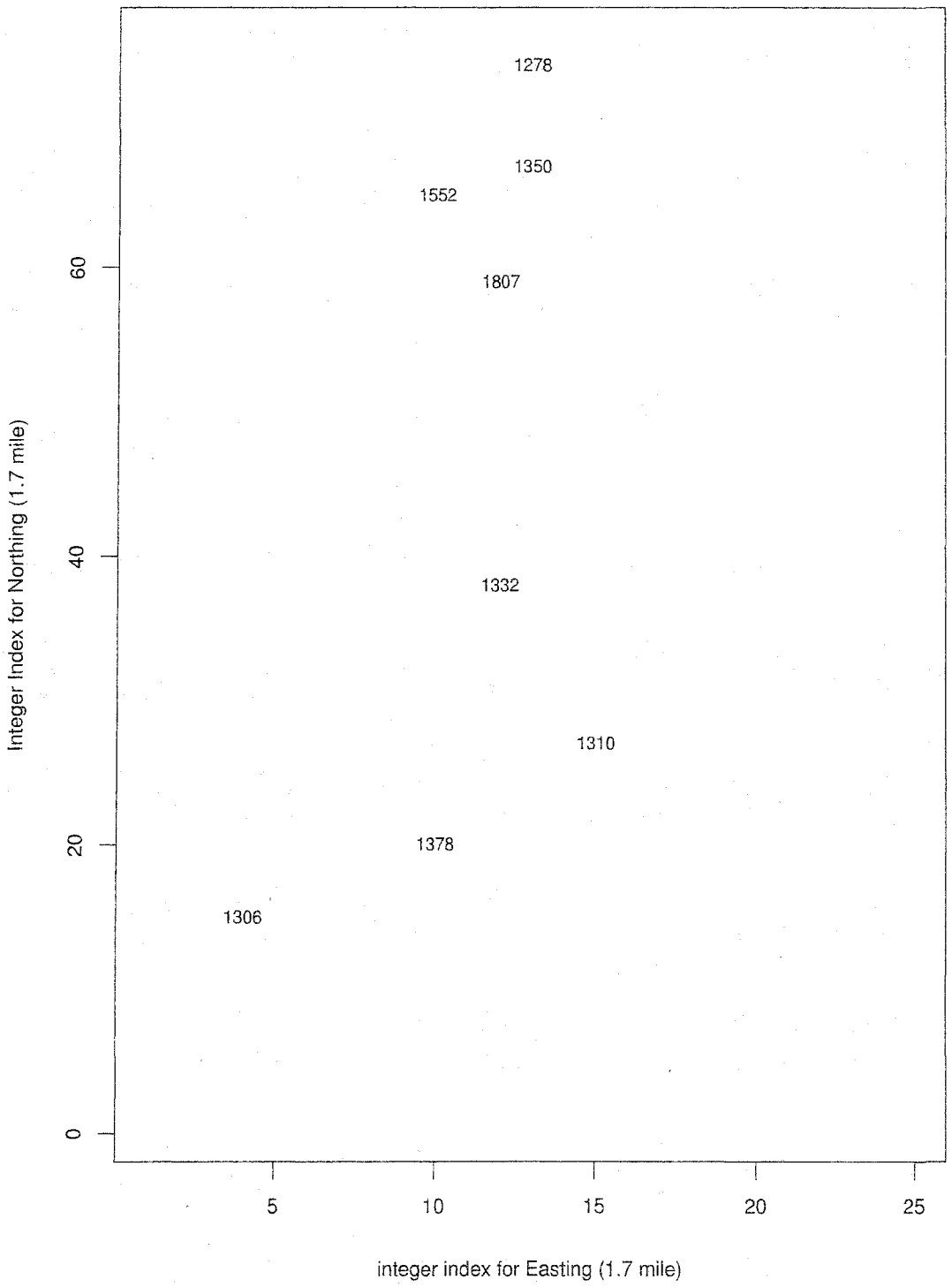


Figure 3.24: Number of live trees of seedlings (Lt_s, trees/ha.) by location

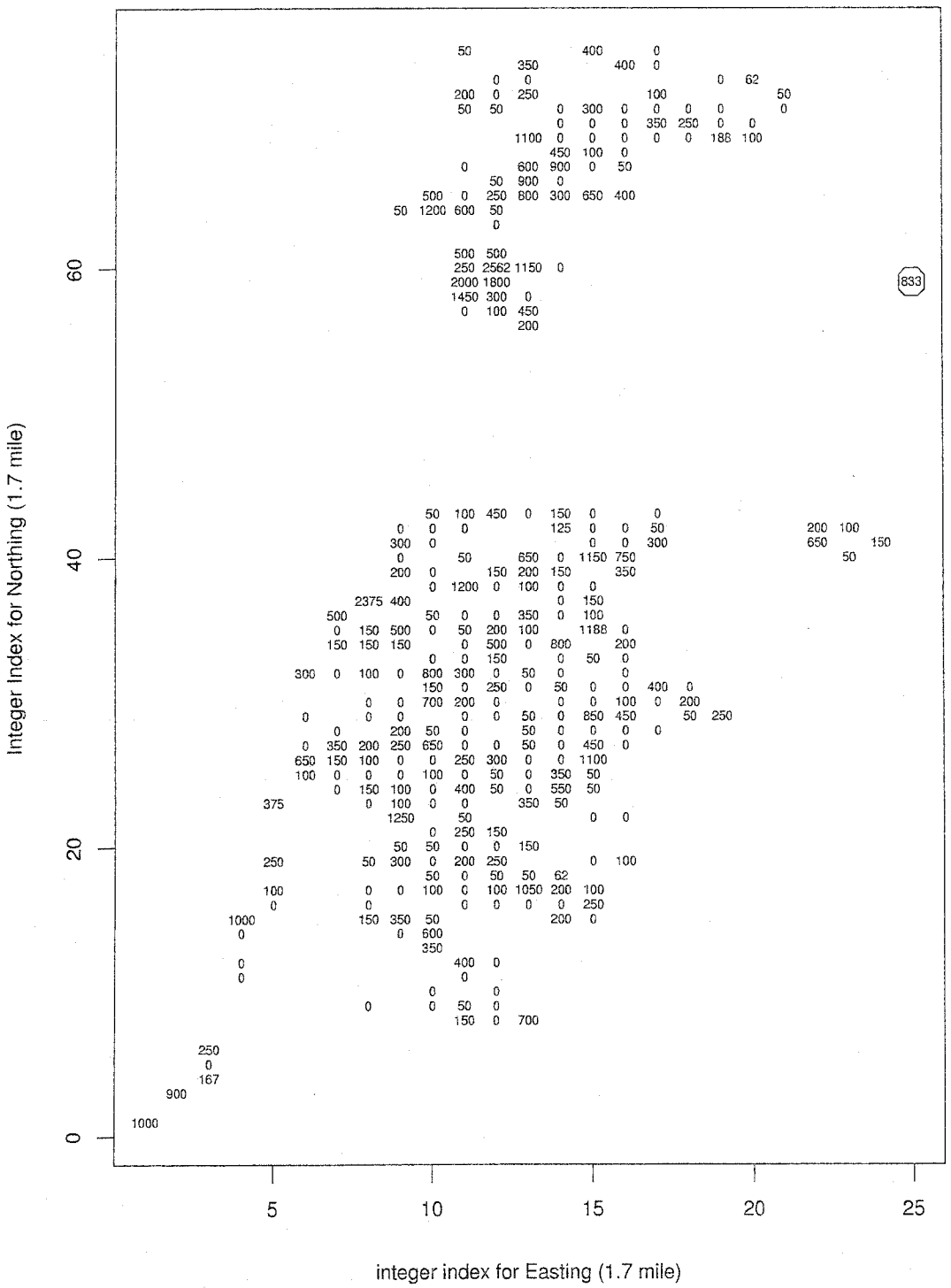


Figure 3.25: Locations for the large Lt_s (trees/ha.) values

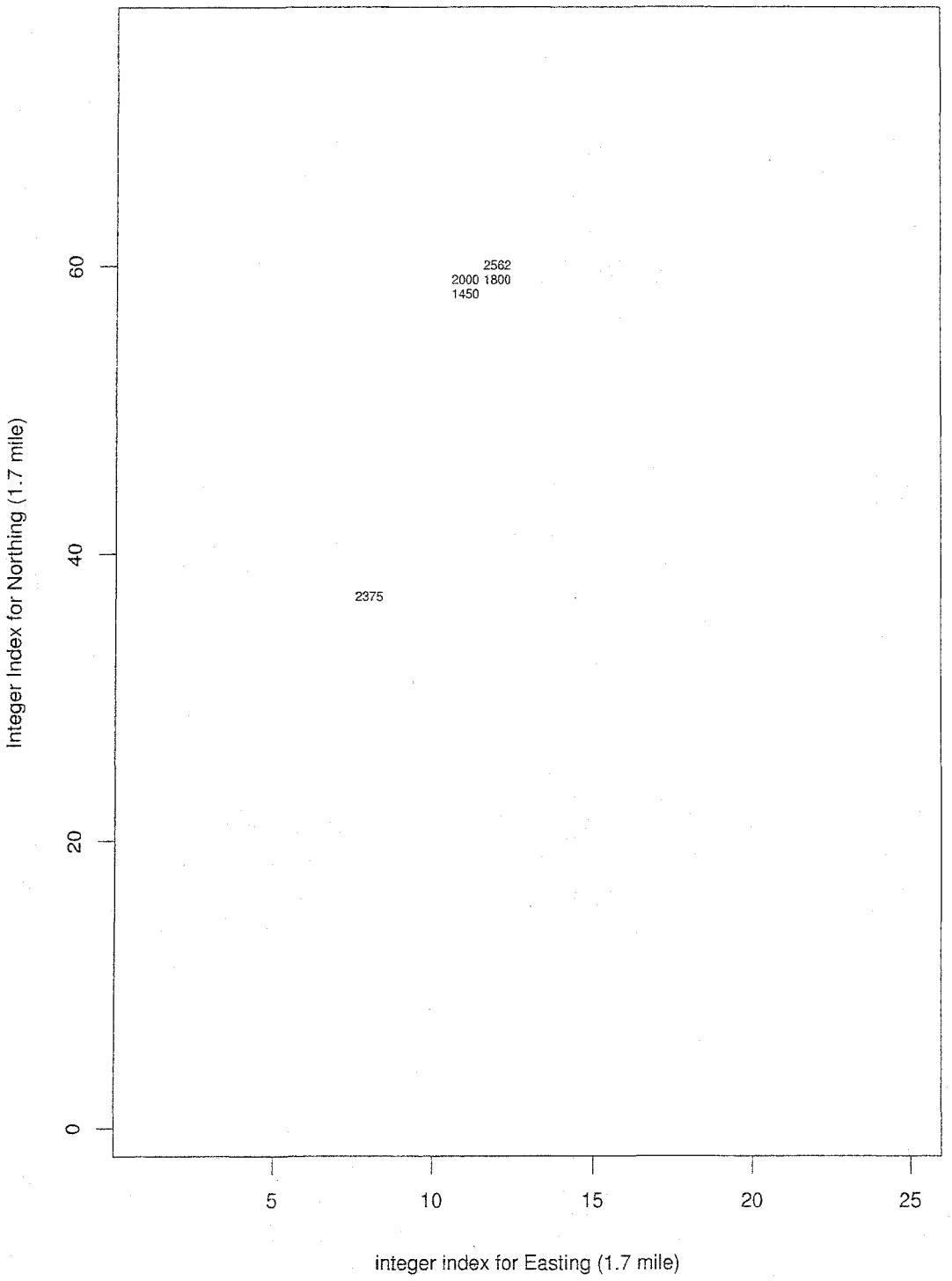


Figure 3.26: Total basal area (Tba, measured in $m^2/ha.$) by location, classes 1=(0, 28.22], 2=(28.22, 44.34], 3=(44.34, 59.07], and 4=($>$ 59.07).

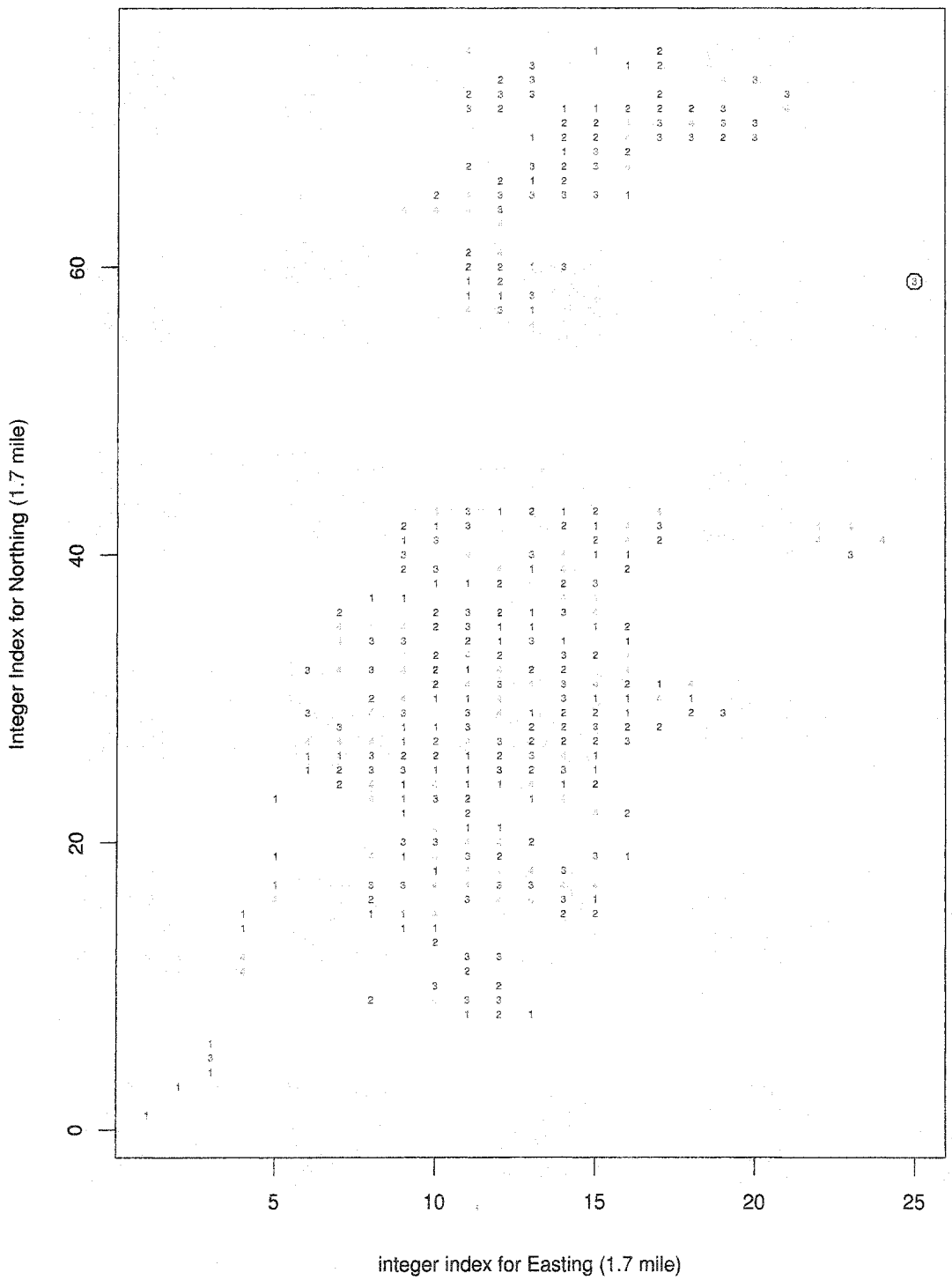


Figure 3.27: Locations for the large Tba ($m^2/ha.$) values

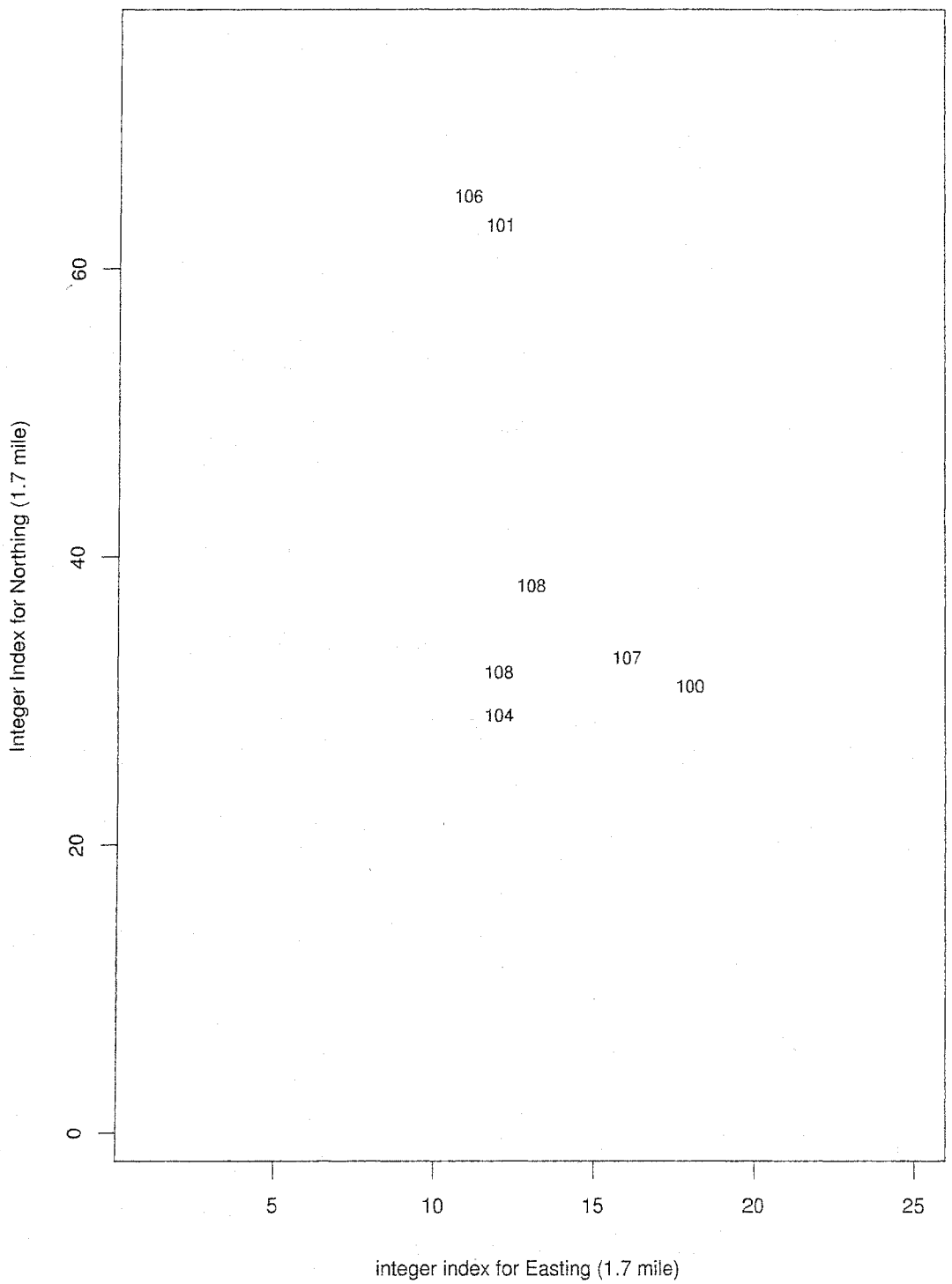


Figure 3.28: Total basal area of seedlings (Tba_s, measured in $m^2/ha.$) by location

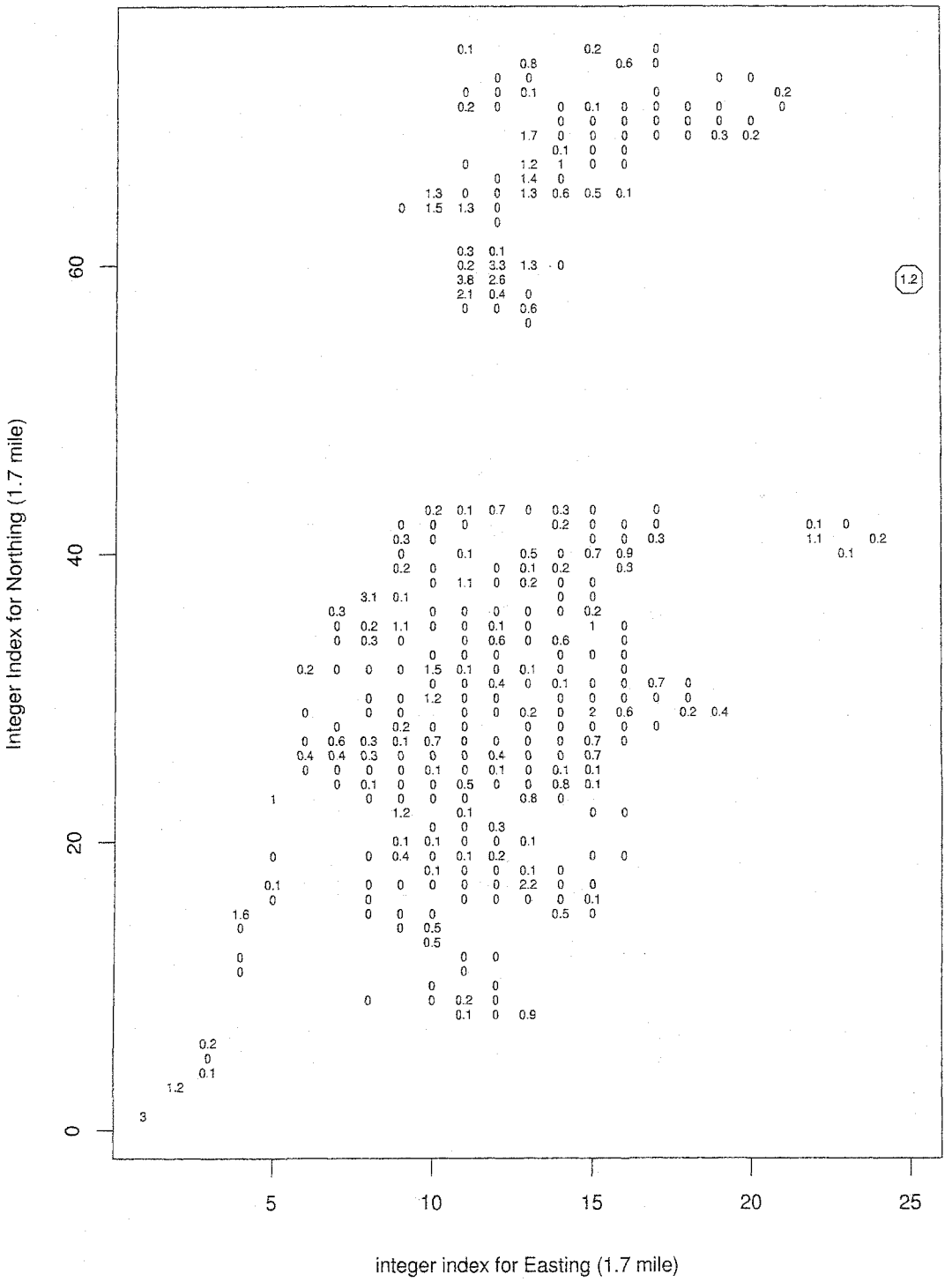


Figure 3.29: Locations for the large Tba_s $(m^2/ha.)$ values

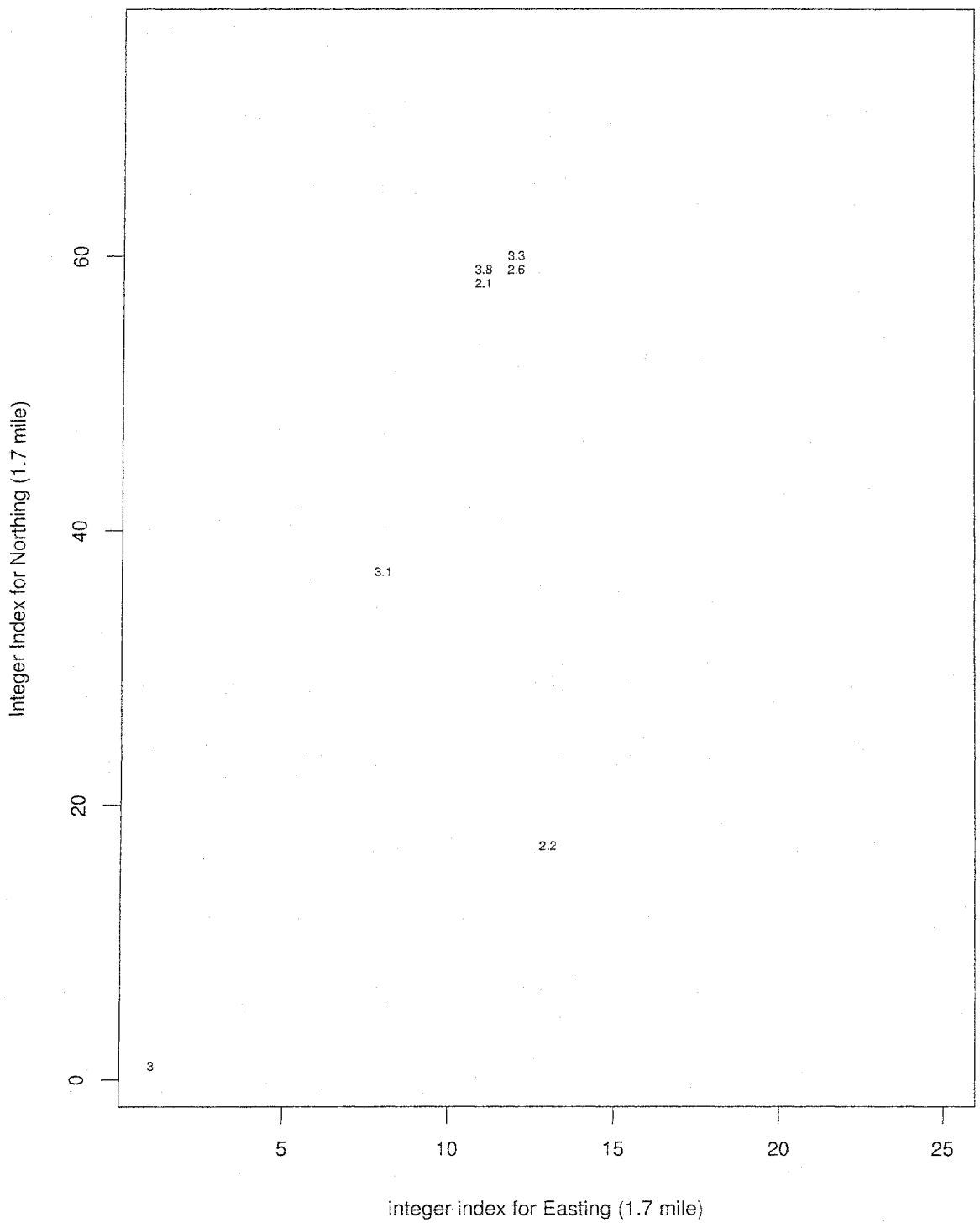
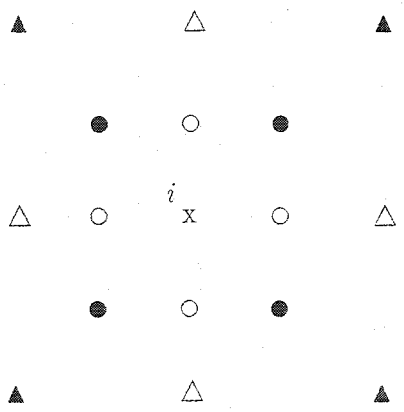


Figure 3.30: The Neighborhood of Plot i



○= one-unit-apart (1.7 mile) horizontal and vertical neighbors
●= one-diagonal-unit-apart (1.7 $\sqrt{2}$ mile) corner neighbors
△= two-unit-apart (3.4 mile) horizontal and vertical neighbors
▲= two-diagonal-unit-apart (3.4 $\sqrt{2}$ mile) corner neighbors

Table 3.4: Sample mean ($\overline{y}$), Sample Variance (s_y^2), and Sample Correlations

	$\overline{y}$	s_y^2	r_h	r_v	r_c	r_{2h}	r_{2v}	r_{2c}
M	62.75	4536.76	0.1363	0.1711	0.0454	0.0155	-0.0104	0.0111
M_s	12.54	1633.88	0.0393	0.0287	0.0727	-0.0541	0.1505	0.0621
Lt	367.64	72091.52	0.0947	0.1190	0.0429	0.0243	0.1151	-0.0103
Lt_s	200.84	128986.53	0.2300	0.3022	0.2101	-0.0306	0.0500	0.0519
Tba	43.50	590.40	-0.0147	0.0322	0.1489	0.0079	0.0169	0.0264
Tba_s	0.23	0.27	0.2243	0.3176	0.1713	-0.0597	0.0904	0.0379

r_h: sample correlation between one-unit-apart (1.7 mile) horizontal neighbors (e.g. x and horizontal-left and horizontal-right ○'s),
r_v: sample correlation between one-unit-apart vertical neighbors (e.g. x and vertical-upper and vertical-down ○'s),
r_c: sample correlation between one-unit-apart (1.7 $\sqrt{2}$ mile) corner neighbors (e.g. x and ●'s),
r_{2h}: sample correlation between two-unit-apart (3.4 mile) horizontal neighbors (e.g. x and horizontal-left and horizontal-right △'s),
r_{2v}: sample correlation between two-unit-apart vertical neighbors (e.g. x and vertical-upper and vertical-down △'s),
r_{2c}: sample correlation between two-unit-apart (3.4 $\sqrt{2}$ mile) corner neighbors. (e.g. x and ▲'s),

Figure 3.31: Scatter plots for plot-level variables in different lag directional neighborhoods

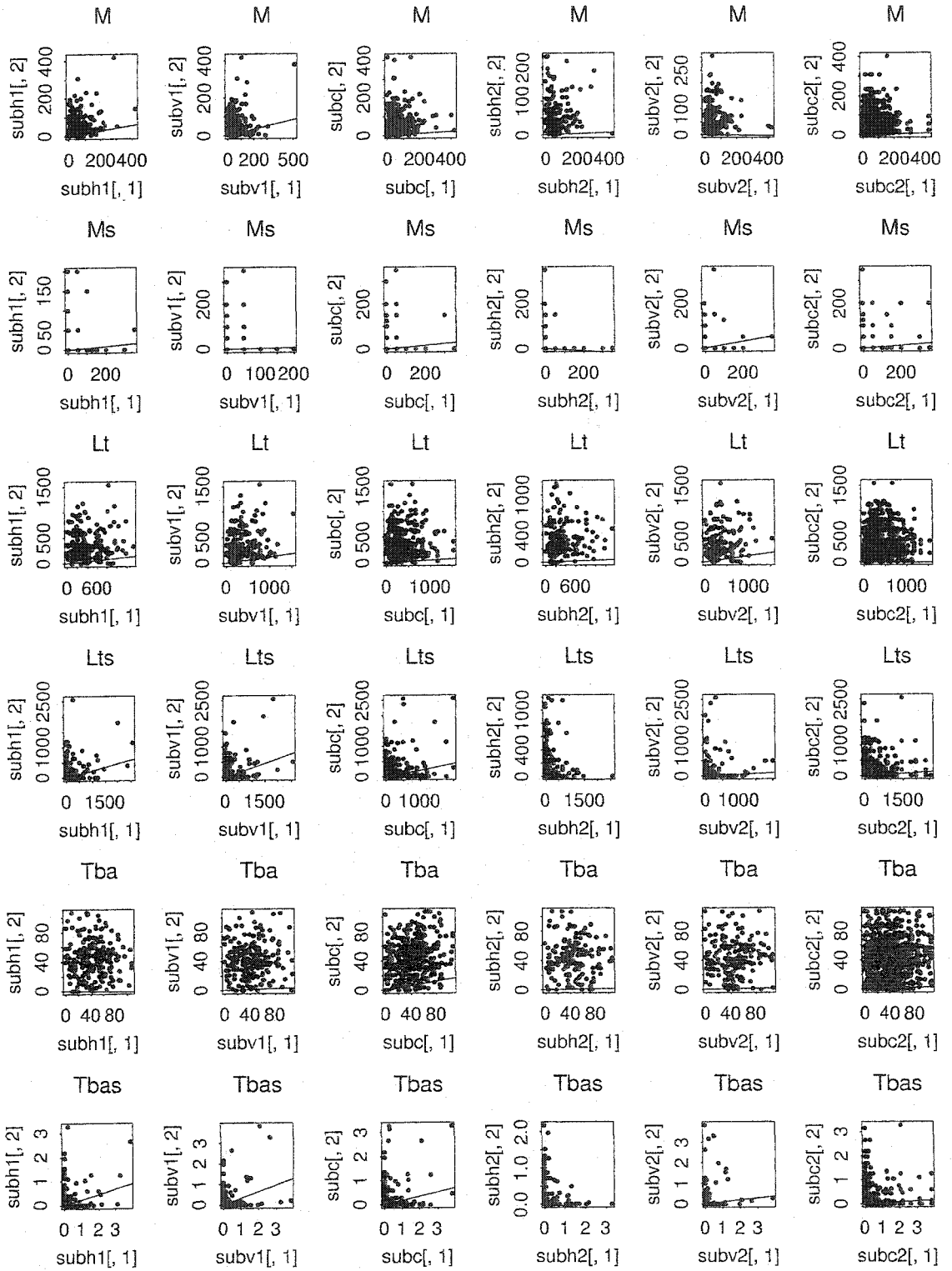
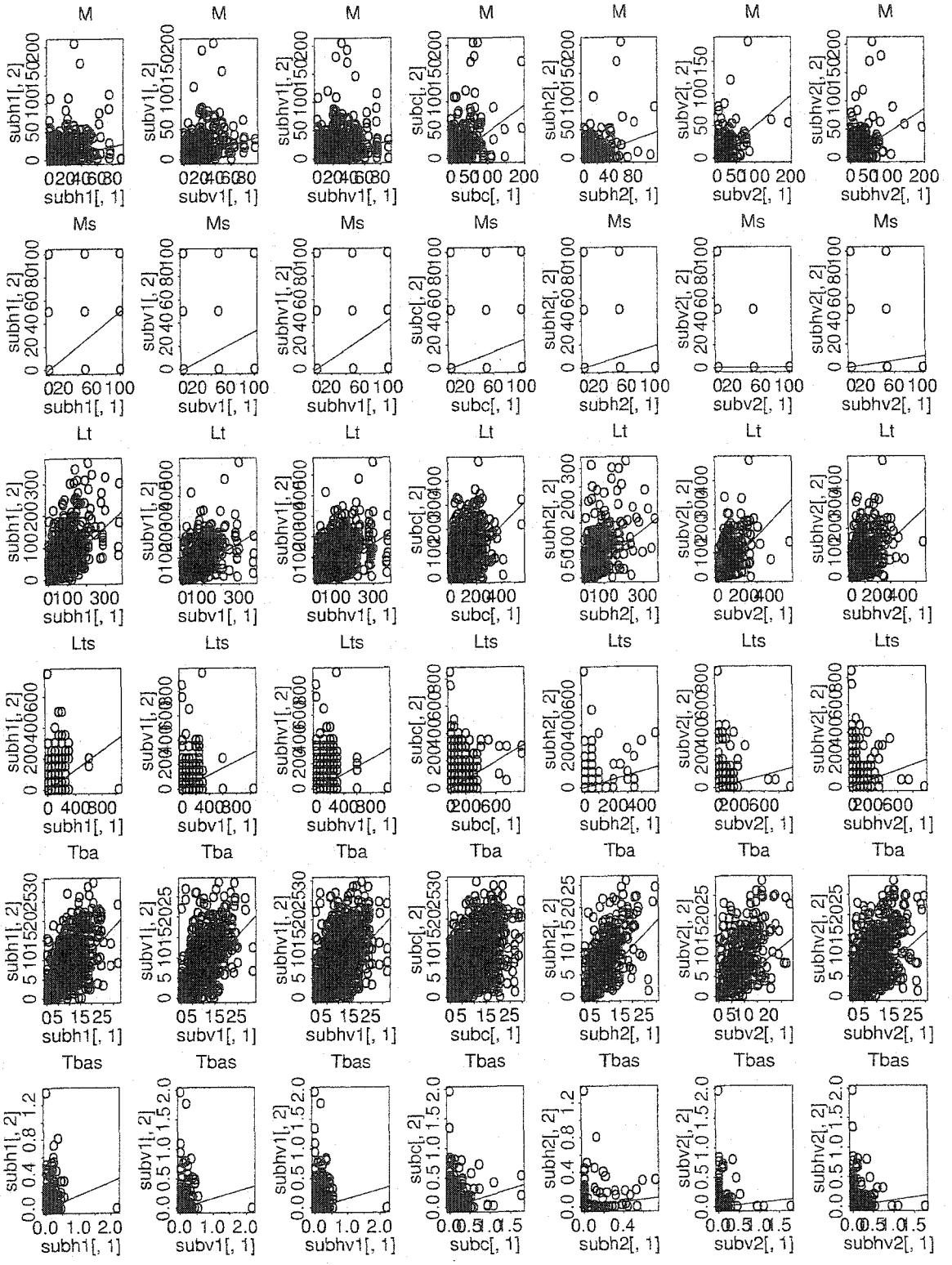


Table 3.5: Sample mean ($\overline{y}'$), Sample Variance (s'^2_y), and Sample Correlations for Subplot Variables

	$\overline{y}'$	s'^2_y	r'_h	r'_v	r'_{hv}	r'_c	r'_{2h}	r'_{2v}	r'_{2hv}
M	12.93	358.59	0.2875	0.3825	0.3350	0.4622	0.3835	0.4807	0.4321
M_s	2.59	159.98	0.4922	0.3204	0.4063	0.2306	0.1857	0.0015	0.0936
Lt	74.23	4553.30	0.5387	0.5674	0.5531	0.5296	0.4283	0.5491	0.4887
Lt_s	36.09	7780.96	0.3131	0.2822	0.2977	0.3205	0.2558	0.1416	0.1987
Tba	8.82	37.06	0.6073	0.6081	0.6077	0.5366	0.5585	0.4814	0.5199
Tba_s	0.04	0.02	0.1542	0.1650	0.1596	0.2033	0.1413	0.0684	0.1048

r'_h: sample correlation between one-unit-apart (40.8 m) horizontal neighbors
(i.e.stake positions (1,3) and (1,5)),
r'_v: sample correlation between one-unit-apart vertical neighbors
(i.e.stake positions (1,2) and (1,4)),
r'_{hv}: sample correlation between one-unit-apart horizontal and vertical neighbors
(i.e.stake positions (1,2), (1,3), (1,4) and (1,5)),
r'_c: sample correlation between one-unit-apart (40.8 m) corner neighbors (i.e.stake positions (2,3), (2,5), (3,4) and (4,5)),
r'_{2h}: sample correlation between two-unit-apart (81.6 m) horizontal neighbors
(i.e.stake positions (3,5)),
r'_{2v}: sample correlation between two-unit-apart vertical neighbors
(i.e.stake positions (2,4)),
where the stake positions shown in Figures 3.2 and 3.3.

Figure 3.32: Scatter plots for subplot-level variables in different lag directional neighborhoods



4. SIMPLE SPATIAL MODELS

This chapter introduces a simple spatial model which regresses the forest variables of interest on themselves with spatial dependence. In this model, we use Euclidean distance to define the neighborhood structure. In the model building, we consider interaction models for Σ, the dispersion matrix of a multivariate (MV) distribution.

Initially, we constructed a spatial autoregressive (SAR) model and a spatial moving average (SMA) model, but these possessed certain deficiencies. We then proceeded with the following alternative approach. Suppose $Y_{\boldsymbol{s_i}}$ is the variable of interest at location $\boldsymbol{s_i}$ for plot i, $i = 1, \ldots, N$, where N is the total number of plots or points observed. Since Σ is a dispersion matrix describing spatial dependence in the errors, $\Sigma_{\boldsymbol{s_i s_j}}$ denotes the covariance between $Y_{\boldsymbol{s_i}}$ and $Y_{\boldsymbol{s_j}}$ and $\Sigma_{\boldsymbol{s_i s_i}}$ denotes the variance of $Y_{\boldsymbol{s_i}}$. The matrix Σ must be symmetric and positive-definite, so any covariance scheme must ensure these conditions.

Let $\boldsymbol{\varepsilon}$ be an $N \times 1$ vector of (unobservable) random variables with mean zero and $N \times N$ dispersion matrix $\sigma^2\ \mathbf{I}$. We consider the schemes of the form

$$A(\mathbf{Y} - \boldsymbol{\mu}) = B\boldsymbol{\varepsilon}. \qquad (*)$$

Then $\mathbf{Y}_{N\times 1} \sim \text{MV}\ [\ \boldsymbol{\mu}_{\ N\times 1},\ \Sigma \equiv \sigma^2\ \Gamma = \sigma^2 BB']$. Since Γ is symmetric and positive-definite, B can be rewritten as $B = P\Lambda^{1/2}P'$, where P is an orthogonal matrix (i.e. $P' = P^{-1}$) and Λ is a diagonal matrix $\Lambda = diag(\lambda_1, \ldots, \lambda_N)$ consisting of the eigenvalues of Γ.

Assume that $A = \mathbf{I}$ and $B = \mathbf{I} - W$. Then the model can be expressed as

$$\mathbf{Y} = \boldsymbol{\mu} + B\boldsymbol{\varepsilon}. \tag{4.1}$$

For easy reference, we call this the "Phase I" model.

Multiplying both sides of the equation (4.1) by $B^{-1} \equiv \mathbf{I} - M$ where $M = (m_{ij})$, it may be rewritten as

$$(\mathbf{I} - M)(\mathbf{Y} - \boldsymbol{\mu}) = \boldsymbol{\varepsilon}.$$

or

$$\mathbf{Y} = \boldsymbol{\mu} + M(\mathbf{Y} - \boldsymbol{\mu}) + \boldsymbol{\varepsilon}. \tag{4.2}$$

For individual observations, equation (4.2) is equivalent to

$$Y_{\boldsymbol{s_i}} = \mu_{\boldsymbol{s_i}} + \sum_{j=1}^{N} m_{ij}(Y_{\boldsymbol{s_j}} - \mu_{\boldsymbol{s_i}}) + \varepsilon_{\boldsymbol{s_i}},\ i = 1, \ldots, N, \tag{4.3}$$

where the m_{ii}'s are zeros for $i = 1, \ldots, N$. The last equation makes the inter-dependency structure between individuals more intuitive. The value at plot i depends on the mean of its distribution at that plot, plus a weighted sum of the errors at the plots neighboring it. The plot location $\boldsymbol{s_i} = (s_{i1}, s_{i2})$ can be conveniently re-indexed by a single number k where k takes on values in $\{1, \ldots, N\}$. This notation of single-number plot index will be used in this thesis. The inter-relationship between plots is maintained by a so-called neighborhood system.

Now let us introduce some basic concepts of spatial neighborhood dependence that will be reflected in the selection of $\boldsymbol{\Gamma}$ in the multivariate normal model. The neighborhood system was described in Section 3.5. The neighboring set of i on the 2D lattice, as illustrated in Figure 3.31 is defined as the set of nearby plots (sites) within a radius of h, $N_i = \{j : d_{ij} \leq h, \forall j \neq i\}$ where d_{ij} denotes the Euclidean distance between plots i and j and h takes the values of 1.7 mile, $1.7\sqrt{2}$ mile,

etc. Since the neighborhood is defined by the neighbors within a specific distance, it is natural to specify the spatial interactions by associating the neighborhood structure with a distance-based structure of the correlation matrix. There are many ways to do this; for example,

(i) $\rho_{ij} \propto d_{ij}^{-k}$, $(k \geq 0)$

where d_{ij} is the distance between plots i and j and k is a constant, or $\rho_{ij} = \theta d_{ij}^{-k}$ where θ is a parameter.

(ii) $\rho_{ij} = exp\,(\theta d_{ij}^{-k})$

where exp denotes the exponential function, k is a constant (one or two in our study) and θ is a parameter.

Not all plots will have the same number of neighbors due to boundary effects and missing data in the lattice network. Plots at or near the boundaries tend to have fewer neighbors. For a regular $m \times n$ rectangular lattice, the domain consists of sites $\boldsymbol{s}_i$=(s_{i1}, s_{i2}) for $s_{i1} = 1, \ldots, m$ and $s_{i2} = 1, \ldots, n$ not ordered in a matrix manner. When the plots constitute a regular rectangular lattice, and a first order neighborhood system is considered, an internal plot has four nearest neighbors, a plot on a non-corner boundary has three and a plot at the corners has two. The neighboring relationship in this kind of lattice has the following properties:

(1) a plot is not a neighbor to itself: $i \notin N_i$,

(2) the neighbor relationship is mutual: $i \in N_j \Leftrightarrow j \in N_i$.

When the plots consist of a regular rectangular lattice with holes, the number of neighbors for an internal, boundary, or corner plot are less or equal to those in a regular rectangular lattice without holes. The above definition of neighborhood is different from the k-nearest neighbors (k-NN) methods which use a fixed number of neighbors (introduced in Chapter 2). Note that k-NN methods don't possess the mutual neighbor property.

4.1 Specifying the Γ Structure of the Multivariate (MV) Distribution

In our MV spatial model, one specification of the covariance-variance matrix $\sigma^2\,\Gamma$ is given by decomposing Γ as $\Gamma \equiv \text{Cor}(\mathbf{Y}) = \mathbf{I}+\rho_1\,\mathbf{H}+\rho_2\,\mathbf{V}+\rho_3\,\mathbf{C}^{(1)}+\rho_4\,\mathbf{C}^{(2)}$. Here $\mathbf{I}$ is an identity matrix, $\mathbf{H}$ and $\mathbf{V}$ denote the corresponding (0,1) matrices with 1's for the horizontal and vertical neighbors, respectively, and $\mathbf{C}^{(1)}$ and $\mathbf{C}^{(2)}$ denote the corresponding (0,1) matrices with 1's for corner neighbors in the directions {(1,1),(-1,-1)} and {(1,-1),(-1,1)}, respectively, and zeros elsewhere in all matrices. The real data is equal distance (1.7 mile) lattice data with missing values. Under this assumed correlation structure, the $N(N-1)$ correlation coefficients among the N spatially scattered observations depend on the four parameters, ρ_1, ρ_2, ρ_3, and ρ_4. For some applications, first order spatial dependence is adequate, so sometimes we will assume some of the ρ_i's to be zero.

A. We will investigate the following schemes:

(1a) symmetric first order scheme, called $S_{\rho_1\rho_100}$, i.e. $\rho_1 = \rho_2 = \rho_{1a}$ and $\rho_3 = \rho_4 = 0$ (Figure 4.1(1a)). The correlation matrix reduces to $\Gamma = \mathbf{I} + \rho_{1a}\,(\mathbf{H}+\mathbf{V})$ $\equiv \mathbf{I}+\ \rho_{1a}\,\mathbf{W}^{(1a)}$, a function of the spatial dependence parameter ρ_{1a}. This assumes row and column correlations are identical and diagonal correlations are zero.

(1b) asymmetric first order scheme, called $S_{\rho_1\rho_200}$, i.e. $\rho_3 = \rho_4 = 0$ (Figure 4.1(1b)). The correlation matrix reduces to $\Gamma = \mathbf{I} + \rho_1\,\mathbf{H} + \rho_2\,\mathbf{V}$. In this scheme, Γ is a function of the spatial dependence parameters ρ_1 and ρ_2. This assumes row and column correlations are different and diagonal correlations are zero.

(2a) symmetric second order scheme, called $S_{\rho_1\rho_1\rho_1\rho_1}$, i.e. $\rho_1 = \rho_2 = \rho_3 = \rho_4 = \rho_{2a}$ (Figure 4.1(2a)). This includes more neighbors than (1a) and (1b). The

correlation matrix then takes the form of $\mathbf{\Gamma} = \mathbf{I} + \rho_{2a}(\mathbf{H}+\mathbf{V}) + \rho_{2a}(\mathbf{C}^{(1)}+\mathbf{C}^{(2)}) \equiv \mathbf{I} + \rho_{2a}\mathbf{W}^{(2a)}$. This assumes row, column, and diagonal correlations are identical.

(2b) asymmetric second order scheme, called $S_{\rho_1\rho_1\rho_3\rho_3}$, i.e. $\rho_2 = \rho_1 = \rho_{1a}, \rho_4 = \rho_3 = \rho_{2b}$ (Figure 4.1(2b)). The correlation matrix reduces to $\mathbf{\Gamma} = \mathbf{I} + \rho_1(\mathbf{H} + \mathbf{V}) + \rho_3(\mathbf{C}^{(1)} + \mathbf{C}^{(2)}) \equiv \mathbf{I} + \rho_{1a}(\mathbf{H} + \mathbf{V}) + \rho_{2b}(\mathbf{C}^{(1)} + \mathbf{C}^{(2)})$, a function of parameters ρ_{1a} and ρ_{2b}. This assumes row and column correlations are identical but different from the identical diagonal correlations.

(2c) asymmetric second order scheme, called $S_{\rho_1\rho_2\rho_3\rho_3}$, i.e. $\rho_4 = \rho_3 = \rho_{2c}$ (Figure 4.1(2c)). The correlation matrix reduces to $\mathbf{\Gamma} = \mathbf{I} + \rho_1\mathbf{H} + \rho_2\mathbf{V} + \rho_3(\mathbf{C}^{(1)} + \mathbf{C}^{(2)}). \equiv \mathbf{I} + \rho_1\mathbf{H} + \rho_2\mathbf{V} + \rho_{2c}\mathbf{W}^{(2c)}$. This assumes unequal row and column correlations with identical diagonal correlations.

(2d) asymmetric second order scheme, called $S_{\rho_1\rho_1\rho_3\rho_4}$, i.e. $\rho_2 = \rho_1 = \rho_{2d}$ (Figure 4.1(2d)). The correlation matrix reduces to $\mathbf{\Gamma} = \mathbf{I} + \rho_{2d}(\mathbf{H} + \mathbf{V}) + \rho_3\mathbf{C}^{(1)} + \rho_4\mathbf{C}^{(2)} \equiv \mathbf{I} + \rho_{2d}\mathbf{W}^{(2d)} + \rho_3\mathbf{C}^{(1)} + \rho_4\mathbf{C}^{(2)}$ where $\mathbf{W}^{(2d)} = \mathbf{W}^{(1a)}$. This assumes equal row and column correlations with different diagonal correlations.

(2e) asymmetric second order scheme, (Figure 4.1(2e)), called $S_{\rho_1\rho_2\rho_3\rho_4}$. The correlation matrix is $\mathbf{\Gamma} = \mathbf{I} + \rho_1\mathbf{H} + \rho_2\mathbf{V} + \rho_3\mathbf{C}^{(1)} + \rho_4\mathbf{C}^{(2)}$, a function of the four parameters ρ_1, ρ_2, ρ_3 and ρ_4. This assumes all row, column, and diagonal correlations to be different. This is the most general case considered.

B. Alternative specifications of the covariance-variance matrix of $\sigma^2\mathbf{\Gamma}$ that are more directly related to distance are: $\rho_{ii} = 1$

(1) $\rho_{ij} = \theta d_{ij}^{-k}$, for $i \neq j$ $(k = 1, 2)$

with no neighborhood constraint. This assumes that the correlations decay

by the proportion of inverse distance or squared inverse distance for all observations.

(2) $\rho_{ij} = \theta d_{ij}^{-k}$, for $i \neq j$ $(k = 1, 2)$

with (a) a first or (b) a second order neighborhood constraint. This is a special case of (1) with neighborhood constraint. The correlations beyond first or second order neighbors are assumed zeros.

(3) $\rho_{ij} = \theta exp(-d_{ij}^{k})$, for $i \neq j$ $(k = 1, 2)$

with no neighborhood constraint. This assumes that the correlation decays by the proportion of the distance exponent function which decays much more quickly than the correlations (1).

(4) $\rho_{ij} = \theta exp(-d_{ij}^{k})$, for $i \neq j$ $(k = 1, 2)$

with (a) a first or (b) a second order neighborhood constraint.

The correlations beyond the first or second order neighborhoods are assumed zeros.

4.2 Constraints on the Phase 1 Model Parameters

Since $\mathbf{\Gamma}$ must be positive-definite, there are restrictions on the possible values of the correlation ρ_1, ρ_2, ρ_3, and ρ_4 for specifications **A** and **B** and θ for the distance-related alternative.

Conditions: $\mathbf{\Gamma}$ is a positive-definite matrix ($a'\mathbf{\Gamma}a > 0$ for all nonzero vectors a) if it satisfies one of the following equivalent conditions

(i) All eigenvalues of $\mathbf{\Gamma}$ are > 0.

(ii) All the upper left submatrices $\mathbf{\Gamma}_k$ have positive determinants.

(iii) All the pivots (without row exchange) satisfy $d_i > 0$ (the d_i's are the diagonal entries on the D matrix of A=LDU factorization where L is a lower triangular matrix, D is a diagonal matrix, and U is an upper triangular matrix).

We discuss the parameter constraints for two cases: (a) a regular $m \times n$ rectangular lattice and (b) a regular $m \times n$ rectangular lattice with holes (i.e. with missing data). The plot locations $\boldsymbol{s}=(s_1, s_2)$ are ordered $(l'_{s_1}, l_1), \ldots, (l'_{s_1}, u_1), (l'_{s_1}+1, l_2), \ldots, (l'_{s_1}+1, u_2), \ldots, (u'_{s_1}, l_m), \ldots, (u'_{s_1}, u_m)$ where l_{s_1} and u_{s_1} respectively denote the smallest and largest of s_2 for $s_1 = l'_{s_1}, \ldots, u'_{s_1}$, $l'_{s_1} = min\{s_1\}$, and $u'_{s_1} = max\{s_1\}$, i.e. the ordering starts with the west most column among all location points, and is indexed from the bottom to the top of that column, and then the following next right column (See Figures 3.4 and 3.5). Define n_{s_1} to be the number of points for $s_1 = 1, \ldots, m$, so $N = \sum_{i=1}^{m} n_{s_i}$.

With either type of data lattice, the points can be ordered in the data vector

$$\begin{aligned} & \mathbf{Y}_{N\times 1} \\ = \; & \Big(Y_{(l'_{s_1},l_1)}, \ldots, Y_{(l'_{s_1},u_1)}, Y_{(l'_{s_1}+1,l_2)}, \ldots, Y_{(l'_{s_1}+1,u_2)}, \ldots\ldots \\ & Y_{(u'_{s_1},l_m)}, \ldots, Y_{(u'_{s_1},u_m)}\Big)' \\ \equiv \; & (Y_1, \ldots, Y_N)' \end{aligned}$$

where l_{s_1} and u_{s_1} denote the smallest and largest of s_2 when $s_1 = 1, \ldots, m$ respectively, and $N = \sum_{s_1=1}^{m} n_{s_1}$.

4.2.1 Regular rectangular lattice

For an $m \times n$ lattice, the correlation matrix $\boldsymbol{\Gamma}$ of scheme $S_{\rho_1\rho_2\rho_3\rho_3}$ as an illustration can be expressed as

$$\boldsymbol{\Gamma} = I_{mn} + (\rho_1 M_m) \otimes I_n + I_m \otimes (\rho_2 M_n) + (\rho_3 M_m) \otimes M_n$$

where the notation $\otimes$ denotes the Kronecker product, I_p is a $p \times p$ identity matrix and M_n is a symmetric $n \times n$ matrix with $M_n(i,j) = 1$ if $\mid i - j \mid = 1$ and zeros

elsewhere. Let λ_i and α_j denote the eigenvalues of M_m and M_n respectively. Then the mn eigenvalues of $\mathbf{\Gamma}$ are $1+\rho_1\lambda_i+\rho_2\alpha_j+\rho_3\lambda_i\alpha_j$ (see Bellman 1970, pp.238). Graybill (1983) provided the explicit form of the eigenvalues of a $n \times n$ Toeplitz matrix with bandwidth 3 (which has constant diagonal, the element a_{ij} depends only on $(i-j)$ and the elements are zeros beyond the minor diagonal). The eigenvalues of M_n are $2cos(\frac{i\pi}{n+1})$ for $i=1,\ldots,n$ and hence the eigenvalues of $\mathbf{\Gamma}$ are

$$\lambda_{ij} = 1+2\rho_1 cos(\frac{i\pi}{m+1})+2\rho_2 cos(\frac{j\pi}{n+1})+4\rho_3 cos(\frac{i\pi}{m+1})cos(\frac{j\pi}{n+1})$$

for $i=1,\ldots,m$ and $j=1,\ldots,n$. The matrix $\mathbf{\Gamma}$ is positive-definite if and only if $\lambda_{ij}>0$. For large $m \times n$ lattices, a sufficient condition for the positive definiteness of $\mathbf{\Gamma}$ is $2(\mid \rho_1 \mid + \mid \rho_2 \mid)+4 \mid \rho_3 \mid < 1$ since the cosine functions in the eigenvalue formula are bounded by 1 for large m and n.

The eigenvalues of the correlation matrices for the cases under **A** are listed in Table 4.1. The alternative correlation matrix $\mathbf{\Gamma}$ (described earlier) can be expressed in the form of $\mathbf{\Gamma}=\mathbf{I}+\theta\ \mathbf{W}$ for all cases, where the elements w_{ij} of $\mathbf{W}$ take values of d_{ij}^{-k} or $exp(-d_{ij}^{k})$ for $i \neq j$ and zero for $i=j$ with no neighborhood constraint; while for 1st or 2nd neighborhood constraints, w_{ij} is zero beyond these 1st or 2nd neighbors. The alternative correlation matrix $\mathbf{\Gamma}$ has a closed form for the eigenvalues when using 1st or 2nd neighborhood constraints, but has only a numerical solution for the matrix with no neighborhood constraint. Assume that $\alpha_1,\ldots,\alpha_N$ are the eigenvalues of $\mathbf{W}$. Thus the eigenvalues of $\mathbf{\Gamma}$ are $1+\theta\alpha_1, 1+\theta\alpha_2, \ldots, 1+\theta\alpha_N$. The eigenvalues of the cases under **B** are listed in Table 4.1.

4.2.2 Regular rectangular lattice with holes

For the symmetric first order $S_{\rho_1\rho_1 00}$ in Figure 4.1 (1a) and second order $S_{\rho_1\rho_1\rho_1\rho_1}$ in Figure 4.1 (2a) schemes as an illustration, the correlation matrix $\mathbf{\Gamma}$

can be rewritten as $\mathbf{\Gamma}=\mathbf{I}+\rho\ \mathbf{W}$. Assume that $\lambda_1, \ldots, \lambda_N$ are the eigenvalues of $\mathbf{W}$ and the largest and smallest of λ_i's are respectively denoted by λ_{max} and λ_{min}. Thus the eigenvalues of $\mathbf{\Gamma}$ are $1+\rho\lambda_1, 1+\rho\lambda_2, \ldots, 1+\rho\lambda_N$. We discuss the following situations:

1. If $-1/\lambda_{max} < \rho < -1/\lambda_{min}, \lambda_{min} < 0$ and $\lambda_{max} > 0$, then all eigenvalues of $\mathbf{\Gamma}$ are positive and hence $\mathbf{\Gamma}$ is positive-definite. $\mathbf{W}$ is symmetric, and so is $\mathbf{\Gamma}$. But we want $-1 < \rho < 1$, so that the constraint is modified to $max(-1, -1/\lambda_{max}) < \rho < min(1, -1/\lambda_{min})$.

2. If $\rho < -1/\lambda_{min}, \lambda_{min} < 0$ and $\lambda_{max} < 0$, then $-1 < \rho < min(1, -1/\lambda_{min})$.

3. If $-1/\lambda_{max} < \rho, \lambda_{min} > 0$ and $\lambda_{max} > 0$, then $max(-1, -1/\lambda_{max}) < \rho < 1$.

For the other schemes, when the correlation matrix is a sparse matrix, grouping eigenvalues with similar values will keep problems manageable. But there is no closed form of the constraints for positive definiteness for the regular lattice data with holes.

4.3 Method of Moment Estimation

In the Phase 1 model (4.1), we only consider $\mu_i = \mu$ for Y_i, so there are at most 6 parameters, namely μ, σ^2, ρ_1, ρ_2, ρ_3, ρ_4 to be estimated for case (2e) under **A** in Section 4.1. We discuss two methods of estimation for those 6 or fewer parameters.

A. There are several methods of parameter estimation, but the method of moment estimation (MOME) is commonly used. We first use MOME. Estimates of ρ_i's change from (1a) to (2e). The MOME estimates for the correlation schemes (2e) with all four of the ρ_i's different in the correlation matrix $\mathbf{\Gamma} = \mathbf{I} + \rho_1 \mathbf{H} +$

$\rho_2 \mathbf{V} + \rho_3 \mathbf{C}^{(1)} + \rho_4 \mathbf{C}^{(2)}$, are defined by

$$\hat{\mu} = \overline{Y} := \frac{1}{N} \sum_{s_1=1}^{m} \sum_{s_2=1}^{n_{s_1}} Y_{(s_1,s_2)} \tag{4.4}$$

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{s_1=1}^{m} \sum_{s_2=1}^{n_{s_1}} [Y_{(s_1,s_2)} - \overline{Y}]^2 \tag{4.5}$$

$$\hat{\rho}_1 = \frac{(\mathbf{Y} - \overline{Y}\mathbf{1})'\mathbf{H}(\mathbf{Y} - \overline{Y}\mathbf{1})}{\hat{\sigma}^2 \sum_{i=1}^{N} \sum_{j=1}^{N} H_{ij}} \equiv \frac{\hat{\gamma}_1}{\hat{\sigma}^2} \tag{4.6}$$

$$\hat{\rho}_2 = \frac{(\mathbf{Y} - \overline{Y}\mathbf{1})'\mathbf{V}(\mathbf{Y} - \overline{Y}\mathbf{1})}{\hat{\sigma}^2 \sum_{i=1}^{N} \sum_{j=1}^{N} V_{ij}} \equiv \frac{\hat{\gamma}_2}{\hat{\sigma}^2} \tag{4.7}$$

$$\hat{\rho}_{i'} = \frac{(\mathbf{Y} - \overline{Y}\mathbf{1})'\mathbf{C}^{(i'-2)}(\mathbf{Y} - \overline{Y}\mathbf{1})}{\hat{\sigma}^2 \sum_{i=1}^{N} \sum_{j=1}^{N} C_{ij}^{(i'-2)}} \equiv \frac{\hat{\gamma}_i'}{\hat{\sigma}^2},\ i' = 3,\ 4 \tag{4.8}$$

where n_i's are the number of sample points for $s_1 = 1, \ldots, m$, $N = \sum_{i=1}^{m} n_i$, $\mathbf{1} = (1, \ldots, 1)'$, $\mathbf{H}$, and $\mathbf{V}$ respectively denotes (0,1) matrices for vertical, horizontal neighbors, and $\mathbf{C}^{(1)}$ and $\mathbf{C}^{(2)}$ denote the corresponding matrices for corner neighbors in the directions {(1,1),(-1,-1)} and {(1,-1),(-1,1)}.

Note that the expected values denoted by $\mathcal{E}$ are:

$$\mathcal{E}(\hat{\mu}) = \frac{1}{N} \sum_{s_1=1}^{m} \sum_{s_2=1}^{n_{s_1}} \mathcal{E}[Y_{(s_1,s_2)}] = \frac{1}{N} \sum_{s_1=1}^{m} \sum_{s_2=1}^{n_{s_1}} (\mu) = \mu$$

$$\mathcal{E}(\hat{\sigma}^2) = \sigma^2 \left(1 - \frac{A}{N^2}\right)$$

$$\mathcal{E}(\hat{\gamma}_1) = \rho_1 \sigma^2 + \frac{\sigma^2}{\sum_{i=1}^{N} \sum_{j=1}^{N} H_{ij}} \left[\frac{\sum_{i=1}^{N} \sum_{j=1}^{N} H_{ij}}{N^2} A - \frac{2}{N} A^{(1)}\right]$$

$$\mathcal{E}(\hat{\gamma}_2) = \rho_2 \sigma^2 + \frac{\sigma^2}{\sum_{i=1}^{N} \sum_{j=1}^{N} V_{ij}} \left[\frac{\sum_{i=1}^{N} \sum_{j=1}^{N} V_{ij}}{N^2} A - \frac{2}{N} A^{(2)}\right]$$

$$\mathcal{E}(\hat{\gamma}_{i'}) = \rho_{i'} \sigma^2 + \frac{\sigma^2}{\sum_{i=1}^{N} \sum_{j=1}^{N} C_{ij}^{(i'-2)}} \left[\frac{\sum_{i=1}^{N} \sum_{j=1}^{N} C_{ij}^{(i'-2)}}{N^2} A - \frac{2}{N} A^{(i)}\right],\ i' = 3, 4.$$

where

$$
\begin{aligned}
A &= N + \rho_1 \sum_i \sum_j H_{ij} + \rho_2 \sum_i \sum_j V_{ij} + \rho_3 \sum_i \sum_j C_{ij}^{(1)} + \rho_4 \sum_i \sum_j C_{ij}^{(2)} \\
A^{(1)} &= \sum_{i=1}^{N} \sum_{k=1}^{N} H_{ik} A_{i.}^{*} \\
A^{(2)} &= \sum_{i=1}^{N} \sum_{k=1}^{N} V_{ik} A_{i.}^{*} \\
A^{(i'-2)} &= \sum_{i=1}^{N} \sum_{k=1}^{N} C_{ik}^{(i'-2)} A_{i.}^{*},\ i' = 3,\ 4, \\
A_{i.}^{*} &= 1 + \rho_1 \sum_{j=1}^{N} H_{ij} + \rho_2 \sum_{j=1}^{N} V_{ij} + \rho_3 \sum_{j=1}^{N} C_{ij}^{(1)} + \rho_4 \sum_{j=1}^{N} C_{ij}^{(2)}
\end{aligned}
$$

As an illustration of the derivation of these expectations,

$$
\begin{aligned}
\mathcal{E}[\hat{\sigma}^2] &= \frac{1}{N} \mathcal{E} \left\{ \sum_{s_1=1}^{m} \sum_{s_2=1}^{n_{s_1}} \left[Y_{(s_1,s_2)} - \overline{Y} \right]^2 \right\} \\
&= \frac{1}{N} \mathcal{E} \left\{ \sum_{i=1}^{N} \left[(Y_i - \mu) - (\overline{Y} - \mu) \right]^2 \right\} \\
&= \frac{1}{N} \left\{ \sum_{i=1}^{N} \mathcal{E} \left[(Y_i - \mu)^2 \right] - N \mathcal{E} \left[(\overline{Y} - \mu)^2 \right] \right\} \\
&= \sigma^2 - \frac{1}{N^2} \mathcal{E} \left\{ \sum_{i=1}^{N} (Y_i - \mu)^2 + \sum_{i=1}^{N} \sum_{i \neq j}^{N} (Y_i - \mu)(Y_j - \mu) \right\} \\
&= \sigma^2 - \frac{1}{N^2} (**)
\end{aligned}
$$

and

$$
\begin{aligned}
\mathcal{E}[(**)] &= \sum_{i=1}^{N} \mathcal{E} \left[(Y_i - \mu)^2 \right] + \sum_{i=1}^{N} \sum_{i \neq j}^{N} \mathcal{E} \left[(Y_i - \mu) H_{ij} (Y_j - \mu) \right] \\
&\quad + \sum_{i=1}^{N} \sum_{i \neq j}^{N} \mathcal{E} \left[(Y_i - \mu) V_{ij} (Y_j - \mu) \right] + \sum_{i=1}^{N} \sum_{i \neq j}^{N} \mathcal{E} \left[(Y_i - \mu) C_{ij}^{(1)} (Y_j - \mu) \right] \\
&\quad + \sum_{i=1}^{N} \sum_{i \neq j}^{N} \mathcal{E} \left[(Y_i - \mu) C_{ij}^{(2)} (Y_j - \mu) \right] \\
&= \sigma^2 \left[N + \rho_1 \sum_i \sum_j H_{ij} + \rho_2 \sum_i \sum_j V_{ij} + \rho_3 \sum_i \sum_j C_{ij}^{(1)} + \rho_4 \sum_i \sum_j C_{ij}^{(2)} \right] \\
&= \sigma^2 A,
\end{aligned}
$$

$$\begin{aligned}
\mathcal{E}[\hat{\gamma}_1] &= \mathcal{E}\left[\frac{(\mathbf{Y}-\overline{Y}\mathbf{1})'\mathbf{H}(\mathbf{Y}-\overline{Y}\mathbf{1})}{\sum_{i=1}^{N}\sum_{j=1}^{N}H_{ij}}\right] \\
&= \frac{1}{\sum_{i=1}^{N}\sum_{j=1}^{N}H_{ij}}\mathcal{E}[(\mathbf{Y}-\mu\mathbf{1})'\mathbf{H}(\mathbf{Y}-\mu\mathbf{1}) - 2(\mathbf{Y}-\mu\mathbf{1})'\mathbf{H}(\overline{Y}\mathbf{1}-\mu\mathbf{1}) \\
&\quad +(\overline{Y}\mathbf{1}-\mu\mathbf{1})'\mathbf{H}(\overline{Y}\mathbf{1}-\mu\mathbf{1})] \\
&= \frac{1}{\sum_{i=1}^{N}\sum_{j=1}^{N}H_{ij}}\mathcal{E}[D_1 - 2\,D_2 + D_3]
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{E}[D_1] &= \sum_{i=1}^{N}\sum_{j=1}^{N}\mathcal{E}\left[(Y_i-\mu)H_{ij}(Y_j-\mu)\right] \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N}H_{ij}\rho_1\sigma^2 \\
\mathcal{E}[D_2] &= \mathcal{E}\left\{\frac{1}{N}\left[\sum_{i=1}^{N}\sum_{k=1}^{N}H_{ik}(Y_i-\mu)\right]\left[\sum_{j=1}^{N}(Y_j-\mu)\right]\right\} \\
&= \frac{1}{N}\left\{\sum_{i=1}^{N}\sum_{k=1}^{N}H_{ik}\mathcal{E}[(Y_i-\mu)^2] + \sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{k=1}^{N}H_{ik}\mathcal{E}[(Y_i-\mu)(Y_j-\mu)]\right\}
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{N}\{\sum_{i=1}^{N}\sum_{k=1}^{N}H_{ik}(\sigma^2) + \sum_{i=1}^{N}\sum_{k=1}^{N}H_{ik}(\rho_1\sigma^2\sum_{j}H_{ij} + \rho_2\sigma^2\sum_{j}V_{ij} + \rho_3\sigma^2\sum_{j}C_{ij}^{(1)} \\
&\quad +\rho_4\sigma^2\sum_{j}C_{ij}^{(2)})\} \\
&= \frac{\sigma^2}{N}A^{(1)} \\
\mathcal{E}[D_3] &= \frac{\sum_{k=1}^{N}\sum_{l=1}^{N}H_{kl}}{N^2}\mathcal{E}\left\{\left[\sum_{i=1}^{N}(Y_i-\mu)\right]\left[\sum_{j=1}^{N}(Y_j-\mu)\right]\right\} \\
&= \frac{\sum_{k=1}^{N}\sum_{l=1}^{N}H_{kl}}{N^2}\mathcal{E}[(**)] \\
&= \frac{\sum_{k=1}^{N}\sum_{l=1}^{N}H_{kl}}{N^2}\sigma^2 A
\end{aligned}$$

The approximations for $\mathcal{E}(\hat{\rho}_i)$, $i = 1,\ldots,4$, using a first order Taylor series expansion are

$$\mathcal{E}(\hat{\rho}_i) \simeq \frac{\mathcal{E}(\hat{\gamma}_i)}{\mathcal{E}(\hat{\sigma}^2)} = \rho_i + \frac{b(\hat{\gamma}_i) - \rho_i\, b(\hat{\sigma}^2)}{\sigma^2 + b(\hat{\sigma}^2)} \equiv \rho_i + b_1(\hat{\rho}_i) \tag{4.9}$$

where the subscript 1 of $b_1(\hat{\rho}_i)$, $i = 1, \ldots, 4$, denotes the first order Taylor series approximation to bias of $\hat{\rho}_i$.

In Eqns. (4.6) - (4.8), we used the overall mean in the covariance product. Next, we use the individual directional means in the product. An alternative MOME approach to $\hat{\gamma}_i$, $i = 1, \ldots, 4$ for cases **A** resulted in

$$\begin{aligned}
\tilde{\gamma}_1' &= \frac{(\mathbf{Y} - \overline{Y}_{(1)}\mathbf{1})'\mathbf{H}(\mathbf{Y} - \overline{Y}_{(1)}\mathbf{1})}{\sum_{i=1}^{N}\sum_{j=1}^{N} H_{ij}} \\
\tilde{\gamma}_2' &= \frac{(\mathbf{Y} - \overline{Y}_{(2)}\mathbf{1})'\mathbf{V}(\mathbf{Y} - \overline{Y}_{(2)}\mathbf{1})}{\sum_{i=1}^{N}\sum_{j=1}^{N} V_{ij}} \\
\tilde{\gamma}_{i'}' &= \frac{(\mathbf{Y} - \overline{Y}_{(i)}\mathbf{1})'\mathbf{C}^{(i'-2)}(\mathbf{Y} - \overline{Y}_{(i)}\mathbf{1})}{\sum_{i=1}^{N}\sum_{j=1}^{N} C^{(i'-2)}}, \; i' = 3, 4.
\end{aligned}$$

where

$$\begin{aligned}
\overline{Y}_{(1)} &= \frac{\mathbf{1H\,Y}}{\sum_{i=1}^{N}\sum_{j=1}^{N} H_{ij}} \\
\overline{Y}_{(2)} &= \frac{\mathbf{1V\,Y}}{\sum_{i=1}^{N}\sum_{j=1}^{N} V_{ij}} \\
\overline{Y}_{(i')} &= \frac{\mathbf{1C}^{(i'-2)}\,\mathbf{Y}}{\sum_{i=1}^{N}\sum_{j=1}^{N} C_{ij}^{(i'-2)}},, \; i' = 3, 4.
\end{aligned}$$

Thus

$$\hat{\rho}_i' = \frac{\hat{\gamma}_i'}{\hat{\sigma}^2}, \; i = 1, \ldots, 4.$$

where $\hat{\sigma}^2$ is defined as before.

B. In the following , we investigate the MOME for cases **B** with the alternative correlation matrix $\mathbf{\Gamma}$=$\mathbf{I}$+ θ $\mathbf{W}$ for all cases, where the w_{ij}'s for $i \neq j$ might be d_{ij}^{-k} or $exp\,(-d_{ij}^{k})$ for $k = 1, 2$ with and without first and second order neighborhood constraints. Once again, the alternative correlation matrix does not affect the estimation of μ and σ^2. The moment estimator of θ is

$$\hat{\theta} = \frac{(\mathbf{Y} - \overline{Y}\mathbf{1})'\mathbf{W}(\mathbf{Y} - \overline{Y}\mathbf{1})}{\hat{\sigma}^2 \sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}} \tag{4.10}$$

where $w_{ij} = w_{ji}$ and w_{ij}=0 unless plots i and j are neighbors via neighborhood specifications and w_{ij}=d_{ij}^{-k} or $exp(-d_{ij}^{k})$ if plots i and j are neighbors.

C. Simulations: To assess some of these results, we conducted a simulation of a population of a 10x10 lattice of points with $\mu = 5$, $\sigma^2 = 4$, $\rho_1 = 0.3$, $\rho_2 = 0.15$, $\rho_3 = 0.02$, $\rho_4 = 0.01$. The simulation gave the means of simulated $\hat{\rho}_i$'s to be (0.2811,0.1302,-0.0012,-0.0011) for the previous MOMEs and (0.2805,0.1294,-0.0027,-0.0126) for the alternative MOMEs. The absolute values of bias for the $\hat{\rho}_i$ from the alternative MOME approach are larger than those from the previous MOME. As mentioned in Chapter 3 for sample correlations, the alternative approach for the cases under **A** is slightly worse than the original one since less information is used in estimating μ. Therefore, we will stay with the original MOME which uses the overall mean in the product of covariances for cases **A** (see Eqns. 4.6-6.8).

The moment estimations for other correlation schemes are similar to that for scheme $S_{\rho_1\rho_2\rho_3\rho_4}$ but the number of parameters are fewer. The estimators of μ and σ^2 are based on all observations with equal weights and ignore distances (see Eqns. (4.4) and (4.5)). Hence, no benefit is derived from the different correlation schemes and so the estimates of those two yield the same ones for all schemes. As to the estimates of ρ_i, the moment estimations are modified to accommodate the same correlations in different directions. For example, if horizontal and vertical neighbors are assumed to have the same effect on the neighboring predictions, then the spatial dependence coefficients between them are equal and the matrix **H** in Eqn. (4.6) and the matrix **V** in Eqn. (4.7) are replaced by the matrix $\mathbf{W}^{(1a)} = \mathbf{H}+\mathbf{V}$. In so doing, more information is taken into account for estimating one spatial correlation parameter in horizontal and vertical neighbors instead of two parameters. Therefore, the estimate of the spatial correlation for horizontal and vertical neighbors is expected to be more efficient. On the other hand, if

effects of vertical and horizontal neighbors are not equal, it is not wise to assume them to be equal. It is best to treat all 4 neighbors as having different effects on neighboring predictions and then merge some neighborhood matrices in a sensible way. Since the estimates of those two directional corner neighbors are generally close, we assumed them to be equal. Scheme $S_{\rho_1\rho_2\rho_3\rho_3}$ assumes that $\rho_4 = \rho_3$ and so the estimated ρ_3 for the neighborhood matrix is

$$\hat{\rho}_3 = \frac{(\mathbf{Y} - \overline{Y}\mathbf{1})'\mathbf{C}(\mathbf{Y} - \overline{Y}\mathbf{1})}{\hat{\sigma}^2 \sum_{i=1}^{N}\sum_{j=1}^{N} C_{ij}} \equiv \frac{\hat{\gamma}_3}{\hat{\sigma}^2} \tag{4.11}$$

where $\mathbf{C}=\mathbf{C}^{(1)} + \mathbf{C}^{(2)}$ denotes (0,1) matrices for corner neighbors.

$$\mathcal{E}(\hat{\gamma}_3) = \rho_3\sigma^2 + \frac{\sigma^2}{\sum_{i=1}^{N}\sum_{j=1}^{N} C_{ij}}\left[\frac{\sum_{i=1}^{N}\sum_{j=1}^{N} C_{ij}}{N^2}A - \frac{2}{N}A^{(3*)}\right].$$

Moment estimation is the simplest method for estimating parameters, but it can fail to produce a positive-definite $\mathbf{\Gamma}$. To overcome the problem and to obtain more efficient estimators, maximum pseudo-likelihood estimators will be derived in the following section.

4.4 Maximum Pseudo-Likelihood Estimators (MPLE)

To obtain maximu likelihood estimation, we assume that the N observations $(Y_1, \ldots, Y_N)'$ have been drawn from a multivariate normal distribution, MVN[$\boldsymbol{\mu}_{N\times 1}$, $\sigma^2\ \mathbf{\Gamma}$]. Then the Gaussian likelihood of $\mathbf{Y} = (Y_1, \ldots, Y_N)'$, under Phase 1 assumptions, is

$$L(\mu, \sigma^2, \rho_i; i = 1, \ldots, 4) = (2\pi\sigma^2)^{-N/2} \mid \mathbf{\Gamma} \mid^{-1/2} exp\left[-\frac{1}{2\sigma^2}(\mathbf{Y} - \mu\mathbf{1})'\mathbf{\Gamma}^{-1}(\mathbf{Y} - \mu\mathbf{1})\right]$$

where $\mathbf{\Gamma} = \mathbf{I} + \rho_1\,\mathbf{H} + \rho_2\,\mathbf{V} + \rho_3\,\mathbf{C}^{(1)} + \rho_4\,\mathbf{C}^{(2)}$. Then the log-likelihood of $\mathbf{Y}$ is

$$\begin{aligned} l(\mu, \sigma^2, \rho_i; i = 1, \ldots, 4) &= log\left(L(\mu, \sigma^2, \rho_i; i = 1, \ldots, 4)\right) \\ &= -\frac{N}{2}log(2\pi\sigma^2) - \frac{1}{2}log(\mid \mathbf{\Gamma} \mid) - \frac{1}{2\sigma^2}(\mathbf{Y} - \mu\mathbf{1})'\mathbf{\Gamma}^{-1}(\mathbf{Y} - \mu\mathbf{1}) \end{aligned}$$

Maximizing with respect to μ and σ^2, we find that

$$\tilde{\mu}_{MLE} = \frac{\mathbf{1}'\boldsymbol{\Gamma}^{-1}\mathbf{Y}}{\mathbf{1}'\boldsymbol{\Gamma}^{-1}\mathbf{1}} \text{ and } \tilde{\sigma}^2_{MLE} = \frac{(\mathbf{Y}-\mu\mathbf{1})'\boldsymbol{\Gamma}^{-1}(\mathbf{Y}-\mu\mathbf{1})}{N}$$

The log-likelihood then reduces to

$$l(\tilde{\mu}, \tilde{\sigma}^2, \rho_i; i = 1, \ldots, 4) = -\frac{N}{2} log(2\pi\tilde{\sigma}^2) - \frac{1}{2} log(\mid \boldsymbol{\Gamma} \mid) - \frac{N}{2}$$

The exact maximum likelihood estimation for the ρ_i's require evaluation of the determinant and the inverse of the correlation matrix $\boldsymbol{\Gamma}$ and it is difficult to get a closed form for the estimators of the parameters, especially for lattice data with holes. However, estimates can be obtained using numerical optimization. The Newton-Raphson method is used. The moment estimates are used as an initial starting guess if they succeed in getting a positive-definite correlation matrix; otherwise we use $\mathbf{0}$ as a starting point. Then continue iterating until the likelihood is maximized with a positive-definite correlation matrix constraint.

Let H_0 and H_a denote two hypotheses where H_0: all ρ_i's are zero vs. H_a: not all ρ_i's are zero. Thus the likelihood ratio (LR) statistic for testing the hypothesis is

$$\text{likelihood ratio } = \frac{L(\mathbf{0}, \hat{\mu}_{MLE}, \hat{\sigma}^2_{MLE})}{L(\hat{\rho}_i; i = 1, \ldots, k, \hat{\mu}_{MLE}, \hat{\sigma}^2_{MLE})} \equiv lrt.$$

where $\hat{\mu}_{MLE}, \hat{\sigma}^2_{MLE}$, and $\hat{\rho}_i, i = 1, \ldots, k$, denote the ML estimators under H_0. If the null hypothesis for the LR test is a model of independence, ρ_i's are zeros, then $-2\,log(lrt)$ is χ^2 distributed with the number of degrees of freedom given by additional parameters estimated under H_a.

For large samples, under H_0, $-2\,log(lrt) \sim \chi^2_k$, where k is the number of quantities jointly estimated. At level α, if $\alpha > p$-value $= Pr[-2\,log(lrt) > \chi^2_k(1-\alpha)]$, then H_0 is rejected. There is some evidence that this approximation is valid for small lattice and non-lattice situations under different models. Haining (1977) using simulation methods, constructed a first order symmetric SAR model against

a model of independence. Haining (1978) compared a moving average model against a model of independence. His simulation results suggest that the χ^2 approximation provides a reasonable guide for lattice situations. Brandsma and Ketellaper (1979), however, suggest that the χ^2 approximation is not satisfied for non-lattice situations.

4.5 Results and Discussions

To see the bias and standard deviation of MOMEs and MPLEs, a simulation study was conducted. We also apply the moment estimation and maximum pseudo-likelihood estimation to real data for the Siuslaw Forest.

4.5.1 Results for simulated data for artificial or real locations

A simulation study was conducted to examine the behavior of moment estimation for parameters. Table 4.3 displays the means, standard deviations and bias of estimated parameters for 10,000 realizations from the process with the true values of $\mu = 5, \sigma^2 = 4, \rho_1 = 0.4, \rho_2 = 0.2, \rho_3 = 0.05$. The bias approximation is not good in the cases of either 5 artificial (integer-indexed coordinates of sites: (1,1), (1,2), (2,1), (2,2), (3,1)) or real point locations (the first five sample plots in the Siuslaw Forest). As the number of sample location points increase, the bias approximation works better. We also calculated the bias of $\hat{\gamma}_i$ and $\hat{\rho}_i$ for the 10,000 realizations. From Table 4.4, we note that the bias of $\hat{\gamma}_i$ is close to the true bias calculated by the formula, but the $b_1(\hat{\rho}_i)$ bias approximations do not work well for small lattices.

For comparison of MOME and MPLE methods, 6,000 realizations for different size lattices and different parameter values were generated and then for each realization moment estimates and maximum pseudo-likelihood estimates are shown in Table 4.5.

So far we only studied simulated data based on artificial or real locations in the Siuslaw National Forest. In the following subsection we deal with real data.

4.5.2 Real data

There is an isolated plot in the Forest which is on a little bit of Forest land near the town of Dallas, Oregon, not near the rest of the Forest. The total neighbors in the neighborhood matrices for the whole Siuslaw area (313 plots) are respectively $\sum_{i=1}^{N}\sum_{j=1}^{N} H_{ij} = 414$, $\sum_{i=1}^{N}\sum_{j=1}^{N} V_{ij} = 400$, $\sum_{i=1}^{N}\sum_{j=1}^{N} C_{ij} = 752$, $\sum_{i=1}^{N}\sum_{j=1}^{N} C_{ij}^{(1)} = 374$, $\sum_{i=1}^{N}\sum_{j=1}^{N} C_{ij}^{(2)} = 378$. In Section 3.5, we pointed out that for Tba data there might be stronger sample spatial dependence for adjacent corner neighbors than that for adjacent horizontal and vertical neighbors. Table 3.4 shows the sample correlations among adjacent horizontal, vertical neighbors, and corner neighbors. We can see that there exists a stronger sample spatial dependence among corner neighbors. This section examines spatial dependence further using spatial model (4.3) with different correlation schemes. Next we apply the moment and maximum pseudo-likelihood estimations to the Tba data fitted by model (4.3).

We examine the model in two ways, with and without the isolated plot. The results for moment estimation are summarized in Tables 4.6-4.11. The differences in results between Tba data with and without the isolated plot are hardly noticed. That's because there's no neighbors in the neighborhood of the isolated plot under our neighborhood structure. Hence including the observation doesn't affect the results much. From Table 4.10, we can see that in scheme $S_{\rho_1\rho_2\rho_3\rho_4}$ with $N = 312$ the estimates of ρ_i are respectively $\hat{\rho}_1 = -0.0146, \hat{\rho}_2 = 0.0320, \hat{\rho}_3 = 0.1802$, and $\hat{\rho}_4 = 0.1170$. As mentioned in Chapter 3 for sample correlations, spatial coefficients ρ_c's for the corner neighbors are larger than for the horizontal and vertical neighbors. An important property of the matrix $\hat{\Gamma}$ is the positive definiteness

exhibited by all estimates of ρ_i for Tba by MOME in Table 4.10. From Table 4.6, under correlation schemes (2d) and (2e) we note that the MOME estimate of spatial dependence ρ_4 is close to zero and quite different from the other estimated ρ_i. Thus it might not be good to treat all ρ_i equal like we did in scheme (2a). Table 4.7 summarized the MOME estimates for M_s; we can see that $\hat{\rho}_1 = 0.0391$, $\hat{\rho}_2 = 0.0287$, $\hat{\rho}_3 = 0.0229$, and $\hat{\rho}_4 = 0.1220$. The $C^{(2)}$ corner neighbors have the strongest spatial dependence. With distance-related spatial dependence function we expected that farther neighbors have less correlation than closer neighbors. But the MOME results of M_s contradict this, suggesting that a spatial correlation function that decays with increasing distance may be invalid.

Table 4.12-4.17 summarized the result of PMLE for M, M_s, Lt, Lt_s, Tba, and Tba_s. In order to obtain PMLE, normality is assumed. For Tba, as shown in Table 4.16, there is a stronger spatial dependence among corner neighbors than the other directional neighbors. But the PMLE estimates of ρ_i's are only slightly different from the MOMEs in Table 4.10. We mentioned in Chapter 3 that Tba corner neighbors seem to have larger sample correlations than horizontal and vertical neighbors. As expected, Table 4.16 verifies this based on the MVN model. The PMLE estimates for M_s shown in Table 4.13 are similar to MOME estimates. Both MOME and PMLE for ρ_i's of M_s point out that the distance-related spatial dependence function doesn't work well for M_s data.

The results of distance-related correlation schemes (i.e. cases **B**) for Tba are summarized in Table 4.18. We note that the estimates of θ are exactly the same for (2a) for different k-values. This is because of equal distance between neighbors included in the neighborhood of scheme case **B**(2a). Similarly, the estimates of θs are the same for (4b) with $k = 1$ and $k = 2$. But the estimated correlation of first order neighbor (2a) is $\hat{\rho}_{ij} = \hat{\theta} d_{ij}^{-k}$, for $i \neq j$ $(k = 1, 2)$, and the maximum of $\hat{\rho}_{ij} = 0.0060$, the same as those of (4a). The estimates coincide with the estimates

of case **A**(1a), $S_{\rho_1\rho_1 00}$ (see Table 4.16 (1a)). Distance-related correlations don't benefit the estimation of correlations because the neighbors are equal-distance under schemes (2) and (4) with a 1st order neighborhood constraint. With distance-related correlations (1) and (3) of case **B** use all the neighbors, but combining one parameter with known distance functions compromises the balance of more information and simplicity of parameterizations. The alternative correlation schemes give less weights to far apart neighbors than to closer neighbors; which means more weights to horizontal and vertical neighbors than to corner neighbors. As mentioned earlier, for Tba the spatial correlation for corner neighbors are larger than those for the horizontal and vertical neighbors. A distance-related correlation is not compatible with this situation and so it might jeopardize the predictions for Tba.

4.5.2.1 Model selection

Arranging the observations in sequence is very important to minimize accumulated prediction errors, and is difficult to generalize in the spatial context. A reliable model selection criterion is crucial. For prediction purposes, a model is often chosen in the sense of the minimum root mean squared prediction error (RMSE). This can be achieved by the apparent prediction error, k-fold cross-validation, or by bootstrap methods such as the simple bootstrap, leave-one-out and .632 bootstrap. We demonstrate the estimates of the prediction errors for several criteria. Akaike (1978) and Schwarz (1978), introduced equivalent consistent model selection criteria conceived from a Bayesian perspective. Schwarz derived SIC for selecting models in the Koopman-Darmois family, whereas Akaike derived his model selection criteria BIC for the problem of selecting a model in linear regression. For time series, models are often selected using the Akaike information criterion (AIC), AICC in which the last C suffix indicates a correction

version of AIC, or Bayesian information criterion (BIC). The criteria are:

$$\begin{aligned}
\text{AIC } (\hat{\mu}, \hat{\sigma}^2) &= -2\, l(\hat{\mu}, \hat{\sigma}^2, \rho_i; i = 1, \ldots, k) + 2(k+1) \\
\text{AICC } (\hat{\mu}, \hat{\sigma}^2) &= -2\, l(\hat{\mu}, \hat{\sigma}^2, \rho_i; i = 1, \ldots, k) + 2(k+1)N/(N-k-2) \\
\text{BIC } (\hat{\mu}, \hat{\sigma}^2) &= -2\, l(\hat{\mu}, \hat{\sigma}^2, \rho_i; i = 1, \ldots, k) + (k+1)ln(N)
\end{aligned}$$

where $l(\hat{\mu}, \hat{\sigma}^2, \rho_i; i = 1, \ldots, k) = -\frac{N}{2} log(2\pi\hat{\sigma}^2) - \frac{1}{2} log(\mid \mathbf{\Gamma} \mid) - \frac{N}{2}$. One can think of the second terms in the three criteria above as penalty terms to discourage over-parameterization. If N is quite large, the penalties of AIC and AICC are about the same, but the BIC criterion penalizes over-parameterization more than the other two.

The following is a description of the approaches compared. The simple bootstrap approach generates b bootstrap samples, estimates the model for each and then applies each fitted model to the original sample to give b estimates of prediction error. The estimate of the overall prediction error is the average of these b estimates. The observations y_i's are treated as assumed i.i.d components, which is not necessarily so for spatial data. Ignoring significant correlation among spatial data could lead to chosing the wrong model by prevalent model selection criteria. If there is spatial dependence among the data, then the observations y_i's should not be treated as independent samples.

A semiparametric bootstrap method does consider spatial dependence among the data. Suppose spatial data follows the linear model

$$\mathbf{Y} = \boldsymbol{\mu} + B\boldsymbol{\varepsilon} \tag{4.12}$$

where $\mathbf{Y} = (Y_1, \ldots, Y_N)'$, $\boldsymbol{\mu}$=$(\mu_1, \ldots, \mu_N)'$, $\mathbf{\Gamma} = \mathbf{I} + \rho_1 \mathbf{H} + \rho_2 \mathbf{V} + \rho_3 \mathbf{C}^{(1)} + \rho_4 \mathbf{C}^{(2)} \equiv B'B$, $\boldsymbol{\varepsilon} \sim \text{MVN}(\mathbf{0}, \sigma^2\, \mathbf{I})$. The model can be expressed as

$$Y_i = \mu + \sum_{j \neq i}^{N} m_{ij}(Y_j - \mu) + \varepsilon_i, i = 1, \ldots, N.$$

where $B^{-1} \equiv \mathbf{I} - M$ and $M = (m_{ij})$.

If a symmetric and positive-definite estimator $\hat{\mathbf{\Gamma}}$ of $\mathbf{\Gamma}$ has been obtained from the data, it can be decomposed as the matrix product $\hat{\mathbf{\Gamma}} = \hat{B}'\hat{B}$. From (4.12), define

$$(\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_N) \equiv \hat{\boldsymbol{\varepsilon}} \equiv \hat{B}^{-1}(\mathbf{Y} - \boldsymbol{\mu}),$$

and $\tilde{\varepsilon}_i \equiv \hat{\varepsilon}_i - (\sum_{j=1}^{N} \hat{\varepsilon}_i/N), i = 1, \ldots, N$. Instead of resampling from the observations y_i's, the assumed i.i.d components $\tilde{\varepsilon}_i$ are bootstrapped. For each bootstrap sample $\boldsymbol{\varepsilon}^*$, transformation back to the y_i is calculated by

$$\mathbf{Y}^* = \hat{\boldsymbol{\mu}} + \hat{B}\boldsymbol{\varepsilon}^*.$$

Then the model is refitted to the bootstrap $\mathbf{Y}^*$ and the estimates from each bootstrap sample are applied to the original samples so the average of b prediction errors is adopted to estimate the overall prediction error.

In leave-one-out cross validation (1CV) which is k-fold cross validation with k=1, one observation is omitted at a time in turn and for each omission the fitted model is applied based on the remaining sample to predict the outcome. The average of errors made in these N predictions is the estimate of the prediction error.

For the simple bootstrap and semiparametric bootstrap methods, 30 bootstrap samples were generated to display the behavior of the different estimation procedure for different correlation matrix schemes. The means and standard deviations of the estimated parameters for 30 bootstrap samples were computed to get a better idea about the error range of the estimated parameters.

The objective of seeking improvement in RMSE, AIC, AICC, and BIC through bootstrapping was to obtain better estimates of those four statistics. Figure 4.2 shows the behavior of the different estimation procedures for different correlation schemes. The RMSEs of Tba fitted by spatial model (4.3) were estimated using

1-fold cross-validation for those 33 plots with all 8 neighbors. For Tba, the estimated RMSEs for model (4.3) with different correlation schemes ranged from 19.6 to 20.1. The model with correlation scheme $S_{\rho_1\rho_1\rho_3\rho_3}$ has the smallest RMSE. For the graphs of AIC, AIC, and BIC, we note that the values drop abruptly when the neighborhood distance increases up to the second order neighborhood. This implies the second order neighborhood works better than the first order neighborhood. Due to overparameterization, the values for correlation scheme $S_{\rho_1\rho_2\rho_3\rho_4}$ go up again. To balance reduction of RMSE and simplicity of models, the model (4.3) with correlation scheme $S_{\rho_1\rho_1\rho_2\rho_2}$ was selected for the Tba data.

To estimate prediction errors of individual plots, we applied simple bootstrap and semiparametric bootstrap methods to the suggested model. The results based on 400 bootstrap samples are depicted in Figure 4.5 and the $\circ$ and $\triangle$ indicate respectively the ratios from semiparametric bootstrap and simple bootstrap methods. The upper graph in Figure 4.5 is the ratios of sample deviation of the prediction errors relative to the sample mean of the prediction errors for each plot. The $\triangle$'s are hidden in the band on Figure 4.5. This implies the ratios from simple bootstrap samples are close to 1 mostly. We can see that several $\circ$'s fall outside the band. So the semiparametric bootstrap method is more sensitive to detect the prediction errors for some hard-to-predict plots. The lower graph in Figure 4.5 is the ratios of sample deviation of the $b = 400$ prediction errors relative to the sample mean of the $b = 400$ predictions for each plot.

For seedlings, there are many zeros among data for several variables. Such data with numerous zeros need to be handled differently. For Tba_s, a possible explanation for the large number of zeros might be the fact that lots of plots don't have any seedlings with a DBH of 1.0" or more. This phenomenon can be handled by a two-component mixture distribution with one degenerated at 0 and the other non-negative distribution. In a subsequent chapter, we will incorporate spatial

dependence in the mixture models for count data such as mortality, mortality of seedlings, number of large live trees (Tba), and Tba of seedlings.

4.5.2.2 Conclusions

The simple spatial model distributional assumption: it models only first and second order moments. Yet it is adequate to enable us to define a simple predictor introduced in Chapter 5. Among the six responses, all except for Tba are clearly nonnormal. Various transformations, such as power and three-parameter lognormal were tried on the data with limited success. Data that is mixed in the sense of being partly discrete and partly continuous can not be transformed so as to be totally continuous, so it is not possible to transform several of our response variables, which have significant mass of zeros, to normality. The modeling of such mixed data is delayed until Chapter 6.

Model selection criteria are usually based on an assumed distributional model. The distributional model assumed here was the multivariate normal. Under such assumption we were able to obtain pseudo maximum likelihood estimators, do some model selection among competing spatial models, and assess prediction errors.

We found for the Tba data, there is not much difference among the RMSEs obtained from various model selection criteria. Therefore, we chose the model with the minimum AICC value for Tba data, i.e. the model with the correlation matrix scheme $S_{\rho_1\rho_1\rho_3\rho_3}$, $\rho_1 = \rho_2 = 0.0027$ and $\rho_3 = \rho_4 = 0.1332$. To obtain a reliable prediction error for each 1-ha plot, we used two bootstrap methods, a simple bootstrap and a semiparametric bootstrap based on the selected model. Since there exists second-order spatial dependence among Tba data, the simple bootstrap may be not reliable. As shown in Figure 4.5, in spite of the fact that the simple bootstrap prediction errors for each plot are lower than those of the

semiparametric bootstrap method, the semiparametric bootstrap prediction errors are recommended since it allows for spatial dependence among the data.

For the other response variables there is little difference in the prediction errors between the simple bootstrap and semiparametric bootstrap methods. This is probably ture because there is little spatial dependence. Hence it is maybe ok to use the simple bootstrap method since it is less time-consuming than the semiparametric bootstrap method.

In summary, the models discussed in this chapter used different correlation schemes based on lattice data with or without holes. Since the non-distance based models were based on the 1.7-mile plot data, they don't allow for prediction of the response variable for non-sampled plots, for example on the 0.85-mile grid. Such models can only make predictions for individual plot responses for non-sampled plots on the 1.7-mile grid. Distance-related correlations can be used to overcome this problem, we investigate distance-related correlation schemes in the next chapter.

Figure 4.1: The correlation matrix schemes

$$\begin{matrix} & \rho_{1\dot{a}} & \\ \rho_{1\dot{a}} & \times & \rho_{1\dot{a}} \\ & \rho_{1\dot{a}} & \end{matrix}$$

(1a). symmetric first order scheme $(S_{\rho_1\rho_1 00})$

$$\begin{matrix} & \rho_{2} & \\ \rho_{1} & \times & \rho_{1} \\ & \rho_{2} & \end{matrix}$$

(1b). asymmetric first order scheme $(S_{\rho_1\rho_2 00})$

$$\begin{matrix} \rho_{2\dot{a}} & \rho_{2\dot{a}} & \rho_{2\dot{a}} \\ \rho_{2\dot{a}} & \times & \rho_{2\dot{a}} \\ \rho_{2\dot{a}} & \rho_{2\dot{a}} & \rho_{2\dot{a}} \end{matrix}$$

(2a). symmetric second order scheme $(S_{\rho_1\rho_1\rho_1\rho_1})$

$$\begin{matrix} \rho_{2\dot{b}} & \rho_{1\dot{a}} & \rho_{2\dot{b}} \\ \rho_{1\dot{a}} & \times & \rho_{1\dot{a}} \\ \rho_{2\dot{b}} & \rho_{1\dot{a}} & \rho_{2\dot{b}} \end{matrix}$$

(2b). asymmetric second order scheme $(S_{\rho_1\rho_1\rho_3\rho_3})$

$$\begin{matrix} \rho_{2\dot{c}} & \rho_{2} & \rho_{2\dot{c}} \\ \rho_{1} & \times & \rho_{1} \\ \rho_{2\dot{c}} & \rho_{2} & \rho_{2\dot{c}} \end{matrix}$$

(2c). asymmetric second order scheme $(S_{\rho_1\rho_2\rho_3\rho_3})$

$$\begin{matrix} \rho_{4} & \rho_{2\dot{d}} & \rho_{3} \\ \rho_{2\dot{d}} & \times & \rho_{2\dot{d}} \\ \rho_{3} & \rho_{2\dot{d}} & \rho_{4} \end{matrix}$$

(2d). asymmetric second order scheme $(S_{\rho_1\rho_1\rho_3\rho_4})$

$$\begin{matrix} \rho_{4} & \rho_{2} & \rho_{3} \\ \rho_{1} & \times & \rho_{1} \\ \rho_{3} & \rho_{2} & \rho_{4} \end{matrix}$$

(2e). asymmetric second order scheme $(S_{\rho_1\rho_2\rho_3\rho_4})$

Table 4.1: The eigenvalues of $\mathbf{\Gamma}$ for $m \times n$ lattices

	Eigenvalues λ_{ij}
$S_{\rho_1\rho_1 00}$	$1+2\rho_{1a}[cos(\frac{i\pi}{m+1})+cos(\frac{j\pi}{n+1})]$
$S_{\rho_1\rho_2 00}$	$1+2[\rho_1 cos(\frac{i\pi}{m+1})+\rho_2 cos(\frac{j\pi}{n+1})]$
$S_{\rho_1\rho_1\rho_1\rho_1}$	$1+2\rho_{2a}[cos(\frac{i\pi}{m+1})+cos(\frac{j\pi}{n+1})+2cos(\frac{i\pi}{m+1})cos(\frac{j\pi}{n+1})]$
$S_{\rho_1\rho_1\rho_3\rho_3}$	$1+2[\rho_{1a}(cos(\frac{i\pi}{m+1})+cos(\frac{j\pi}{n+1}))+2\rho_{2b}cos(\frac{i\pi}{m+1})cos(\frac{j\pi}{n+1})]$
$S_{\rho_1\rho_2\rho_3\rho_3}$	$1+2[\rho_1 cos(\frac{i\pi}{m+1})+\rho_2 cos(\frac{j\pi}{n+1})+2\rho_c cos(\frac{i\pi}{m+1})cos(\frac{j\pi}{n+1})]$
$S_{\rho_1\rho_1\rho_3\rho_4}$	$1+2[\rho_{2d}(cos(\frac{i\pi}{m+1})+cos(\frac{j\pi}{n+1}))+\rho_3 cos(\frac{i\pi}{m+1}+\frac{j\pi}{n+1})+\rho_4 cos(\frac{i\pi}{m+1}-\frac{j\pi}{n+1})]$
$S_{\rho_1\rho_2\rho_3\rho_4}$	$1+2[\rho_1(cos(\frac{i\pi}{m+1})+cos(\frac{j\pi}{n+1}))+\rho_3 cos(\frac{i\pi}{m+1}+\frac{j\pi}{n+1})+\rho_4 cos(\frac{i\pi}{m+1}-\frac{j\pi}{n+1})]$

Table 4.2: The eigenvalues of alternative correlation schemes of $\mathbf{\Gamma}$ for $m \times n$ lattices

	Eigenvalues λ_{ij}
(1) $k=1$	no closed form
(1) $k=2$	no closed form
(2a) $k=1$	$1+\frac{2}{1.7}\theta[cos(\frac{i\pi}{m+1})+cos(\frac{j\pi}{n+1})]$
(2a) $k=2$	$1+\frac{2}{(1.7)^2}\theta[cos(\frac{i\pi}{m+1})+cos(\frac{j\pi}{n+1})]$
(2b) $k=1$	$1+\frac{2}{1.7}\theta[cos(\frac{i\pi}{m+1})+cos(\frac{j\pi}{n+1})+\sqrt{2}cos(\frac{i\pi}{m+1})cos(\frac{j\pi}{n+1})]$
(2b) $k=2$	$1+\frac{2}{(1.7)^2}\theta[cos(\frac{i\pi}{m+1})+cos(\frac{j\pi}{n+1})+cos(\frac{i\pi}{m+1})cos(\frac{j\pi}{n+1})]$
(3) $k=1$	no closed form
(3) $k=2$	no closed form
(4a) $k=1$	$1+\frac{2}{exp(1.7)}\theta[cos(\frac{i\pi}{m+1})+cos(\frac{j\pi}{n+1})]$
(4a) $k=2$	$1+\frac{2}{exp((1.7)^2)}\theta[cos(\frac{i\pi}{m+1})+cos(\frac{j\pi}{n+1})]$
(4b) $k=1$	$1+\frac{2}{exp(1.7)}\theta[cos(\frac{i\pi}{m+1})+cos(\frac{j\pi}{n+1})+\sqrt{2}cos(\frac{i\pi}{m+1})cos(\frac{j\pi}{n+1})]$
(4b) $k=2$	$1+\frac{2}{exp((1.7)^2)}\theta[cos(\frac{i\pi}{m+1})+cos(\frac{j\pi}{n+1})+cos(\frac{i\pi}{m+1})cos(\frac{j\pi}{n+1})]$

Table 4.3: The means and standard deviations of 10,000 estimated parameters from the 10,000 realizations with $\mu = 5$, $\sigma^2 = 4$, $\rho_1 = .4$, $\rho_2 = .2$, $\rho_3 = .05$ for the artificial or real locations points

	$\hat{\mu}$ mean (s.d.) $b(\hat{\mu})$	$\hat{\sigma}^2$ mean (s.d.) $b(\hat{\sigma}^2)$	$\hat{\rho}_1$ mean (s.d.) $b_1(\hat{\rho}_1)$	$\hat{\rho}_2$ mean (s.d.) $b_1(\hat{\rho}_2)$	$\hat{\rho}_3$ mean (s.d.) $b_1(\hat{\rho}_3)$
artificial 5 points	5.0029	2.7017	-0.0626	-0.2611	-0.3700
. .	(1.1363)	(2.0419)	(0.4884)	(0.6159)	(0.4168)
. . .	0.0000	-1.3600	-0.3445	-0.4500	-.4894
real 5 points	4.9997	3.0556	NA[1]	-0.1518	-0.2503
	(0.9716)	(2.1711)	(NA[1])	(0.4820)	(0.8470)
real 24 points	4.9981	3.7838	0.2612	0.1261	-0.0068
	(0.4650)	(1.1599)	(0.5109)	(0.2984)	(0.3878)
real 43 points	5.0006	3.8537	0.3211	0.1452	-0.0021
	(0.3848)	(0.9048)	(0.2414)	(0.1918)	(0.1771)
real 98 points	4.9982	3.9341	0.3632	0.1688	0.0336
	(0.2712)	(0.6232)	(0.1334)	(0.1264)	(0.1089)
	0.0000	-0.0173	-0.0169	-0.0177	-0.0217
real 190 points	4.9994	3.9629	0.3753	0.1793	0.0392
	(0.1963)	(0.4595)	(0.0837)	(0.0917)	(0.0738)
real 319 points	4.9980	3.9743	0.3808	0.1845	0.0436
	(0.1549)	(0.3597)	(0.0629)	(0.0722)	(0.0567)
	0.0000	-0.0239	-0.0048	-0.0057	-0.0065

$b(.)$ denotes the true bias of estimators and $b_1(.)$ denotes the bias from Taylor approximation through first order.

[1] "NA" indicates the estimate $\hat{\rho}_1$ is not available because there is no first-order horizontal neighbor among the first five sample plot points in the Siuslaw Forest (see the points in the west and south coner of Figure 3.5).

Table 4.4: The means and standard deviations of 10,000 estimated parameters from the 10,000 realizations with $\mu = 5$, $\sigma^2 = 4$, $\rho_1 = .3$, $\rho_2 = .15$, $\rho_3 = .02$, $\rho_4 = .01$ for the artificial points

	$\hat{\mu}$ mean (s.d.) $\mu + b(\hat{\mu})$	$\hat{\sigma}^2$ mean (s.d.) $\sigma^2 + b(\hat{\sigma}^2)$	$\hat{\rho}_1$ mean (s.d.) $\rho_1 + b_1(\hat{\rho}_1)$	$\hat{\rho}_2$ mean (s.d.) $\rho_2 + b_1(\hat{\rho}_2)$	$\hat{\rho}_3$ mean (s.d.) $\rho_3 + b_1(\hat{\rho}_3)$	$\hat{\rho}_4$ mean (s.d.) $\rho_4 + b_1(\hat{\rho}_4)$
3x2	5.0162	2.9779	-0.0401	-0.1661	-0.3096	-0.3236
	(1.0140)	(1.9605)	(0.4206)	(0.5380)	(0.5913)	(0.5957)
	5.0000	-1.0467	-0.2718	-0.3012	-0.3721	-0.3734
10x10	5.0046	3.9268	0.2801	0.1309	-0.0021	-0.0101
	(0.2692)	(0.6203)	(0.0958)	(0.1085)	(0.1196)	(0.1174)
	5.0000	3.9257	0.2862	0.1336	0.0006	-0.0097
			$b(\hat{\gamma}_1)$ $b_s(\hat{\gamma}_1)$	$b(\hat{\gamma}_2)$ $b_s(\hat{\gamma}_2)$	$b(\hat{\gamma}_3)$ $b_s(\hat{\gamma}_3)$	$b(\hat{\gamma}_4)$ $b_s(\hat{\gamma}_4)$
3x2			-1.1167	-1.0466	-1.1200	-1.1133
			-1.1395	-1.0898	-1.1287	-1.1060
10x10			-0.0807	-0.0773	-0.0814	-0.0741
			-0.0767	-0.0756	-0.0779	-0.0779

$b(.)$ denotes the true bias of estimators, $b_1(.)$ denotes the bias from Taylor approximation through first order, and $b_s(.)$ denotes the bias from the means of 10,000 estimated parameters.

Table 4.5: The means and standard deviations of 6,000 estimated parameters from the 6,000 realizations with $\mu = 5$, $\sigma^2 = 4$, $\rho_1 = .3$, $\rho_2 = .15$, $\rho_3 = .02$, $\rho_4 = .01$ for the artificial points

	$\hat{\mu}$ mean (s.d.)	$\hat{\sigma}^2$ mean (s.d.)	$\hat{\rho}_1$ mean (s.d.)	$\hat{\rho}_2$ mean (s.d.)	$\hat{\rho}_3$ mean (s.d.)	$\hat{\rho}_4$ mean (s.d.)
3x2						
MOME	5.0060	2.9436	-0.0405	-0.1460	-0.3218	-0.3319
	(1.0126)	(1.9417)	(0.4188)	(0.5401)	(0.5945)	(0.5960)
PMLE	5.0023	2.4343	-0.1118	-0.1144	-0.2062	-0.2119
	(1.0781)	(1.9048)	(0.3335)	(0.4510)	(0.3391)	(0.3391)
10x10						
MOME	5.0021	3.9405	0.2801	0.1298	-0.0015	-0.0117
	(0.2689)	(0.6067)	(0.0958)	(0.1085)	(0.1168)	(0.1176)
PMLE	5.0016	3.8328	0.2545	0.1170	-0.0046	-0.0137
	(0.2682)	(0.5583)	(0.0682)	(0.0936)	(0.0912)	(0.0910)

Table 4.6: Moment estimates of μ, σ^2, ρ_1, ρ_2, ρ_3, and ρ_4 for the M data under the correlation schemes (1a)-(2e)

	The number of plots with the isolated plot (N=313)					
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$
(1a)	62.75	4536.76	0.1534	0.1534		
(1b)	62.75	4536.76	0.1363	0.1711		
(2a)	62.75	4536.76	0.1016	0.1016	0.1016	0.1016
(2b)	62.75	4536.76	0.1534	0.1534	0.0455	0.0455
(2c)	62.75	4536.76	0.1363	0.1711	0.0455	0.0455
(2d)	62.75	4536.76	0.1534	0.1534	0.1180	-0.0263
(2e)	62.75	4536.76	0.1363	0.1711	0.1180	-0.0263
	The number of plots without the isolated plot (N=312)					
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$
(1a)	62.89	4544.73	0.1532	0.1532		
(1b)	62.89	4544.73	0.1361	0.1709		
(2a)	62.89	4544.73	0.1014	0.1014	0.1014	0.1014
(2b)	62.89	4544.73	0.1532	0.1532	0.0455	0.0455
(2c)	62.89	4544.73	0.1361	0.1709	0.0455	0.0455
(2d)	62.89	4544.73	0.1532	0.1532	0.1180	-0.0262
(2e)	62.89	4544.73	0.1361	0.1709	0.1180	-0.0262

The correlation matrix schemes are

$(1a)$ $\rho_2 = \rho_1$, $\rho_3 = \rho_4 = 0$, $(1b)$ $\rho_1 \neq \rho_2$, $\rho_3 = \rho_4 = 0$,

$(2a)$ $\rho_i = \rho_1$, $i = 2, 3, 4$, $(2b)$ $\rho_2 = \rho_1$, $\rho_3 = \rho_4$,

$(2c)$ $\rho_1 \neq \rho_2$, $\rho_4 = \rho_3$, $(2d)$ $\rho_2 = \rho_1$, $\rho_3 \neq \rho_4$,

$(2e)$ $\rho_i, i = 1, \ldots, 4$.

Table 4.7: Moment estimates of μ, σ^2, ρ_1, ρ_2, ρ_3, and ρ_4 for the M_s data under the correlation schemes (1a)-(2e)

	The number of plots with the isolated plot (N=313)					
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$
(1a)	12.54	1633.88	0.0341	0.0341		
(1b)	12.54	1633.88	0.0393	0.0287		
(2a)	12.54	1633.88	0.0525	0.0525	0.0525	0.0525
(2b)	12.54	1633.88	0.0341	0.0341	0.0727	0.0727
(2c)	12.54	1633.88	0.0393	0.0287	0.0727	0.0727
(2d)	12.54	1633.88	0.0341	0.0341	0.0230	0.1220
(2e)	12.54	1633.88	0.0393	0.0287	0.0230	0.1220
	The number of plots without the isolated plot (N=312)					
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$
(1a)	12.58	1638.62	0.0340	0.0340		
(1b)	12.58	1638.62	0.0391	0.0287		
(2a)	12.58	1638.62	0.0525	0.0525	0.0525	0.0525
(2b)	12.58	1638.62	0.0340	0.0340	0.0725	0.0725
(2c)	12.58	1638.62	0.0391	0.0287	0.0725	0.0725
(2d)	12.58	1638.62	0.0340	0.0340	0.0229	0.1220
(2e)	12.58	1638.62	0.0391	0.0287	0.0229	0.1220

The correlation matrix schemes are

$(1a)$ $\rho_2 = \rho_1$, $\rho_3 = \rho_4 = 0$, $(1b)$ $\rho_1 \neq \rho_2$, $\rho_3 = \rho_4 = 0$,

$(2a)$ $\rho_i = \rho_1$, $i = 2, 3, 4$, $(2b)$ $\rho_2 = \rho_1$, $\rho_3 = \rho_4$,

$(2c)$ $\rho_1 \neq \rho_2$, $\rho_4 = \rho_3$, $(2d)$ $\rho_2 = \rho_1$, $\rho_3 \neq \rho_4$,

$(2e)$ $\rho_i, i = 1, \ldots, 4$.

Table 4.8: Moment estimates of μ, σ^2, ρ_1, ρ_2, ρ_3, and ρ_4 for the Lt data under the correlation schemes (1a)-(2e)

	The number of plots with the isolated plot (N=313)					
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$
(1a)	367.64	72091.52	0.1067	0.1067		
(1b)	367.64	72091.52	0.0947	0.1190		
(2a)	367.64	72091.52	0.0760	0.0760	0.0760	0.0760
(2b)	367.64	72091.52	0.1067	0.1067	0.0429	0.0429
(2c)	367.64	72091.52	0.0947	0.1190	0.0429	0.0429
(2d)	367.64	72091.52	0.1067	0.1067	0.0727	0.0133
(2e)	367.64	72091.52	0.0947	0.1190	0.0727	0.0133
	The number of plots without the isolated plot (N=312)					
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$
(1a)	367.36	72298.19	0.1064	0.1064		
(1b)	367.36	72298.19	0.0945	0.1187		
(2a)	367.36	72298.19	0.0758	0.0758	0.0758	0.0758
(2b)	367.36	72298.19	0.1064	0.1064	0.0427	0.0427
(2c)	367.36	72298.19	0.0945	0.1187	0.0427	0.0427
(2d)	367.36	72298.19	0.1064	0.1064	0.0725	0.0132
(2e)	367.36	72298.19	0.0945	0.1187	0.0725	0.0132

The correlation matrix schemes are

(1a) $\rho_2 = \rho_1$, $\rho_3 = \rho_4 = 0$, (1b) $\rho_1 \neq \rho_2$, $\rho_3 = \rho_4 = 0$,
(2a) $\rho_i = \rho_1$, $i = 2, 3, 4$, (2b) $\rho_2 = \rho_1$, $\rho_3 = \rho_4$,
(2c) $\rho_1 \neq \rho_2$, $\rho_4 = \rho_3$, (2d) $\rho_2 = \rho_1$, $\rho_3 \neq \rho_4$,
(2e) $\rho_i, i = 1, \ldots, 4$.

Table 4.9: Moment estimates of μ, σ^2, ρ_1, ρ_2, ρ_3, and ρ_4 for the Lt_s data under the correlation schemes (1a)-(2e)

	The number of plots with the isolated plot (N=313)					
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$
(1a)	200.84	128986.53	0.2654	0.2654		
(1b)	200.84	128986.53	0.2300	0.3022		
(2a)	200.84	128986.53	0.2389	0.2389	0.2389	0.2389
(2b)	200.84	128986.53	0.2654	0.2654	0.2101	0.2101
(2c)	200.84	128986.53	0.2300	0.3022	0.2101	0.2101
(2d)	200.84	128986.53	0.2654	0.2654	0.3957	0.0265
(2e)	200.84	128986.53	0.2300	0.3022	0.3957	0.0265
	The number of plots without the isolated plot (N=312)					
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$
(1a)	198.81	128113.63	0.2671	0.2671		
(1b)	198.81	128113.63	0.2315	0.3040		
(2a)	198.81	128113.63	0.2402	0.2402	0.2402	0.2402
(2b)	198.81	128113.63	0.2671	0.2671	0.2101	0.2101
(2c)	198.81	128113.63	0.2315	0.3040	0.2101	0.2101
(2d)	198.81	128113.63	0.2671	0.2671	0.3982	0.0261
(2e)	198.81	128113.63	0.2315	0.3040	0.3982	0.0261

The correlation matrix schemes are

(1a) $\rho_2 = \rho_1$, $\rho_3 = \rho_4 = 0$, (1b) $\rho_1 \neq \rho_2$, $\rho_3 = \rho_4 = 0$,

(2a) $\rho_i = \rho_1$, $i = 2, 3, 4$, (2b) $\rho_2 = \rho_1$, $\rho_3 = \rho_4$,

(2c) $\rho_1 \neq \rho_2$, $\rho_4 = \rho_3$, (2d) $\rho_2 = \rho_1$, $\rho_3 \neq \rho_4$,

(2e) $\rho_i, i = 1, \ldots, 4$.

Table 4.10: Moment estimates of μ, σ^2, ρ_1, ρ_2, ρ_3, and ρ_4 for the Tba data under the correlation schemes (1a)-(2e)

	The number of plots with the isolated plot (N=313)					
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$
(1a)	43.50	590.40	0.0083	0.0083		
(1b)	43.50	590.40	-0.0147	0.0321		
(2a)	43.50	590.40	0.0758	0.0758	0.0758	0.0758
(2b)	43.50	590.40	0.0083	0.0083	0.1489	0.1489
(2c)	43.50	590.40	-0.0147	0.0322	0.1489	0.1489
(2d)	43.50	590.40	0.0083	0.0083	0.1808	0.1173
(2e)	43.50	590.40	-0.0147	0.0322	0.1808	0.1173
	The number of plots without the isolated plot (N=312)					
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$
(1a)	43.47	592.06	0.0083	0.0083		
(1b)	43.47	592.06	-0.0146	0.0320		
(2a)	43.47	592.06	0.0756	0.0756	0.0756	0.0756
(2b)	43.47	592.06	0.0083	0.0083	0.1484	0.1484
(2c)	43.47	592.06	-0.0146	0.0320	0.1484	0.1484
(2d)	43.47	592.06	0.0083	0.0083	0.1802	0.1170
(2e)	43.47	592.06	-0.0146	0.0320	0.1802	0.1170

The correlation matrix schemes are

$(1a)$ $\rho_2 = \rho_1$, $\rho_3 = \rho_4 = 0$, $(1b)$ $\rho_1 \neq \rho_2$, $\rho_3 = \rho_4 = 0$,

$(2a)$ $\rho_i = \rho_1$, $i = 2, 3, 4$, $(2b)$ $\rho_2 = \rho_1$, $\rho_3 = \rho_4$,

$(2c)$ $\rho_1 \neq \rho_2$, $\rho_4 = \rho_3$, $(2d)$ $\rho_2 = \rho_1$, $\rho_3 \neq \rho_4$,

$(2e)$ $\rho_i, i = 1, \ldots, 4$.

Table 4.11: Moment estimates of μ, σ^2, ρ_1, ρ_2, ρ_3, and ρ_4 for the Tba_s data under the correlation schemes (1a)-(2e)

	The number of plots with the isolated plot (N=313)					
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$
(1a)	0.23	0.27	0.2701	0.2701		
(1b)	0.23	0.27	0.2243	0.3175		
(2a)	0.23	0.27	0.2226	0.2226	0.2226	0.2226
(2b)	0.23	0.27	0.2701	0.2701	0.1713	0.1713
(2c)	0.23	0.27	0.2243	0.3175	0.1713	0.1713
(2d)	0.23	0.27	0.2701	0.2701	0.3537	-0.0092
(2e)	0.23	0.27	0.2243	0.3175	0.3537	-0.0092
	The number of plots without the isolated plot (N=312)					
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$
(1a)	0.23	0.26	0.2726	0.2726		
(1b)	0.23	0.26	0.2264	0.3203		
(2a)	0.23	0.26	0.2226	0.2226	0.2226	0.2226
(2b)	0.23	0.26	0.2726	0.2726	0.1725	0.1725
(2c)	0.23	0.26	0.2264	0.3203	0.1725	0.1725
(2d)	0.23	0.26	0.2726	0.2726	0.3568	-0.0099
(2e)	0.23	0.26	0.2264	0.3203	0.3568	-0.0099

The correlation matrix schemes are

$(1a)$ $\rho_2 = \rho_1$, $\rho_3 = \rho_4 = 0$, $(1b)$ $\rho_1 \neq \rho_2$, $\rho_3 = \rho_4 = 0$,

$(2a)$ $\rho_i = \rho_1$, $i = 2, 3, 4$, $(2b)$ $\rho_2 = \rho_1$, $\rho_3 = \rho_4$,

$(2c)$ $\rho_1 \neq \rho_2$, $\rho_4 = \rho_3$, $(2d)$ $\rho_2 = \rho_1$, $\rho_3 \neq \rho_4$,

$(2e)$ $\rho_i, i = 1, \ldots, 4$.

Table 4.12: Maximum Pseudo-likelihood estimates (PMLE) of μ, σ^2, ρ_1, ρ_2, ρ_3, and ρ_4 for the M data under the correlation schemes (1a)-(2e)

	The number of plots with the isolated plot (N=313)									
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$	p-value	AIC	AICC	BIC
(1a)	63.0	4533.55	0.1600	0.1600			0.001	3516.08	3516.12	3523.57
(1b)	63.0	4533.79	0.1620	0.1582			0.003	3518.08	3518.16	3529.32
(2a)	63.4	4545.52	0.1250	0.1250	0.1250	0.1250	0.001	3517.11	3517.15	3524.60
(2b)	63.4	4613.25	0.2122	0.2122	0.0946	0.0946	0.001	3515.08	3515.16	3526.32
(2c)	63.4	4617.91	0.2295	0.1968	0.0957	0.0957	0.002	3516.94	3517.07	3531.92
(2d)	63.9	4609.32	0.2000	0.2000	0.2136	0.0072	0.000	3512.94	3513.07	3527.92
(2e)	63.9	4615.17	0.2074	0.1982	0.2133	0.0098	0.001	3512.92	3513.05	3527.91
	The number of plots without the isolated plot (N=312)									
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$	p-value	AIC	AICC	BIC
(1a)	62.2	4541.85	0.1602	0.1602			0.001	3505.37	3505.37	3512.85
(1b)	62.2	4542.11	0.1623	0.1583			0.003	3507.36	3507.44	3518.59
(2a)	63.6	4553.32	0.1250	0.1250	0.1250	0.1250	0.001	3506.38	3506.42	3513.87
(2b)	63.5	4631.84	0.2115	0.2115	0.0917	0.0917	0.001	3494.01	3494.09	3505.23
(2c)	63.7	4627.43	0.2301	0.1972	0.0961	0.0961	0.002	3506.20	3506.33	3521.18
(2d)	63.2	4617.85	0.2000	0.2000	0.2142	0.0071	0.000	3502.19	3502.33	3517.17
(2e)	64.2	4624.66	0.2078	0.1986	0.2139	0.0100	0.001	3502.18	3502.31	3517.15

The correlation matrix schemes are

(1a) $\rho_2 = \rho_1$, $\rho_3 = \rho_4 = 0$, (1b) $\rho_1 \neq \rho_2$, $\rho_3 = \rho_4 = 0$,

(2a) $\rho_i = \rho_1$, $i = 2, 3, 4$, (2b) $\rho_2 = \rho_1$, $\rho_3 = \rho_4$,

(2c) $\rho_1 \neq \rho_2$, $\rho_4 = \rho_3$, (2d) $\rho_2 = \rho_1$, $\rho_3 \neq \rho_4$,

(2e) $\rho_i, i = 1, \ldots, 4$.

Table 4.13: Maximum Pseudo-likelihood estimates (PMLE) of μ, σ^2, ρ_1, ρ_2, ρ_3, and ρ_4 for the M_s data under the correlation schemes (1a)-(2e)

	The number of plots with the isolated plot (N=313)									
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$	p-value	AIC	AICC	BIC
(1a)	12.6	1633.66	0.0301	0.0301			0.52	3207.64	3207.68	3215.13
(1b)	12.5	1633.77	0.0402	0.0202			0.80	3209.61	3209.69	3220.85
(2a)	12.6	1632.08	0.0405	0.0405	0.0405	0.0405	0.21	3206.48	3206.52	3213.97
(2b)	12.5	1631.71	0.0186	0.0186	0.0583	0.0583	0.40	3208.22	3208.30	3219.46
(2c)	12.5	1632.37	0.0378	0.0022	0.0600	0.0600	0.58	3210.09	3210.22	3225.07
(2d)	12.6	1645.61	-0.0283	-0.0283	-0.0063	0.1690	0.30	3208.43	3208.55	3223.41
(2e)	12.5	1647.75	0.0031	-0.0574	-0.0122	0.1743	0.39	3207.95	3208.08	3222.94
	The number of plots without the isolated plot (N=312)									
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$	p-value	AIC	AICC	BIC
(1a)	12.6	1638.40	0.0302	0.0302			0.52	3198.31	3198.34	3205.79
(1b)	12.6	1638.51	0.0404	0.0202			0.80	3200.28	3200.36	3211.51
(2a)	12.6	1636.83	0.0406	0.0406	0.0406	0.0406	0.21	3197.15	3197.19	3204.63
(2b)	12.7	1641.18	0.0183	0.0183	0.0602	0.0602	0.39	3189.48	3189.56	3200.70
(2c)	12.6	1637.14	0.0379	0.0022	0.0602	0.0602	0.58	3200.76	3200.89	3215.73
(2d)	12.7	1650.80	-0.0288	-0.0288	-0.0065	0.1700	0.30	3199.08	3199.21	3214.05
(2e)	12.6	1652.91	0.0027	-0.0578	-0.0125	0.1752	0.39	3198.61	3198.74	3213.58

The correlation matrix schemes are

$(1a)$ $\rho_2 = \rho_1$, $\rho_3 = \rho_4 = 0$, $(1b)$ $\rho_1 \neq \rho_2$, $\rho_3 = \rho_4 = 0$,

$(2a)$ $\rho_i = \rho_1$, $i = 2, 3, 4$, $(2b)$ $\rho_2 = \rho_1$, $\rho_3 = \rho_4$,

$(2c)$ $\rho_1 \neq \rho_2$, $\rho_4 = \rho_3$, $(2d)$ $\rho_2 = \rho_1$, $\rho_3 \neq \rho_4$,

$(2e)$ $\rho_i, i = 1, \ldots, 4$.

Table 4.14: Maximum Pseudo-likelihood estimates (PMLE) of μ, σ^2, ρ_1, ρ_2, ρ_3, and ρ_4 for the Lt data under the correlation schemes (1a)-(2e)

	The number of plots with the isolated plot (N=313)									
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$	p-value	AIC	AICC	BIC
(1a)	366.7	71900.7	0.0815	0.0815			0.06	4389.8	4389.8	4397.3
(1b)	366.7	71895.3	0.0769	0.0855			0.17	4391.8	4391.9	4403.0
(2a)	367.9	71924.5	0.0591	0.0591	0.0591	0.0591	0.07	4390.1	4390.1	4397.6
(2b)	367.3	71915.3	0.0868	0.0868	0.0322	0.0322	0.13	4391.3	4391.4	4402.6
(2c)	367.3	71907.9	0.0807	0.0921	0.0326	0.0326	0.26	4393.3	4393.5	4408.3
(2d)	367.2	71882.7	0.0847	0.0847	0.0551	0.0038	0.24	4393.1	4393.3	4408.1
(2e)	367.1	71877.5	0.0739	0.0936	0.0581	0.0017	0.37	4393.1	4393.2	4408.1
	The number of plots without the isolated plot (N=312)									
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$	p-value	AIC	AICC	BIC
(1a)	366.4	72109.5	0.0817	0.0817			0.06	4376.7	4376.7	4384.1
(1b)	366.4	72103.9	0.0770	0.0857			0.17	4378.7	4378.7	4389.9
(2a)	367.6	72133.5	0.0593	0.0593	0.0593	0.0593	0.07	4377.0	4377.0	4384.4
(2b)	367.2	72345.7	0.0872	0.0872	0.0321	0.0321	0.13	4365.1	4365.2	4376.4
(2c)	367.0	72117.5	0.0809	0.0923	0.0326	0.0326	0.26	4380.2	4380.3	4395.2
(2d)	366.8	72092.2	0.0849	0.0849	0.0552	0.0038	0.24	4380.0	4380.1	4395.0
(2e)	366.8	72087.0	0.0741	0.0938	0.0582	0.0016	0.37	4380.0	4380.1	4395.0

The correlation matrix schemes are

$(1a)$ $\rho_2 = \rho_1$, $\rho_3 = \rho_4 = 0$, $(1b)$ $\rho_1 \neq \rho_2$, $\rho_3 = \rho_4 = 0$,

$(2a)$ $\rho_i = \rho_1$, $i = 2, 3, 4$, $(2b)$ $\rho_2 = \rho_1$, $\rho_3 = \rho_4$,

$(2c)$ $\rho_1 \neq \rho_2$, $\rho_4 = \rho_3$, $(2d)$ $\rho_2 = \rho_1$, $\rho_3 \neq \rho_4$,

$(2e)$ $\rho_i, i = 1, \ldots, 4$.

Table 4.15: Maximum Pseudo-likelihood estimates (PMLE) of μ, σ^2, ρ_1, ρ_2, ρ_3, and ρ_4 for the Lt_s data under the correlation schemes (1a)-(2e)

	The number of plots with the isolated plot (N=313)									
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$	p-value	AIC	AICC	BIC
(1a)	200.4	127534.31	0.2029	0.2029			0.00	4552.3	4552.3	4559.8
(1b)	201.3	127942.09	0.1287	0.2768			0.00	4553.0	4553.0	4564.2
(2a)	204.8	121901.22	0.1250	0.1250	0.1250	0.1250	0.00	4546.6	4546.6	4554.1
(2b)	206.0	128222.78	0.2458	0.2458	0.1766	0.1766	0.00	4542.8	4542.9	4554.1
(2c)	206.1	128241.87	0.2241	0.2775	0.1701	0.1701	0.00	4544.2	4544.3	4559.2
(2d)	204.1	123670.96	0.2000	0.2000	0.2574	0.0289	0.00	4538.6	4538.7	4553.6
(2e)	204.6	125009.59	0.2012	0.2555	0.2466	0.0434	0.00	4537.9	4538.0	4552.8
	The number of plots without the isolated plot (N=312)									
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$	p-value	AIC	AICC	BIC
(1a)	197.4	126584.84	0.2022	0.2022			0.00	4535.6	4535.6	4543.0
(1b)	198.3	126975.98	0.1291	0.2751			0.00	4536.2	4536.3	4547.4
(2a)	201.6	121019.17	0.1250	0.1250	0.1250	0.1250	0.00	4529.8	4529.8	4537.3
(2a)	203.1	125952.06	0.1958	0.1958	0.1958	0.1958	0.00	4525.2	4525.2	4532.6
(2b)	201.9	127629.79	0.2448	0.2448	0.1753	0.1753	0.00	4512.6	4512.7	4523.8
(2c)	202.2	127190.08	0.2223	0.2757	0.1691	0.1691	0.00	4527.5	4527.6	4542.5
(2d)	200.5	122737.68	0.2000	0.2000	0.2561	0.0281	0.00	4521.8	4521.9	4536.8
(2e)	200.8	123973.03	0.1996	0.2536	0.2458	0.0418	0.00	4521.1	4521.2	4536.1

The correlation matrix schemes are

$(1a)$ $\rho_2 = \rho_1$, $\rho_3 = \rho_4 = 0$, $(1b)$ $\rho_1 \neq \rho_2$, $\rho_3 = \rho_4 = 0$,

$(2a)$ $\rho_i = \rho_1$, $i = 2, 3, 4$, $(2b)$ $\rho_2 = \rho_1$, $\rho_3 = \rho_4$,

$(2c)$ $\rho_1 \neq \rho_2$, $\rho_4 = \rho_3$, $(2d)$ $\rho_2 = \rho_1$, $\rho_3 \neq \rho_4$,

$(2e)$ $\rho_i, i = 1, \ldots, 4$.

Table 4.16: Maximum Pseudo-likelihood estimates (PMLE) of μ, σ^2, ρ_1, ρ_2, ρ_3, and ρ_4 for the Tba data under the correlation schemes (1a)-(2e)

	The number of plots with the isolated plot (N=313)									
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$	p-value	AIC	AICC	BIC
(1a)	43.50	590.39	0.0060	0.0060			0.89	2889.42	2889.46	2896.92
(1b)	43.56	590.75	-0.0312	0.0414			0.84	2891.08	2891.16	2902.32
(2a)	43.56	590.64	0.0790	0.0790	0.0790	0.0790	0.04	2885.04	2885.08	2892.53
(2b)	43.55	588.14	0.0021	0.0021	0.1363	0.1363	0.01	2882.82	2882.90	2894.06
(2c)	43.63	588.41	-0.0302	0.0308	0.1386	0.1386	0.02	2884.02	2884.15	2899.01
(2d)	43.55	589.61	0.0052	0.0052	0.2018	0.0799	0.02	2883.58	2883.71	2898.57
(2e)	43.64	590.41	-0.0292	0.0366	0.2105	0.0749	0.03	2882.60	2882.73	2897.58
	The number of plots without the isolated plot (N=312)									
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$	p-value	AIC	AICC	BIC
(1a)	43.47	592.05	0.0060	0.0060			0.89	2889.42	2889.46	2896.92
(1b)	43.53	592.42	-0.0312	0.0415			0.84	2891.08	2891.16	2902.32
(2a)	43.52	592.34	0.0793	0.0793	0.0793	0.0793	0.04	2885.04	2885.08	2892.53
(2b)	43.41	587.94	0.0021	0.0027	0.1332	0.1332	0.02	2882.82	2882.90	2894.06
(2c)	43.59	590.12	-0.0302	0.0307	0.1389	0.1389	0.02	2884.02	2884.15	2899.01
(2d)	43.51	591.35	0.0052	0.0052	0.2023	0.0799	0.02	2883.58	2883.71	2898.57
(2e)	43.60	592.17	-0.0292	0.0366	0.2110	0.0749	0.03	2882.60	2882.73	2897.58

The correlation matrix schemes are

$(1a)$ $\rho_2 = \rho_1$, $\rho_3 = \rho_4 = 0$, $(1b)$ $\rho_1 \neq \rho_2$, $\rho_3 = \rho_4 = 0$,

$(2a)$ $\rho_i = \rho_1$, $i = 2, 3, 4$, $(2b)$ $\rho_2 = \rho_1$, $\rho_3 = \rho_4$,

$(2c)$ $\rho_1 \neq \rho_2$, $\rho_4 = \rho_3$, $(2d)$ $\rho_2 = \rho_1$, $\rho_3 \neq \rho_4$,

$(2e)$ $\rho_i, i = 1, \ldots, 4$.

Table 4.17: Maximum Pseudo-likelihood estimates (PMLE) of μ, σ^2, ρ_1, ρ_2, ρ_3, and ρ_4 for the Tba_s data under the correlation schemes (1a)-(2e)

	The number of plots with the isolated plot (N=313)									
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$	p-value	AIC	AICC	BIC
(1a)	0.23	0.26	0.1764	0.1764			0.00	458.02	458.06	465.51
(1b)	0.23	0.26	0.1225	0.2269			0.00	459.33	459.41	470.57
(2a)	0.24	0.26	0.1250	0.1250	0.1250	0.1250	0.00	454.08	454.12	461.57
(2b)	0.24	0.26	0.2345	0.2345	0.1419	0.1419	0.00	451.49	451.57	462.73
(2c)	0.24	0.26	0.2195	0.2472	0.1390	0.1390	0.00	453.33	453.46	468.31
(2d)	0.23	0.26	0.2000	0.2000	0.2360	-0.0101	0.00	446.07	446.20	461.06
(2e)	0.23	0.26	0.1927	0.2184	0.2326	-0.0045	0.00	445.94	446.06	460.92
	The number of plots without the isolated plot (N=312)									
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$	p-value	AIC	AICC	BIC
(1a)	0.23	0.26	0.1757	0.1757			0.00	453.55	453.59	461.04
(1b)	0.23	0.26	0.1228	0.2254			0.00	454.87	454.95	466.10
(2a)	0.23	0.25	0.1250	0.1250	0.1250	0.1250	0.00	449.59	449.63	457.08
(2b)	0.23	0.26	0.2331	0.2331	0.1406	0.1406	0.00	446.63	446.70	457.84
(2c)	0.23	0.26	0.2173	0.2453	0.1375	0.1375	0.00	448.94	449.07	463.91
(2d)	0.23	0.26	0.2000	0.2000	0.2347	-0.0103	0.00	441.58	441.71	456.55
(2e)	0.23	0.26	0.1907	0.2167	0.2314	-0.0061	0.00	441.45	441.58	456.42

The correlation matrix schemes are

$(1a)$ $\rho_2 = \rho_1$, $\rho_3 = \rho_4 = 0$, $(1b)$ $\rho_1 \neq \rho_2$, $\rho_3 = \rho_4 = 0$,

$(2a)$ $\rho_i = \rho_1$, $i = 2, 3, 4$, $(2b)$ $\rho_2 = \rho_1$, $\rho_3 = \rho_4$,

$(2c)$ $\rho_1 \neq \rho_2$, $\rho_4 = \rho_3$, $(2d)$ $\rho_2 = \rho_1$, $\rho_3 \neq \rho_4$,

$(2e)$ $\rho_i, i = 1, \ldots, 4$.

Table 4.18: Maximum Pseudo-likelihood estimates (PMLE) of μ, σ^2 and θ under alternative correlation schemes (1)-(4)

	The number of plots without the isolated plot (N=312)							
	M							
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\theta}$	$max(\hat{\rho}_{ij})$	p-value	AIC	AICC	BIC
(1) $k=1$	66.77	5035.21	0.3031	0.3031	0.0002	3503.43	3503.47	3510.91
(1) $k=2$	64.84	4691.00	0.2405	0.2405	0.0001	3501.95	3501.99	3509.44
(2a) $k=1$	63.19	4541.85	0.1602	0.1602	0.0006	3505.37	3505.40	3512.85
(2a) $k=2$	63.19	4541.85	0.1602	0.1602	0.0006	3505.37	3505.40	3512.85
(2b) $k=1$	63.73	4615.51	0.1846	0.1846	0.0003	3503.76	3503.80	3511.24
(2b) $k=2$	63.69	4629.88	0.2121	0.2121	0.0001	3502.42	3502.46	3509.90
(3) $k=1$	65.07	4779.92	0.7185	0.2643	0.0000	3500.54	3500.58	3508.03
(3) $k=2$	63.59	4607.36	0.5733	0.2109	0.0001	3502.41	3502.44	3509.89
(4a) $k=1$	63.19	4541.85	0.4355	0.1602	0.0006	3505.37	3505.40	3512.85
(4a) $k=2$	63.19	4541.85	0.4355	0.1602	0.0006	3505.37	3505.40	3512.85
(4b) $k=1$	63.73	4622.44	0.5246	0.1930	0.0002	3503.36	3503.40	3510.85
(4b) $k=2$	63.57	4605.78	0.5662	0.2083	0.0001	3502.49	3502.53	3509.98
	M_s							
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\theta}$	$max(\hat{\rho}_{ij})$	p-value	AIC	AICC	BIC
(1) $k=1$	13.52	1634.29	0.0771	0.0771	0.0188	3193.20	3193.24	3200.68
(1) $k=2$	12.91	1633.56	0.0670	0.0670	0.0729	3195.50	3195.54	3202.99
(2a) $k=1$	12.60	1638.40	0.0302	0.0302	0.5203	3198.31	3198.35	3205.79
(2a) $k=2$	12.60	1638.40	0.0302	0.0302	0.5203	3198.31	3198.35	3205.79
(2b) $k=1$	12.61	1637.17	0.0438	0.0438	0.2433	3197.36	3197.40	3204.84
(2b) $k=2$	12.61	1637.49	0.0439	0.0439	0.2859	3197.58	3197.62	3205.07
(3) $k=1$	12.82	1632.79	0.1858	0.0683	0.0501	3194.88	3194.92	3202.37
(3) $k=2$	12.62	1637.48	0.1224	0.0450	0.2920	3197.61	3197.65	3205.10
(4a) $k=1$	12.60	1638.40	0.0821	0.0302	0.5203	3198.31	3198.35	3205.79
(4a) $k=2$	12.60	1638.40	0.0821	0.0302	0.5203	3198.31	3198.35	3205.79
(4b) $k=1$	12.61	1637.24	0.1198	0.0441	0.2510	3197.40	3197.44	3204.89
(4b) $k=2$	12.61	1637.72	0.1150	0.0423	0.3263	3197.76	3197.79	3205.24

Alternative correlation matrix schemes are

(1) $\rho_{ij} = \theta d^{-k}, k = 1, 2$ without neighborhood constraint,

(2a) $\rho_{ij} = \theta d^{-k}, k = 1, 2$ with 1st order neighborhood constraint,

(2b) $\rho_{ij} = \theta d^{-k}, k = 1, 2$ with 2nd order neighborhood constraint,

(3) $\rho_{ij} = \theta exp(-d^k), k = 1, 2$ without neighborhood constraint,

(4a) $\rho_{ij} = \theta exp(-d^k), k = 1, 2$ with 1st neighborhood constraint,

(4b) $\rho_{ij} = \theta exp(-d^k), k = 1, 2$ with 2nd neighborhood constraint,

(Continued)

	The number of plots without the isolated plot (N=312)							
	Lt							
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\theta}$	$max(\hat{\rho}_{ij})$	p-value	AIC	AICC	BIC
(1) $k=1$	368.97	72864.30	0.101	0.1014	0.026	4375.27	4375.30	4382.75
(1) $k=2$	367.99	72149.00	0.097	0.0974	0.023	4375.05	4375.09	4382.53
(2a) $k=1$	366.38	72109.45	0.082	0.0817	0.058	4376.66	4376.70	4384.15
(2a) $k=2$	366.38	72109.45	0.082	0.0817	0.058	4376.66	4376.70	4384.15
(2b) $k=1$	367.38	72133.58	0.074	0.0742	0.053	4376.50	4376.54	4383.99
(2b) $k=2$	367.14	72132.14	0.084	0.0836	0.046	4376.26	4376.30	4383.75
(3) $k=1$	368.25	72163.87	0.258	0.0949	0.018	4374.67	4374.71	4382.15
(3) $k=2$	366.95	72116.56	0.244	0.0897	0.039	4376.00	4376.04	4383.49
(4a) $k=1$	366.38	72109.45	0.222	0.0817	0.058	4376.66	4376.70	4384.15
(4a) $k=2$	366.38	72109.45	0.222	0.0817	0.058	4376.66	4376.70	4384.15
(4b) $k=1$	367.33	72133.93	0.208	0.0765	0.051	4376.44	4376.48	4383.92
(4b) $k=2$	366.94	72124.69	0.237	0.0871	0.045	4376.21	4376.25	4383.70
	Lt_s							
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\theta}$	$max(\hat{\rho}_{ij})$	p-value	AIC	AICC	BIC
(1) $k=1$	215.35	131066.10	0.270	0.2699	0.000	4528.48	4528.51	4535.96
(1) $k=2$	206.02	126737.90	0.256	0.2555	0.000	4527.35	4527.39	4534.84
(2a) $k=1$	197.40	126584.80	0.202	0.2022	0.000	4535.55	4535.59	4543.03
(2a) $k=2$	197.40	126584.80	0.202	0.2022	0.000	4535.55	4535.59	4543.03
(2b) $k=1$	201.99	127221.60	0.246	0.2459	0.000	4524.13	4524.17	4531.62
(2b) $k=2$	200.36	127404.00	0.263	0.2626	0.000	4525.40	4525.44	4532.89
(3) $k=1$	204.26	128015.70	0.705	0.2592	0.000	4527.47	4527.51	4534.95
(3) $k=2$	199.47	126524.60	0.677	0.2491	0.000	4528.21	4528.25	4535.69
(4a) $k=1$	197.40	126584.80	0.550	0.2022	0.000	4535.55	4535.59	4543.03
(4a) $k=2$	197.40	126584.80	0.550	0.2022	0.000	4535.55	4535.59	4543.03
(4b) $k=1$	201.68	127399.40	0.687	0.2526	0.000	4524.21	4524.24	4531.69
(4b) $k=2$	199.31	126842.20	0.689	0.2534	0.000	4527.47	4527.51	4534.96

Alternative correlation matrix schemes are

(1) $\rho_{ij} = \theta d^{-k}, k = 1, 2$ without neighborhood constraint,

(2a) $\rho_{ij} = \theta d^{-k}, k = 1, 2$ with 1st order neighborhood constraint,

(2b) $\rho_{ij} = \theta d^{-k}, k = 1, 2$ with 2nd order neighborhood constraint,

(3) $\rho_{ij} = \theta exp(-d^k), k = 1, 2$ without neighborhood constraint,

(4a) $\rho_{ij} = \theta exp(-d^k), k = 1, 2$ with 1st neighborhood constraint,

(4b) $\rho_{ij} = \theta exp(-d^k), k = 1, 2$ with 2nd neighborhood constraint,

(Continued)

	The number of plots without the isolated plot (N=312)							
	Tba							
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\theta}$	$max(\hat{\rho}_{ij})$	p-value	AIC	AICC	BIC
(1) $k=1$	43.57	595.24	0.0469	0.0469	0.56	2880.77	2880.81	2888.25
(1) $k=2$	43.54	592.19	0.0501	0.0501	0.26	2879.88	2879.92	2887.36
(2a) $k=1$	43.47	592.05	0.0060	0.0060	0.89	2881.08	2881.12	2888.57
(2a) $k=2$	43.47	592.05	0.0060	0.0060	0.89	2881.08	2881.12	2888.57
(2b) $k=1$	43.51	591.76	0.0685	0.0685	0.10	2878.35	2878.39	2885.84
(2b) $k=2$	43.50	591.59	0.0531	0.0531	0.21	2879.51	2879.56	2887.00
(3) $k=1$	43.56	592.77	0.1954	0.0719	0.11	2878.48	2878.52	2885.97
(3) $k=2$	43.49	591.64	0.1150	0.0423	0.31	2880.09	2880.13	2887.58
(4a) $k=1$	43.47	592.05	0.0163	0.0060	0.89	2881.08	2881.12	2888.57
(4a) $k=2$	43.47	592.05	0.0163	0.0060	0.89	2881.08	2881.12	2888.57
(4b) $k=1$	43.51	591.69	0.1780	0.0655	0.12	2878.62	2878.66	2886.11
(4b) $k=2$	43.49	591.67	0.1124	0.0413	0.33	2880.14	2880.18	2887.63
	Tba_s							
	$\hat{\mu}$	$\hat{\sigma}^2$	$\hat{\theta}$	$max(\hat{\rho}_{ij})$	p-value	AIC	AICC	BIC
(1) $k=1$	0.25	0.27	0.2456	0.2456	0.00	448.16	448.20	455.65
(1) $k=2$	0.24	0.26	0.2345	0.2345	0.00	446.69	446.73	454.18
(2a) $k=1$	0.23	0.26	0.1757	0.1757	0.00	453.55	453.59	461.04
(2a) $k=2$	0.23	0.26	0.1757	0.1757	0.00	453.55	453.59	461.04
(2b) $k=1$	0.23	0.26	0.2188	0.2188	0.00	445.38	445.42	452.86
(2b) $k=2$	0.23	0.26	0.2366	0.2366	0.00	445.45	445.48	452.93
(3) $k=1$	0.23	0.26	0.6492	0.2388	0.00	447.53	447.57	455.01
(3) $k=2$	0.23	0.26	0.6136	0.2257	0.00	447.23	447.27	454.72
(4a) $k=1$	0.23	0.26	0.4777	0.1757	0.00	453.55	453.59	461.04
(4a) $k=2$	0.23	0.26	0.4777	0.1757	0.00	453.55	453.59	461.04
(4b) $k=1$	0.23	0.26	0.6136	0.2257	0.00	445.19	445.23	452.68
(4b) $k=2$	0.23	0.26	0.6168	0.2269	0.00	446.86	446.91	454.35

Alternative correlation matrix schemes are

(1) $\rho_{ij} = \theta d^{-k}, k = 1, 2$ without neighborhood constraint,

(2a) $\rho_{ij} = \theta d^{-k}, k = 1, 2$ with 1st order neighborhood constraint,

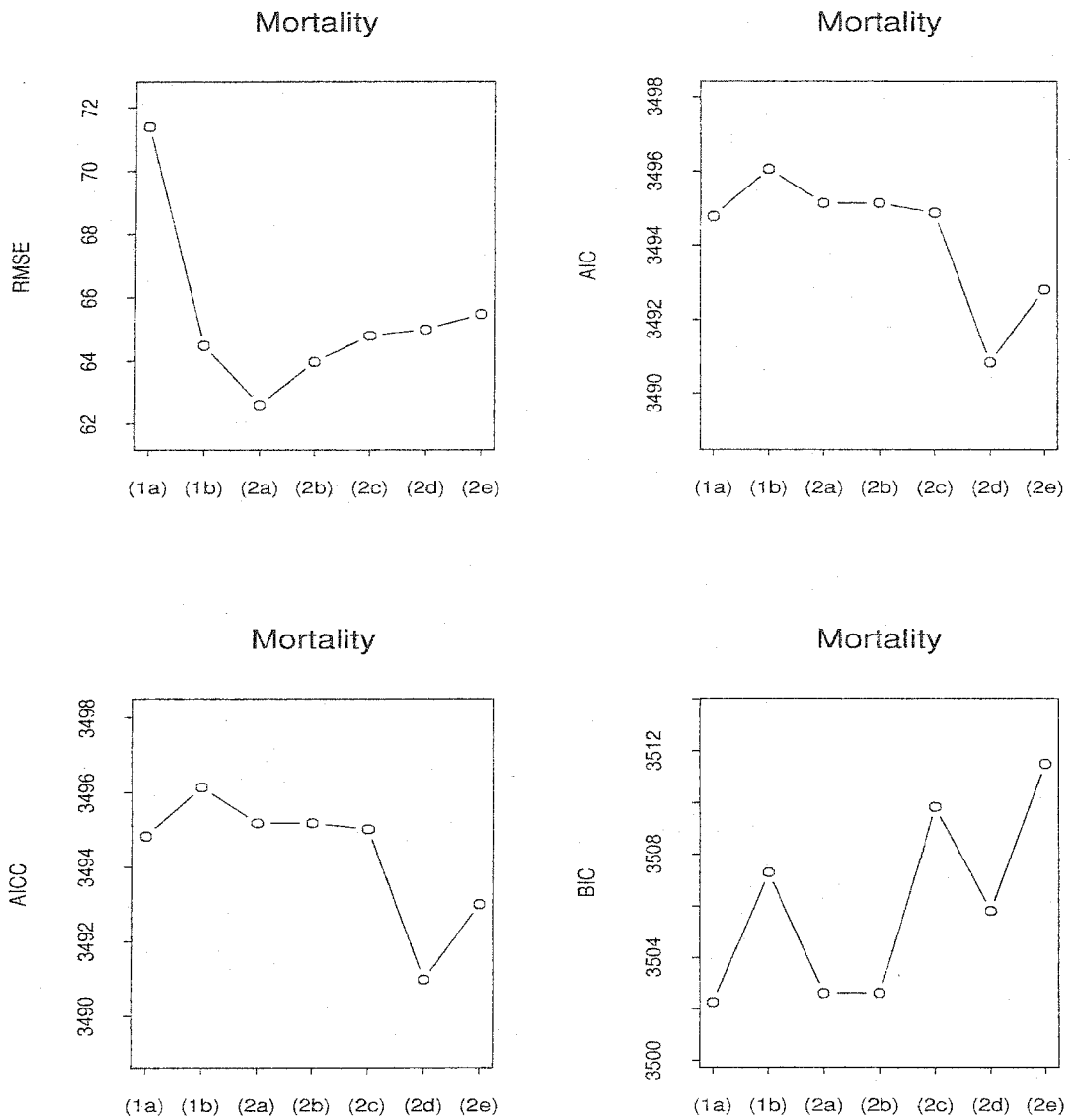
(2b) $\rho_{ij} = \theta d^{-k}, k = 1, 2$ with 2nd order neighborhood constraint,

(3) $\rho_{ij} = \theta exp(-d^k), k = 1, 2$ without neighborhood constraint,

(4a) $\rho_{ij} = \theta exp(-d^k), k = 1, 2$ with 1st neighborhood constraint,

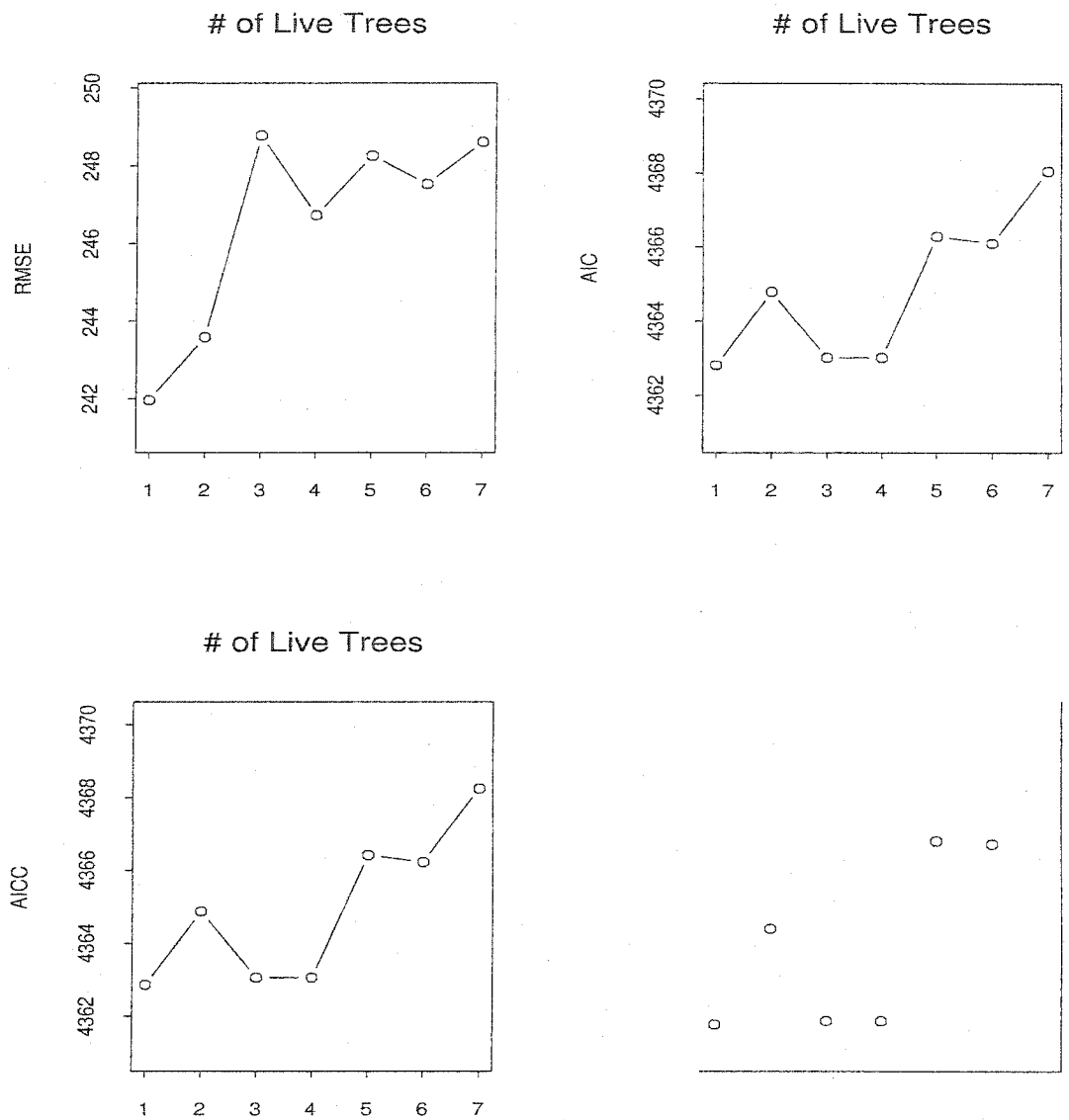
(4b) $\rho_{ij} = \theta exp(-d^k), k = 1, 2$ with 2nd neighborhood constraint.

Figure 4.2: RMSE, AIC, AICC, and BIC for the M data fitted by spatial "Phase 1" models



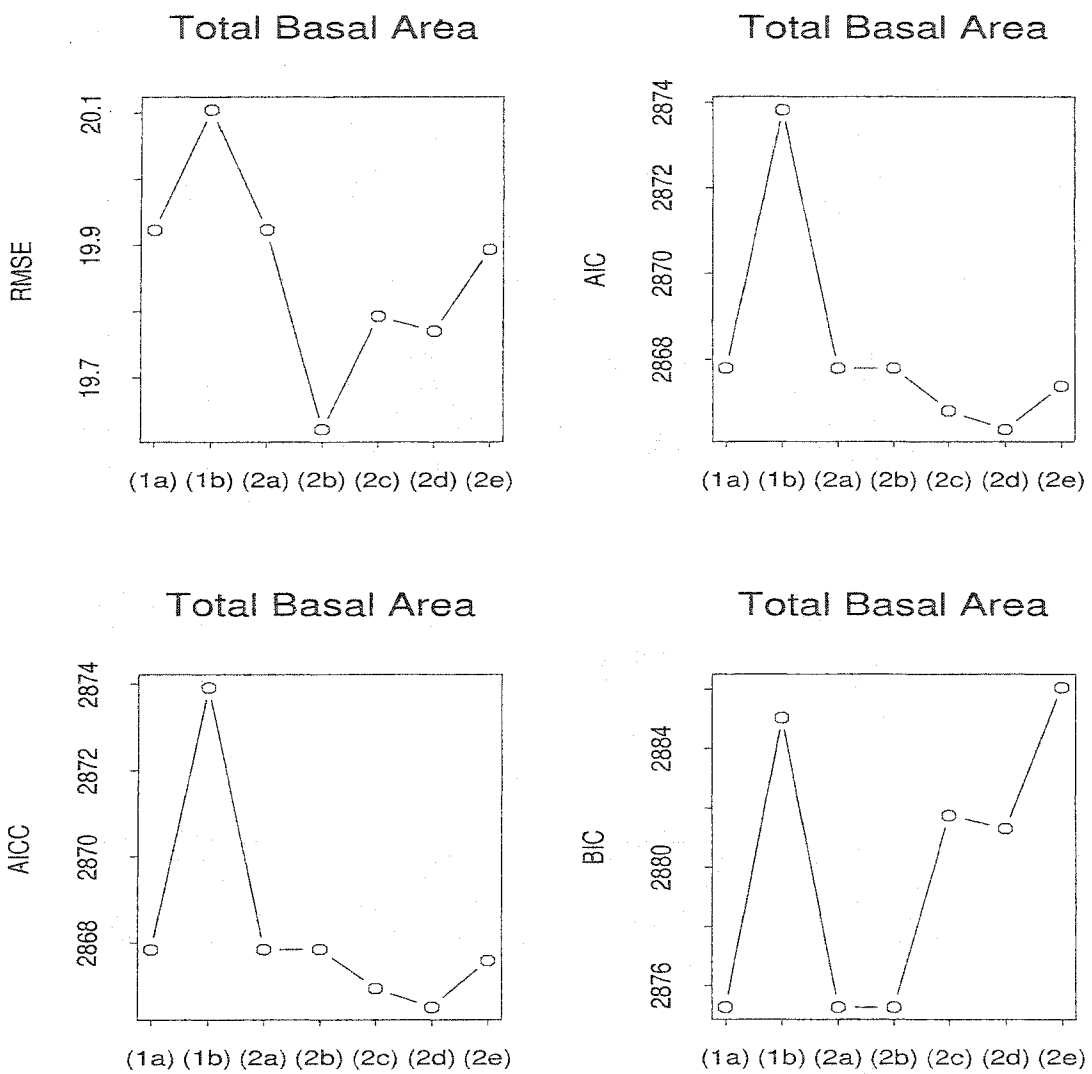
The RMSE, AIC, AICC and BIC statistics are estimated by leave-one-out cross-validation method for the 33 plots with 8 neighbors. The values of x-axis are (1a) assume that $\rho_h = \rho_v = \rho_{hv}$ and $\rho_{c(1)} = \rho_{c(2)} = 0$, (1b) assume that $\rho_{c(1)} = \rho_{c(2)} = 0$, (2a) assume that $\rho_h = \rho_v = \rho_{c(1)} = \rho_{c(2)} = \rho_{hvc}$, (2b) assume that $\rho_h = \rho_v = \rho_{hv}$ and $\rho_{c(1)} = \rho_{c(2)} = \rho_c$, (2c) assume that $\rho_{c(1)} = \rho_{c(2)} = \rho_c$, (2d) assume that $\rho_h = \rho_v = \rho_{hv}$, (2e) assume that $\rho_h, \rho_v, \hat{\rho}_{c(1)}$ and $\hat{\rho}_{c(2)}$ are all different.

Figure 4.3: RMSE, AIC, AICC, and BIC for the Lt data fitted by spatial "Phase 1" models



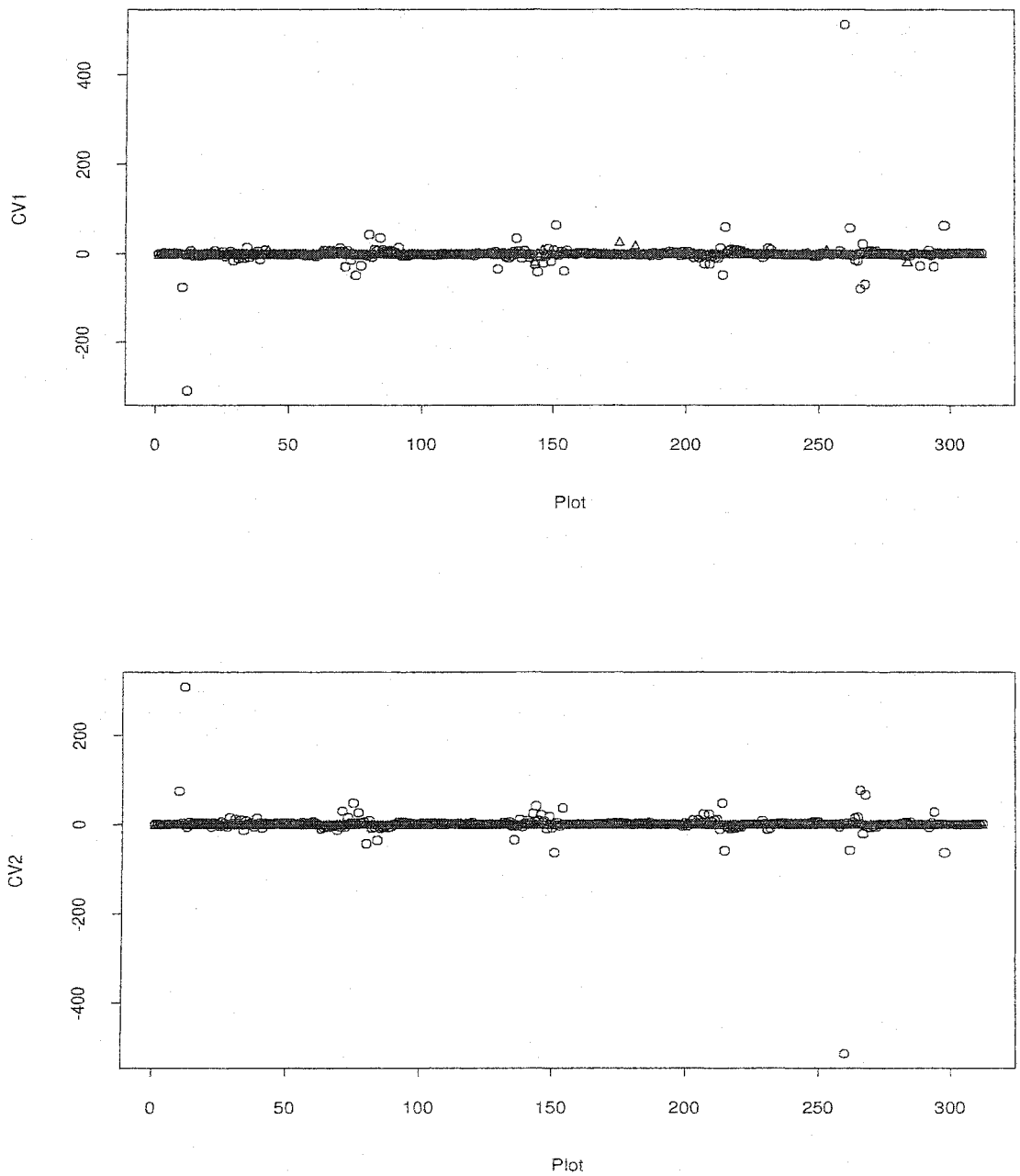
The RMSE, AIC, AICC and BIC statistics are estimated by leave-one-out cross-validation method for the 33 plots with 8 neighbors. The values of x-axis are (1a) assume that $\rho_h = \rho_v = \rho_{hv}$ and $\rho_{c(1)} = \rho_{c(2)} = 0$, (1b) assume that $\rho_{c(1)} = \rho_{c(2)} = 0$, (2a) assume that $\rho_h = \rho_v = \rho_{c(1)} = \rho_{c(2)} = \rho_{hvc}$, (2b) assume that $\rho_h = \rho_v = \rho_{hv}$ and $\rho_{c(1)} = \rho_{c(2)} = \rho_c$, (2c) assume that $\rho_{c(1)} = \rho_{c(2)} = \rho_c$, (2d) assume that $\rho_h = \rho_v = \rho_{hv}$, (2e) assume that $\rho_h, \rho_v, \hat{\rho}_{c(1)}$ and $\hat{\rho}_{c(2)}$ are all different.

Figure 4.4: RMSE, AIC, AICC, and BIC for the Tba data fitted by spatial "Phase 1" models



The RMSE, AIC, AICC and BIC statistics are estimated by leave-one-out cross-validation method for the 33 plots with 8 neighbors. The values of x-axis are (1a) assume that $\rho_h = \rho_v = \rho_{hv}$ and $\rho_{c(1)} = \rho_{c(2)} = 0$, (1b) assume that $\rho_{c(1)} = \rho_{c(2)} = 0$, (2a) assume that $\rho_h = \rho_v = \rho_{c(1)} = \rho_{c(2)} = \rho_{hvc}$, (2b) assume that $\rho_h = \rho_v = \rho_{hv}$ and $\rho_{c(1)} = \rho_{c(2)} = \rho_c$, (2c) assume that $\rho_{c(1)} = \rho_{c(2)} = \rho_c$, (2d) assume that $\rho_h = \rho_v = \rho_{hv}$, (2e) assume that $\rho_h, \rho_v, \hat{\rho}_{c(1)}$ and $\hat{\rho}_{c(2)}$ are all different.

Figure 4.5: Coefficients of Variation of Tba data for individual plots using 400 bootstrap samples



Note: the $\triangle$ denotes the estimates using the simple bootstrap method which is resampling y's and the o denotes the estimates using the semiparametric bootstrap which is resampling the residuals ε's.

5. PREDICTIONS FOR NON-SAMPLED LOCATIONS USING ONLY FIELD SAMPLE PLOT OR SUBPLOT INFORMATION

To make predictions is one of our objectives for this study. In Chapter 4 we only discussed estimations for primary sample units (1-ha plots) on the 1.7-mile grid. This chapter investigates predictions of spatially varying response variables at non-sampled plots on the 0.85-mile grid based on the observations at sampled plots on the 1.7 mile grid, i.e. fill-in predictions or interpolations. Intuitively, a linear combination of the observations is suggested, i.e. a linear prediction, and minimization of mean-squared error (MSE) is used leading to the classical best linear prediction (BLP).

5.1 Best Linear Unbiased Prediction (BLUP)

Suppose we observe $\mathbf{Y}=(Y_{s_1}, Y_{s_2}, \ldots, Y_{s_N}) \equiv (Y_1, Y_2, \ldots, Y_N) \sim (\mu, \sigma^2)$ and wish to predict Y_{s_0} at a non-sampled plot location s_0. Assume $Y_{s_0}, Y_{s_1}, Y_{s_2}, \ldots, Y_{s_N} \sim (\mu, \sigma^2)$ with correlations $\rho_{Y_{s_0},Y_{s_j}} \equiv \rho_{0j}$, $\rho_{Y_{s_i},Y_{s_j}} \equiv \rho_{ij}$ for $i, j = 1, \ldots, N$, $\rho_{Y_{s_i},Y_{s_i}} \equiv \rho_{ii} = 1$; these correlations could be distance based.

Using a linear combination of $\mathbf{Y}$ to predict Y_{s_0}, the prediction can be expressed as

$$\widehat{Y}_{s_0} = \sum_{i=1}^{N} a_i Y_{s_i} \tag{5.1}$$

$$\equiv \sum_{i=1}^{N} a_i Y_i \tag{5.2}$$

If $\sum_i a_i = 1$, then $\mathcal{E}[\widehat{Y}_{s_0}] = \mathcal{E}[Y_{s_0}] = \mu$ and the predictor $\widehat{Y}_{s_0}$ is called unbiased.

Then the mean squared error (MSE) of the prediction is

$$\begin{aligned}\mathcal{E}[(Y_{s_0}-\widehat{Y}_{s_0})^2] &= \mathcal{E}[(Y_{s_0}-\sum_{i=1}^{N}a_iY_i)^2] \\ &= Var[Y_{s_0}]-2Cov[Y_{s_0},\sum_{i=1}^{N}a_iY_i]+Var[\sum_{i=1}^{N}a_iY_i] \\ &= \sigma^2\left\{1-2\sum_{i=1}^{N}a_i\rho_{0i}+\sum_{i=1}^{N}\sum_{j=1}^{N}a_ia_j\rho_{ij}\right\}. \qquad (5.3)\end{aligned}$$

The MSE of prediction is also known as the variance of prediction error or mean squared prediction error of prediction.

The optimal weights $a_1,\ldots,a_N$ can be found using the following Lagrangian function,

$$\sigma^2\left\{1-2\sum_{i=1}^{N}a_i\rho_{0i}+\sum_{i=1}^{N}\sum_{j=1}^{N}a_ia_j\rho_{ij}\right\}-\lambda\left(\sum_{i}^{N}a_i-1\right) \qquad (5.4)$$

which utilizes the constraint $\sum a_i = 1$ ensuring unbiasedness.

Minimizing this expression with respect to $a_1,\ldots,a_N$ and λ yields (Cressie, 1993)

$$\widehat{Y}_{s_0} = \sum_{i=1}^{N}a_iY_i, \qquad (5.5)$$

$$\begin{aligned}\sigma^2_{Y,p}(s_0) &= min\,\mathcal{E}[(Y_{s_0}-\widehat{Y}_{s_0})^2], \\ &= \sigma^2-\mathbf{a}'\mathbf{c}+\lambda/2, \qquad (5.6)\end{aligned}$$

where the subscript p in $\sigma^2_{Y,p}$ indicates the predictor error,

$$\mathbf{a}=(a_1,\ldots,a_N)'\left(\mathbf{c}+\mathbf{1}\frac{1-\mathbf{1}'\Sigma^{-1}\mathbf{c}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}\right)\Sigma^{-1}, \qquad \lambda=-2\frac{1-\mathbf{1}'\Sigma^{-1}\mathbf{c}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}. \qquad (5.7)$$

Here, $\mathbf{c}\equiv\sigma^2(\rho_{01},\ldots,\rho_{0N})'$ and Σ is an $N\times N$ matrix whose (i,j)th elements are $\sigma^2\rho_{ij}$. In order for $\widehat{Y}_{s_0}$ to be a data based predictor, the unknown ρ_{ij}'s in $\mathbf{a}$ and $\mathbf{c}$ will be estimated. It is possible to have a negative value for $\widehat{Y}_{s_0}$ even if all data values are positive. One way to ensure the predictors $\widehat{Y}_{s_0}=\sum_{i=1}^{N}a_iY_i\geq 0$

is to specify $a_i \geq 0$, $i = 1, \ldots, N$, as a further constraint when minimizing Eqn. (5.4). The extra constraint leads to a (perhaps unnecessarily large) increase in mean-squared prediction error. In fact, negative weights a_i's can be advantageous because they don't constrain $\widehat{Y}_{s_0}$ by maximum and minimum data values. And hence, we allow a_i's to be negative.

5.2 3-Parameter Lognormal Transformation

Now suppose that the $Y_{(\cdot)}$ process is obtained from

$$Y_{s_i} = g(Z_{s_i}), \qquad s_i \in D, \tag{5.8}$$

where $Z_{(\cdot)}$ is a Gaussian process that is assumed to be intrinsically stationary and $g(\cdot)$ is a twice-differentiable measurable function. If $d^2g(z)/dz^2 = a + b\,g(z)$, then $Y_{(\cdot)}$ is a generalized lognormal process. The transformation to obtain approximate normality that we applied to M, Lt, and Tba data is the 3-parameter lognormal transformation; that is,

$$Z = \gamma + \delta\, log(Y - \theta'), \tag{5.9}$$

where Z is $N(0,1)$. Thus

$$\begin{aligned} Y &= exp(\frac{Z-\gamma}{\delta}) + \theta', & (5.10)\\ &\equiv g(Z). & (5.11) \end{aligned}$$

The first two derivatives of $g(\cdot)$ are

$$\begin{aligned} g'(z) &= \frac{1}{\delta} exp(\frac{z-\gamma}{\delta}), & (5.12)\\ g''(z) &= \frac{1}{\delta^2} exp(\frac{z-\gamma}{\delta}), & (5.13)\\ &\equiv a + b\,g(z), & (5.14) \end{aligned}$$

where $a = -\frac{\theta'}{\delta^2}$ and $b = \frac{1}{\delta^2}$. So $Y_{(\cdot)}$ is a generalized lognormal process. Shimizu and Iwase (1987) provided an unbiased estimator and its variance for the autocovariance function of a stationary generalized lognormal process.

If $Z_{(\cdot)}$ is $N(\mu_z, \sigma_z^2)$, then the first and second moments of $Y_{(\cdot)} = g\left(Z_{(\cdot)}\right)$ are

$$\begin{aligned} \mathcal{E}[Y] &= \mathcal{E}\left[g\left(Z_{(\cdot)}\right)\right] \\ &= \mathcal{E}\left[exp\left(\frac{Z-\gamma}{\delta}\right)+\theta'\right] \\ &= exp\left(\frac{\mu_z-\gamma}{\delta}+\frac{\sigma_z^2}{2\delta^2}\right)+\theta' \end{aligned} \tag{5.15}$$

and

$$\begin{aligned} \mathcal{E}[Y^2] &= \mathcal{E}\left\{\left[exp\left(\frac{Z-\gamma}{\delta}\right)+\theta'\right]^2\right\} \\ &= \mathcal{E}\left\{\left[exp\left(\frac{Z-\gamma}{\delta}\right)\right]^2+2\theta'\left[exp\left(\frac{Z-\gamma}{\delta}\right)\right]+(\theta')^2\right\} \\ &= exp\left(\frac{2(\mu_z-\gamma)}{\delta}+\frac{2\sigma_z^2}{\delta^2}\right)+2\theta' exp\left(\frac{\mu_z-\gamma}{\delta}+\frac{\sigma_z^2}{2\delta^2}\right)+(\theta')^2. \end{aligned}$$

And so

$$Var[Y] = exp\left(\frac{2(\mu_z-\gamma)}{\delta}+\frac{\sigma_z^2}{\delta^2}\right)\left\{exp\left(\frac{\sigma_z^2}{\delta^2}\right)-1\right\}. \tag{5.16}$$

The covariance function is

$$Cov[Y_{s_i}, Y_{s_j}] = exp\left(\frac{2(\mu_z-\gamma)}{\delta}+\frac{\sigma_z^2}{\delta^2}\right)\left\{exp\left(\frac{\sigma_z^2\rho_{ij}}{\delta^2}\right)-1\right\}. \tag{5.17}$$

5.3 Correlation Function Fitting

Suppose an autocorrelation function is distance-related, say $\rho(d)$. Possible criteria for fitting this autocorrelation function are the Least Squares options, i.e. ordinary least squares (O.L.S) and weighted least squares (W.L.S).

Using the sum of squares of the difference between the sample correlation estimator $\rho^{\#}(d)$ and the model $\rho(d)$ that has parameters θ and k, the method of ordinary least squares specifies that θ and k are estimated by minimizing

$$\sum_{"d"}\left\{\rho^{\#}(d)-\rho(d)\right\}^2 \tag{5.18}$$

for lag distance d.

The W.L.S. method of autocorrelation-model fitting is finding the parameters θ and k that minimize

$$\sum_{"d"} N(d) \left\{ \rho^{\#}(d) - \rho(d) \right\}^2 \tag{5.19}$$

for lag distance d and $N(d)$ is the number of pairs of observations with lag distance d. The estimates $\hat{\theta}$ and $\hat{k}$ can be obtained from an iterative, nonlinear estimation routine such as the Gauss-Newton algorithm in the Splus nlminb function.

5.3.1 Correlation function fitting to the Y scale

First, we fit the autocorrelation function to the sample correlation among the Y scale observations. The autocorrelation functions we consider are listed as follows:

(I)(1) $\rho_{ij}(Y) = exp(-\theta d_{ij}^k)$, $\theta, k > 0$,

(I)(2) $\rho_{ij}(Y) = (1 + \theta d_{ij}^2)^{(-k)}$, $\theta, k > 0$.

where d_{ij} is the distance between sites s_i and s_j. Such a correlation function fitting we call "direct" correlation function fitting to the Y scale.

5.3.2 Correlation function fitting to the Z scale

If the process generating Y is 3-parameter lognormal, then there is a relation between the autocorrelation function of the transformed Z scale and that of the untransformed Y scale. The autocorrelations ρ_{ij} and $\rho_{ij}(Y)$ respectively denote the correlations of (Z_{s_i}, Z_{s_j}) and (Y_{s_i}, Y_{s_j}). If we fit the following autocorrelation functions to the Y scale,

(II)(1) $\rho_{ij}(Y) = exp(-\theta d_{ij}^k)$, $\theta, k > 0$,

(II)(2) $\rho_{ij}(Y) = (1 + \theta d_{ij}^2)^{(-k)}$, $\theta, k > 0$

then the corresponding autocorrelation functions for the Z scale are respectively

(III)(1) $\rho_{ij} = \frac{\delta^2}{\sigma_z^2} log\left(1 + e^{-\theta d_{ij}^k}(e^{\sigma_z^2/\delta^2} - 1)\right), \theta, k > 0,$

(III)(2) $\rho_{ij} = \frac{\delta^2}{\sigma_z^2} log\left(1 + (1 + \theta d_{ij}^2)^{-k}(e^{\sigma_z^2/\delta^2} - 1)\right), \theta, k > 0.$

The procedure is to fit the autocorrelation functions (III)(1) and (III)(2) to the sample correlations of the Z scale and then substitute the estimated θ and k into the autocorrelation functions of Y's, (II)(1) and (II)(2). Such a correlation function fitting we call "indirect" correlation function fitting to the Y scale.

The autocorrelation model fitting for M, Lt, and Tba on 1-ha plots are portrayed in Figures 5.1-5.3. The O.L.S. and W.L.S. methods yield parameter estimates that are very close. The sample correlations decay quickly and so the weights don't change the sum of squares much. From Figures 5.1-5.3, we can see that there is not much difference between the O.L.S. fitting and the W.L.S. fitting. And hence, we only use the weighted least square method in our further inference.

5.4 Predictions

5.4.1 Predictions based on the Y scale

There are no specific distribution assumption for the Y's in Eqns. (5.5)-(5.7). If the observed process is a 3-parameter lognormal process, do predictions based on the Y's with the indirect fitting correlation functions work better than those with the direct correlation fitting?

5.4.1.1 Predictions based on the Y scale with direct correlation fitting

The coefficients a_i in Eqn. (5.2) are estimated from the data and so a data based predictor is given by

$$\widetilde{Y}_{s_0} = \sum_{i=1}^{N} \hat{a}_i Y_i \tag{5.20}$$

$$\widehat{\sigma}^2_{Y,p}(s_0) = \widehat{\sigma}^2 - \widehat{\mathbf{a}}\,\widehat{\mathbf{c}} + \widehat{\lambda}/2, \tag{5.21}$$

where

$$\widehat{\mathbf{a}} = \left(\widehat{\mathbf{c}} + \mathbf{1}\frac{1 - \mathbf{1}'\widehat{\Sigma}^{-1}\widehat{\mathbf{c}}}{\mathbf{1}'\widehat{\Sigma}^{-1}\mathbf{1}}\right)\widehat{\Sigma}^{-1}, \qquad \hat{\lambda} = -2\frac{1 - \mathbf{1}'\widehat{\Sigma}^{-1}\mathbf{c}}{\mathbf{1}'\widehat{\Sigma}^{-1}\mathbf{1}},$$

$\widehat{\mathbf{c}} = \widehat{\sigma}^2(\hat{\rho}_{01}(Y), \ldots, \hat{\rho}_{0N}(Y))'$ and $\widehat{\Sigma} = \widehat{\sigma}^2\hat{\rho}_{ij}(Y)$. Here, $\hat{\rho}_{ij}(Y)$ is an estimate by direct autocorrelation fittings to the Y scale, (I)(1) and (I)(2).

5.4.1.2 Predictions based on the Y scale with indirect correlation fitting

a_i is a function of σ_Y^2 and $\rho_{ij}(Y)$, so $\widetilde{Y}_{s_0,I}$ is

$$\widetilde{Y}_{s_0,I} = \sum_{i=1}^{N} \hat{a}_i Y_i \tag{5.22}$$

$$\widehat{\sigma}^2_{Y,p,I}(s_0) = \widehat{\sigma}^2 - \widehat{\mathbf{a}}\,\widehat{\mathbf{c}} + \widehat{\lambda}/2, \tag{5.23}$$

where the subscript I in $\widetilde{Y}_{s_0,I}$ indicates the prediction based on the Y scale with the indirect autocorrelation fitting,

$$\widehat{\mathbf{a}} = \left(\widehat{\mathbf{c}} + \mathbf{1}\frac{1 - \mathbf{1}'\widehat{\Sigma}^{-1}\widehat{\mathbf{c}}}{\mathbf{1}'\widehat{\Sigma}^{-1}\mathbf{1}}\right)\widehat{\Sigma}^{-1}, \qquad \hat{\lambda} = -2\frac{1 - \mathbf{1}'\widehat{\Sigma}^{-1}\mathbf{c}}{\mathbf{1}'\widehat{\Sigma}^{-1}\mathbf{1}},$$

$\widehat{\mathbf{c}} = \widehat{\sigma}^2(\hat{\rho}_{01}(Y), \ldots, \hat{\rho}_{0N}(Y))'$ and $\widehat{\Sigma} = \widehat{\sigma}^2\hat{\rho}_{ij}(Y)$. Here, $\hat{\rho}_{ij}(Y)$ is estimated by indirect autocorrelation fittings to the Y scale, (II)(1) and (II)(2). The correlation fittings of (II)(1) and (II)(2) are based on the assumption that Y is a 3-parameter lognormal process, so the prediction $\widetilde{Y}_{s_0,I}$ requires a stronger assumption than the previous one, $\widetilde{Y}_{s_0}$.

5.4.2 Back-transformed predictions based on the Z scale

5.4.2.1 "Plain" Back-Transformed Prediction

To transform the prediction from the Z scale to the Y scale, the back-transformed value $g(\widehat{Z}) \equiv g(\sum_{i=1}^{N} a_i Z_{s_i})$ is used to predict Y_{s_0}, where $a_1, \ldots, a_N$ satisfy Eqn. (5.7) on the transformed Z scale and need to be estimated.

The expectation of $g\left(\widehat{Z}_{s_0}\right)$ is

$$\begin{aligned}
\mathcal{E}\left[g\left(\widehat{Z}_{s_0}\right)\right] &= exp\left(\frac{\mathcal{E}[\widehat{Z}_{s_0}]-\gamma}{\delta}+\frac{Var[\widehat{Z}_{s_0}]}{2\delta^2}\right)+\theta', \\
&= exp\left(\frac{\mu_z-\gamma}{\delta}+\frac{Var[\widehat{Z}_{s_0}]}{2\delta^2}\right)+\theta', \qquad (5.24) \\
&\neq g(\mu)
\end{aligned}$$

where $\mathcal{E}[\widehat{Z}_{s_0}] = \mathcal{E}[\sum_{i=1}^{N} a_i Z_{s_i}] = \mu_z$. And so the back-transformed predictor,

$$\begin{aligned}
\widehat{Y}_{s_0,b} &\equiv g\left(\widehat{Z}_{s_0};\Theta\right). \qquad \text{where } \Theta=(\gamma,\delta,\theta') \\
&= g\left(\sum_{i=1}^{N} a_i Z_{s_i}\right) \\
&= exp\left(\frac{\widehat{Z}_{s_0}-\gamma}{\delta}\right)+\theta' \qquad (5.25)
\end{aligned}$$

is not an unbiased predictor. The bias of $\widehat{Y}_{s_0,b}$ is

$$\begin{aligned}
& bias\left(\widehat{Y}_{s_0,b}\right) \\
&= \mathcal{E}[\widehat{Y}_{s_0,b}]-\mathcal{E}[Y_{s_0}] \\
&= exp\left(\frac{\mu_z-\gamma}{\delta}\right)\left\{exp\left(\frac{Var[\widehat{Z}_{s_0}]}{2\delta^2}\right)-exp\left(\frac{\sigma_z^2}{2\delta^2}\right)\right\} \\
&= exp\left(\frac{\mu_z-\gamma}{\delta}\right)\left\{exp\left(\frac{\sigma_z^2-\sigma_{z,p}^2(s_0)+\lambda_z}{2\delta^2}\right)-exp\left(\frac{\sigma_z^2}{2\delta^2}\right)\right\} \qquad (5.26)
\end{aligned}$$

Thus an alternative back-transformed predictor is:

$$\begin{aligned}
\widetilde{Y}_{s_0,b} &\equiv g\left(\widetilde{Z}_{s_0};\Theta\right), \qquad \text{where } \Theta=(\gamma,\delta,\theta') \\
&= g\left(\sum_{i=1}^{N} \hat{a}_i Z_{s_i}\right) \\
&= exp\left(\frac{\widetilde{Z}_{s_0}-\gamma}{\delta}\right)+\theta' \qquad (5.27)
\end{aligned}$$

where $\hat{a}_i$'s are estimators of a_i's.

The mean-squared prediction error of $\widetilde{Y}_{s_0,b}$ is

$$\begin{aligned}
&\sigma^2_{Y,p,b}(s_0) \\
&= \mathcal{E}[(Y_{s_0} - \widetilde{Y}_{s_0,b})^2] \\
&= Var[Y_{s_0}] - 2Cov[Y_{s_0}, \widetilde{Y}_{s_0,b}] + Var[\widetilde{Y}_{s_0,b}] + \left(\mathcal{E}[(Y_{s_0}] - \mathcal{E}[\widetilde{Y}_{s_0,b}]\right)^2 \\
= \; & exp\left(\frac{2(\mu_z-\gamma)}{\delta} + \frac{\sigma_z^2}{\delta^2}\right)\left\{exp\left(\frac{\sigma_z^2}{\delta^2}\right) - 1\right\} \\
& -2\,exp\left(\frac{2(\mu_z-\gamma)}{\delta} + \frac{\sigma_z^2 + Var[\widetilde{Z}_{s_0}]}{2\delta^2}\right)\left\{exp\left(Cov\left[\frac{Z_{s_0}-\gamma}{\delta}, \frac{\widetilde{Z}_{s_0}-\gamma}{\delta}\right]\right) - 1\right\} \\
& +exp\left(\frac{2(\mu_z-\gamma)}{\delta} + \frac{Var[\widetilde{Z}_{s_0}]}{\delta^2}\right)\left\{exp\left(\frac{Var[\widetilde{Z}_{s_0}]}{\delta^2}\right) - 1\right\} \\
& +exp\left(\frac{2(\mu_z-\gamma)}{\delta}\right)\left\{exp\left(\frac{\sigma_z^2}{2\delta^2}\right) - exp\left(\frac{\sigma_z^2 - \sigma^2_{z,p}(s_0) + \lambda_z}{2\delta^2}\right)\right\}^2 \\
= \; & exp\left(\frac{2(\mu_z-\gamma)}{\delta} + \frac{\sigma_z^2}{\delta^2}\right)\left\{\left[exp\left(\frac{\sigma_z^2}{\delta^2}\right) - 1\right] - 2exp\left(\frac{\lambda_z - \sigma^2_{z,p}(s_0)}{2\delta^2}\right)\right. \\
& \left[exp\left(\frac{\sigma_z^2 - \sigma^2_{z,p}(s_0) + \lambda_z/2}{\delta^2}\right) - 1\right] + exp\left(\frac{\lambda_z - \sigma^2_{z,p}(s_0)}{\delta^2}\right) \\
& \left.\left[exp\left(\frac{\sigma_z^2 - \sigma^2_{z,p}(s_0) + \lambda_z}{\delta^2}\right) - 1\right] + \left\{1 - exp\left(\frac{-\sigma^2_{z,p}(s_0) + \lambda_z}{2\delta^2}\right)\right\}^2\right\} \\
= \; & exp\left(\frac{2(\mu_z-\gamma)}{\delta} + \frac{2\sigma_z^2}{\delta^2}\right)\left\{1 - 2exp\left(\frac{-3\sigma^2_{z,p}(s_0) + 2\lambda_z}{2\delta^2}\right)\right. \\
& \left. +exp\left(\frac{-2(\sigma^2_{z,p}(s_0) - \lambda_z)}{\delta^2}\right)\right\} \qquad (5.28)
\end{aligned}$$

Thus, the estimator of $\sigma^2_{Y,p,b}(s_0)$ is given by

$$\begin{aligned}
\hat{\sigma}^2_{Y,p,b}(s_0) \;=\; & exp\left(\frac{2(\hat{\mu}_z - \hat{\gamma})}{\hat{\delta}} + \frac{2\hat{\sigma}_z^2}{\hat{\delta}^2}\right)\left\{1 - 2exp\left(\frac{-3\hat{\sigma}^2_{z,p}(s_0) + 2\hat{\lambda}_z}{2\hat{\delta}^2}\right)\right. \\
& \left. +exp\left(\frac{-2(\hat{\sigma}^2_{z,p}(s_0) - \hat{\lambda}_z)}{\hat{\delta}^2}\right)\right\} \qquad (5.29)
\end{aligned}$$

In real data, we don't know the actual values of Θ and so the estimated Θ is used. Thus we investigate the following two predictors,

$$\widehat{\widehat{Y}}_{s_0,b} \equiv g\left(\widehat{Z}_{s_0}; \widehat{\Theta}\right) \qquad (5.30)$$

$$\widehat{\widehat{Y}}_{s_0,b} \equiv g\left(\widetilde{Z}_{s_0}; \widehat{\Theta}\right) \qquad (5.31)$$

where $\widehat{Z}_{s_0} = \sum_{i=1}^{N} a_i Z_{s_i}$, $\widetilde{Z}_{s_0} = \sum_{i=1}^{N} \hat{a}_i Z_{s_i}$ and $\widehat{\Theta} = (\hat{\gamma}, \hat{\delta}, \widehat{\theta'})$

Since the straightforward back-transformed predictor, $\widehat{Y}_{s_0,b}$ is a biased predictor, a bias correction is made in the following subsection.

5.4.2.2 "Approximately" Unbiased Back-Transformed (Trans-Gaussian) Prediction

Using the δ-method around $\mu_z = \mathcal{E}[Z_{s_i}]$ to obtain a bias correction, the approximately unbiased prediction of Y_{s_0} is (See Cressie, 1993, p.137 for details),

$$\begin{aligned}
\widehat{Y}_{s_0,a} &= g(\widehat{Z}_{s_0}; \Theta) + \frac{g''(\mu_z)}{2}\{\sigma^2_{z,p}(s_0) - \lambda_z\} \\
&= exp\left(\frac{\widehat{Z}_{s_0} - \gamma}{\delta}\right) + \theta' + \frac{1}{2\delta^2} exp\left(\frac{\mu_z - \gamma}{\delta}\right)\{\sigma^2_{z,p}(s_0) - \lambda_z\}, \quad (5.32) \\
&= \widehat{Y}_{s_0,b} + \frac{1}{2\delta^2} exp\left(\frac{\mu_z - \gamma}{\delta}\right)\{\sigma^2_{z,p}(s_0) - \lambda_z\}, \quad (5.33)
\end{aligned}$$

where the subscript a in $\widehat{Y}_{s_0,a}$ indicates the approximately unbiased prediction, where (see Eqns. (5.5)-(5.7)), $\mu_z = \mathcal{E}[Z]$, $\sigma^2_{z,p}(s_0) = {\sigma_z}^2 - \mathbf{ac} + \lambda_z/2$, and $\Theta = (\gamma, \delta, \theta')$ are parameters in 3-parameter lognormal likelihood of the Y's to be estimated by the maximum likelihood estimation from the 3-parameter lognormal likelihood based on the Y observations. When $\sigma^2_{z,p}(s_0) > \lambda_z$, $\widehat{Y}_{s_0,a} > \widehat{Y}_{s_0,b}$. From Eqns. (5.21) and (5.28), the bias of $\widehat{Y}_{s_0,a}$ is

$$\begin{aligned}
& bias\left(\widehat{Y}_{s_0,a}\right) \\
&= bias\left(\widehat{Y}_{s_0,b}\right) + \frac{1}{2\delta^2} exp\left(\frac{\mu_z - \gamma}{\delta}\right)\{\sigma^2_{z,p}(s_0) - \lambda_z\}. \quad (5.34)
\end{aligned}$$

The mean-squared prediction error is, approximately,

$$\begin{aligned}
\sigma^2_{Y,p,a}(s_0) &= \{g'(\mu_z)\}^2 \sigma^2_{z,p}(s_0) \\
&= \frac{1}{\delta^2} exp\left(\frac{2(\mu_z - \gamma)}{\delta}\right)\sigma^2_{z,p}(s_0). \quad (5.35)
\end{aligned}$$

The approximations (5.32) and (5.35) rely on $\sigma^2_y(s_0)$ being small. The higher order approximation to $\sigma^2_{Y,p,a}(s_0)$ is needed when $\sigma^2_y(s_0)$ is not small.

The following predictors are also considered for Trans-Gaussian setting

$$\widetilde{Y}_{s_0,a} = g(\widetilde{Z};\Theta) + \frac{g''(\mu_z)}{2}\{\sigma^2_{z,p}(s_0) - \lambda_z\} \quad (5.36)$$

$$\widehat{\widetilde{Y}}_{s_0,a} = g(\widehat{Z};\widehat{\Theta}) + \frac{g''(\hat{\mu}_z)}{2}\{\hat{\sigma}^2_{z,p}(s_0) - \hat{\lambda}_z\} \quad (5.37)$$

$$\widehat{\widetilde{Y}}_{s_0,a} = g(\widetilde{Z};\widehat{\Theta}) + \frac{g''(\hat{\mu}_z)}{2}\{\hat{\sigma}^2_{z,p}(s_0) - \hat{\lambda}_z\} \quad (5.38)$$

The estimator of $\sigma^2_{Y,p,a}(s_0)$ is

$$\hat{\sigma}^2_{Y,p,a}(s_0) = \frac{1}{\hat{\delta}^2} exp\left(\frac{2(\hat{\mu}_z - \hat{\gamma})}{\hat{\delta}}\right)\hat{\sigma}^2_{z,p}(s_0), \quad (5.39)$$

where the estimators from Eqns. (5.5)-(5.7) $\hat{\mu}_z = \overline{Z}$, $\hat{\sigma}^2_{z,p}(s_0) = \hat{\sigma}^2_z - \hat{\mathbf{a}}\hat{\mathbf{c}} + \hat{\lambda}_z/2$ and $\hat{\gamma}$, $\hat{\delta}$, and $\widehat{\theta'}$ are given by maximum likelihood estimation from the 3-parameter lognormal likelihood based on the Y observations.

5.4.2.3 Unbiased Back-Transformed Prediction

Cressie (1993, p.135) discussed lognormal kriging. In the following, we extend his results to 3-parameter lognormal unbiased predictions. Multiplying Eqn. (5.19) by a correction constant, an unbiased predictor of Y_{s_0} is given by

$$\begin{aligned}\widehat{Y}_{s_0,u} &= exp\left\{\frac{\widehat{Z}_{s_0} - \gamma}{\delta} + \frac{\sigma^2_z - Var[\widehat{Z}_{s_0}]}{2\delta^2}\right\} + \theta' \\ &= exp\left\{\frac{\widehat{Z}_{s_0} - \gamma}{\delta} + \frac{1}{2\delta^2}\left(\sigma^2_{z,p}(s_0) - \lambda_z\right)\right\} + \theta'. \end{aligned} \quad (5.40)$$

When $\theta' = 0$, γ=1, and $\delta = 1$, the unbiased 3-parameter lognormal predictor yields the lognormal kriging in Cressie (1993, p.135).

The following predictors are also considered for the 3-parameter lognormal situation

$$\widetilde{Y}_{s_0,u} = exp\left\{\frac{\widetilde{Z}_{s_0} - \gamma}{\delta} + \frac{1}{2\delta^2}\left(\sigma^2_{z,p}(s_0) - \lambda_z\right)\right\} + \theta' \quad (5.41)$$

$$\widehat{\widetilde{Y}}_{s_0,u} = exp\left\{\frac{\widehat{Z}_{s_0} - \hat{\gamma}}{\hat{\delta}} + \frac{1}{2\hat{\delta}^2}\left(\hat{\sigma}^2_{z,p}(s_0) - \hat{\lambda}_z\right)\right\} + \widehat{\theta'} \quad (5.42)$$

$$\widehat{\widetilde{Y}}_{s_0,u} = exp\left\{\frac{\widetilde{Z}_{s_0} - \hat{\gamma}}{\hat{\delta}} + \frac{1}{2\hat{\delta}^2}\left(\hat{\sigma}^2_{z,p}(s_0) - \hat{\lambda}_z\right)\right\} + \widehat{\theta'} \quad (5.43)$$

The mean-squared prediction error is

$$
\begin{aligned}
& \sigma^2_{Y,\,p\,u}(s_0) \\
= \;& \mathcal{E}[(Y_{s_0} - \widehat{Y}_{s_0,\,u})] \\
= \;& Var[Y_{s_0}] - 2Cov[Y_{s_0}, \widehat{Y}_{s_0,\,u}] + Var[\widehat{Y}_{s_0,\,u}] \\
= \;& exp\left(\frac{2(\mu_z - \gamma)}{\delta} + \frac{\sigma_z^2}{\delta^2}\right)\left\{exp\left(\frac{\sigma_z^2}{\delta^2}\right) - 1\right\} - 2\,exp\left(\frac{\sigma_z^2 - Var[\widehat{Z}_{s_0}]}{2\delta^2}\right) \\
& exp\left(\frac{2(\mu_z - \gamma)}{\delta} + \frac{\sigma_z^2 + Var[\widehat{Z}_{s_0}]}{2\delta^2}\right)\left\{exp\left(Cov\left[\frac{Z_{s_0} - \gamma}{\delta}, \frac{\widehat{Z}_{s_0} - \gamma}{\delta}\right]\right) - 1\right\} \\
& + exp\left(\frac{2(\mu_z - \gamma)}{\delta} + \frac{\sigma_z^2}{\delta^2}\right)\left\{exp\left(\frac{Var[\widehat{Z}_{s_0}]}{\delta^2}\right) - 1\right\} \\
= \;& exp\left(\frac{2(\mu_z - \gamma)}{\delta} + \frac{\sigma_z^2}{\delta^2}\right)\left\{exp\left(\frac{\sigma_z^2}{\delta^2}\right) - 2exp\left(\frac{1}{\delta^2}Cov[Z_{s_0}, \widehat{Z}_{s_0}]\right)\right. \\
& \left. + exp\left(\frac{Var[\widehat{Z}_{s_0}]}{\delta^2}\right)\right\} \\
= \;& exp\left(\frac{2(\mu_z - \gamma)}{\delta} + \frac{\sigma_z^2}{\delta^2}\right)\left\{exp\left(\frac{\sigma_z^2}{\delta^2}\right) - 2exp\left(\frac{\sigma_z^2 - \sigma^2_{z,\,p}(s_0) + \lambda_z/2}{\delta^2}\right)\right. \\
& \left. + exp\left(\frac{\sigma_z^2 - \sigma^2_{z,\,p}(s_0) + \lambda_z}{\delta^2}\right)\right\}. \qquad (5.44)
\end{aligned}
$$

Then the estimator of $\sigma^2_{Y,\,p,\,u}(s_0)$ is

$$
\begin{aligned}
& \hat{\sigma}^2_{Y,\,p,\,u}(s_0) \\
= \;& exp\left(\frac{2(\hat{\mu}_z - \hat{\gamma})}{\hat{\delta}} + \frac{\hat{\sigma}_z^2}{\hat{\delta}^2}\right)\left\{exp\left(\frac{\hat{\sigma}_z^2}{\hat{\delta}^2}\right) - 2exp\left(\frac{\hat{\sigma}_z^2 - \hat{\sigma}^2_{z,\,p}(s_0) + \hat{\lambda}_z/2}{\hat{\delta}^2}\right)\right. \\
& \left. + exp\left(\frac{\hat{\sigma}_z^2 - \hat{\sigma}^2_{z,\,p}(s_0) + \hat{\lambda}_z}{\hat{\delta}^2}\right)\right\} \qquad (5.45)
\end{aligned}
$$

where the estimators from Eqns. (5.5)-(5.7), $\hat{\mu}_z = \overline{Z}$, $\hat{\sigma}^2_{z,\,p}(s_0) = \hat{\sigma}_z^2 - \hat{\mathbf{a}}\hat{\mathbf{c}} + \hat{\lambda}_z/2$, and $\hat{\gamma}$, $\hat{\delta}$, and $\hat{\theta}'$ are given by maximum likelihood estimation from the 3-parameter lognormal likelihood based on the Y observations.

To compare the predictors, we rewrite the unbiased predictor as

$$\begin{aligned}\widehat{Y}_{s_0,u} &= \widehat{Y}_{s_0,a} + exp\left(\frac{\widehat{Z}_{s_0}-\gamma}{\delta}\right)\left\{exp\left(\frac{\sigma^2_{z,p}(s_0)-\lambda_z}{2\delta^2}\right)\right. \\ &\quad \left. -1-exp\left(\frac{\mu_z-\widehat{Z}_{s_0}}{\delta}\right)\frac{\sigma^2_{z,p}(s_0)-\lambda_z}{2\delta^2}\right\} \qquad (5.46) \\ &\equiv \widehat{Y}_{s_0,a} + D_{\widehat{Y}_{s_0,u},\widehat{Y}_{s_0,a}}. \qquad (5.47)\end{aligned}$$

$exp\left(\frac{\mu_z-\widehat{Z}_{s_0}}{\delta}\right)$ is not always greater than 1 and so the second term, $D_{\widehat{Y}_{s_0,u},\widehat{Y}_{s_0,a}}$, is not always positive. Hence, $\widehat{Y}_{s_0,u}$ is not always larger than $\widehat{Y}_{s_0,a}$.

The relation between $\widehat{Y}_{s_0,u}$ and $\widehat{Y}_{s_0,b}$ is

$$\begin{aligned}\widehat{Y}_{s_0,u} &= \widehat{Y}_{s_0,b} + exp\left(\frac{\widehat{Z}_{s_0}-\gamma}{\delta}\right)\left\{exp\left(\frac{\sigma^2_{z,p}(s_0)-\lambda_z}{2\delta^2}\right)-1\right\} \qquad (5.48) \\ &\equiv \widehat{Y}_{s_0,b} + D_{\widehat{Y}_{s_0,u},\widehat{Y}_{s_0,b}}. \qquad (5.49)\end{aligned}$$

When $\sigma^2_{z,p}(s_0) > \lambda_z$, $D_{\widehat{Y}_{s_0,u},\widehat{Y}_{s_0,b}} > 0$ and $\widehat{Y}_{s_0,u} > \widehat{Y}_{s_0,b}$. As mentioned earlier, $\widehat{Y}_{s_0,a} > \widehat{Y}_{s_0,b}$ when $\sigma^2_{z,p}(s_0) > \lambda_z$. In general, if $\sigma^2_{z,p}(s_0) > \lambda_z$, then both $\widehat{Y}_{s_0,u}$ and $\widehat{Y}_{s_0,a}$ are larger than $\widehat{Y}_{s_0,b}$.

The estimated mean sqaured prediction errors (m.s.p.e.) of those three predictors seem quite different. To examine them more closely, the Taylor series expansion is used to break down some $exp(\cdot)$ functions of Eqn. (5.39). Through first order expansion,

$$\begin{aligned}&\sigma^2_{Y,p,u}(s_0) \\ =\ & exp\left(\frac{2(\mu_z-\gamma)}{\delta}\right)\left\{exp\left(\frac{2\sigma_z^2}{\delta^2}\right)-2exp\left(\frac{2\sigma_z^2-\sigma^2_{z,p}(s_0)+\lambda_z/2}{\delta^2}\right)\right. \\ & \left. +exp\left(\frac{2\sigma_z^2-\sigma^2_{z,p}(s_0)+\lambda_z}{\delta^2}\right)\right\} \\ \approx\ & exp\left(\frac{2(\mu_z-\gamma)}{\delta}\right)\left\{\left(1+\frac{2\sigma_z^2}{\delta^2}\right)-2\left(1+\frac{2\sigma_z^2-\sigma^2_{z,p}(s_0)+\frac{\lambda}{2}}{\delta^2}\right)\right. \\ & \left. +\left(1+\frac{2\sigma_z^2-\sigma^2_{z,p}(s_0)+\lambda}{\delta^2}\right)\right\} \\ =\ & \sigma^2_{Y,p,a}(s_0). \qquad (5.50)\end{aligned}$$

When $\sigma_z^2(s_0)$ is not small, the remainder of first order Taylor series expansion for $\sigma^2_{Y,p,u}(s_0)$ might be a large-scale quantity compared to $\sigma^2_{Y,p,a}(s_0)$.

5.5 Results

5.5.1 Simulated data

To assess how well the various predictors introduced in this chapter behave, we conducted several simulations generating populations with the estimated parameters from the real data. In the first place, six non-sampled plots on the 0.85 mile grid, shown by black numbers in Figure 5.4, were randomly selected for making predictions and then a realization Z_i, i= $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N, \boldsymbol{s}_{N+1}, \ldots, \boldsymbol{s}_{N+6}$ from a multivariate normal was generated by

$$\mathbf{Z} = \boldsymbol{\mu} + B\boldsymbol{\varepsilon} \tag{5.51}$$

where $\mathbf{Z} = (Z_{\boldsymbol{s}_1}, \ldots, Z_{\boldsymbol{s}_{N+6}})'$, $\boldsymbol{\mu}$=$(\mu_{\boldsymbol{s}_1}, \ldots, \mu_{\boldsymbol{s}_{N+6}})' = (\mu_z, \ldots, \mu_z)$, $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N$ are sampled plot locations, $\boldsymbol{s}_{N+1}, \ldots, \boldsymbol{s}_{N+6}$ are six non-sampled plot locations, B is defined via the $(N+6) \times (N+6)$ correlation matrix $\Gamma = \left(exp(-\theta d_{ij}^k)\right)_{i,j} = BB'$, and $\boldsymbol{\varepsilon} \sim$ MVN($\mathbf{0}$, σ_z^2 $\mathbf{I}$). Secondly, the Z values are back-transformed by the function $g(\cdot)$

$$\begin{aligned} Y &= g(Z) \\ &= exp(\frac{Z-\gamma}{\delta}) + \theta' \end{aligned}$$

As mentioned earlier, the actual $\Theta = (\gamma, \delta, \theta')$ is unknown, so the predictors, $\widehat{Y}_{s_0,(.)}$ and $\widetilde{Y}_{s_0,(.)}$ are not available for real data. So the predictors $\widehat{\widehat{Y}}_{s_0,(.)}$ and $\widehat{\widetilde{Y}}_{s_0,(.)}$ are used instead. To assess the loss, we conduct a simulation study with chosen values of all parameters, μ, σ^2, Θ and (θ, k) in the ρ_{ij} function.

For M data, the MLE's of γ, δ, and θ' are -4.50, 1.16, and -7.50, respectively. The W.S.L. estimates of θ and k in the exponential autocorrelation for

3-parameter lognormal transformed M data are $\hat{\theta} = 0.12$ and $\hat{k} = 0.39$. To mimic the process we had for the actual M data, the values of parameters chosen were $\mu_z = 0, \sigma_z = 1, \theta = 0.12, k = 0.39, \gamma = -4.50, \delta = 1.16$, and $\theta' = -7.50$. The back-transformed $\mathbf{Y}$ process is a generalized lognormal process with mean $\mu_y = 62.67$ and variance $\sigma_y^2 = 5428.58$. The results of three predictors based on 1,000 realizations of $\mathbf{Y}$ are summarized in Table 5.1. $\widehat{\mathcal{E}}[\widehat{Y}_{s_0,b}] = \overline{\widehat{Y}}_{s_0,b} = \sum_{i=1}^{1,000} \widehat{Y}^i_{s_0,b}$ is listed in Table 5.1(a) where i indicates the $\widehat{Y}^i_{s_0,b}$ value obtained from the ith generated $\mathbf{Y}_i = (Y^i_{s_1}, \ldots, Y^i_{s_N})$ realization. $\overline{\widehat{Y}}_{s_0,a}$ and $\overline{\widehat{Y}}_{s_0,u}$ are defined similarly and $\overline{\widetilde{Y}}_{s_0,(\cdot)}$, $\overline{\widehat{\widetilde{Y}}}_{s_0,(\cdot)}$, and $\overline{\widehat{\widehat{Y}}}_{s_0,(\cdot)}$ denote the average of 1,000 $\widetilde{Y}_{s_0,(\cdot)}$, $\widehat{\widetilde{Y}}_{s_0,(\cdot)}$, and $\widehat{\widehat{Y}}_{s_0,(\cdot)}$, respectively.

For each realization of $\mathbf{Y}$, the simulated $Y^i_{s_j}$'s, $j = N+1, \ldots, N+6$, were put aside and the predictions, $\widehat{Y}^i_{s_0,b}$, $\widehat{Y}^i_{s_0,a}$, and $\widehat{Y}^i_{s_0,u}$, $s_0 = s_{N+1}, \ldots, s_{N+6}$, were obtained by substituting the simulated $\mathbf{Y}_i = (Y^i_{s_1}, \ldots, Y^i_{s_N})$ in the prediction and mean sqaured prediction error equations. The values of $\overline{\widehat{Y}}_{s_0,u}$ at each non-sampled plot locations are almost the same as the actual mean $\mu_y = 62.67$; which is not surprising since $\widehat{Y}_{s_0,u}$ is an "unbiased" predictor. The quotation for unbiased indicates that $\widehat{Y}_{s_0,u}$ is asymptotically unbiased since the a's are estimated. Also, we note that the values of $\overline{\widehat{Y}}_{s_0,a}$ are close to the actual mean and the bias is about 6% of the actual mean. The biggest bias belongs to the back-transformation predictor, $\widehat{Y}_{s_0,b}$, which is about one-third of the actual mean.

The estimated standard error (s.e.) for $\overline{\widehat{Y}}_{s_0,b}$ is calculated by

$$\widehat{s.e.}(\widehat{Y}_{s_0,b}) = \frac{\sum_{i=1}^{1,000} \left(\widehat{Y}^i_{s_0,b} - \overline{\widehat{Y}}_{s_0,b}\right)^2}{1,000 - 1}.$$

The estimated standard errors (s.e.) for $\overline{\widehat{Y}}_{s_0,a}$ and $\overline{\widehat{Y}}_{s_0,u}$ are calculated similarly. $\widehat{s.e.}(\widehat{Y}_{s_0,(\cdot)})$ is not a good estimator of $s.e.(\widehat{Y}_{s_0,(\cdot)})$ since the $\widehat{Y}^i_{s_0,(\cdot)}$'s are not independent. We note that the values for each plot in columns of $\overline{\hat{\sigma}}_{Y,p,(\cdot)}$ are almost the same for each m.s. prediction error. That's because the $\hat{\sigma}_{Y,p,(\cdot)}$'s don't depend

directly on the $\mathbf{Y}_i$'s like the $\widehat{Y}^i_{s_0,(.)}$'s do and only rely on the $\hat{\mu}_z$, $\hat{\sigma}^2_z$, and $\hat{\Gamma}$. Those selected non-sampled plots have the common estimated γ, δ, θ', μ_z, σ^2_z, θ, and k for each generated M realization and so $\hat{\lambda}$ is not different.

Table 5.1(b) presents the estimated m.s. prediction errors from the simulated data which is calculated by the following equation

$$\widehat{m.s.e.}(\widehat{Y}_{s_0,(.)}) = \frac{1}{1,000}\sum_{i=1}^{1,000}\left(Y^i_{s_0} - \widehat{Y}_{s_0,(.)}\right)^2$$

where $s_0 = s_{N+1}, \ldots, s_{N+6}$.

The approximately unbiased predictor, $\widehat{Y}_{s_0,a}$, has the smallest estimated m.s. prediction error, $\widehat{m.s.e.}(\widehat{Y}_{s_0,a})$, but differs little from the others. Compared to the values in column 6 of Table 5.1(a), the approximately estimated m.s. prediction error, $\hat{\sigma}^2_{Y,(.)}(s_0)$ is very different from $\widehat{m.s.e.}(\widehat{Y}_{s_0,a})$. As mentioned before, the approximation $\hat{\sigma}^2_{Y,p,a}$ relies on small σ^2_y. The actual value of σ^2_y for the simulated M realization is 5428.58, so $\hat{\sigma}^2_{Y,p,b}(s_0)$ and $\hat{\sigma}^2_{Y,p}(s_0)$ are more reliable. For plot 1, the m.s. prediction error of $\widehat{Y}_{s_0,b}$, has interval estimate given by $(5594.98 - 1.96 \times 1426.76, 5594.98 + 1.96 \times 1426.76) = (2741.46, 8448.50)$; the m.s. prediction error of $\widehat{Y}_{s_0,a}$, has interval estimate given by $(1659.15 - 1.96 \times 283.17, 1659.15 + 1.96 \times 283.17) = (1092.81, 2225.49)$; the m.s. prediction error of $\widehat{Y}_{s_0,u}$, has interval estimate given by $(5511.03 - 1.96 \times 1302.55, 5511.03 + 1.96 \times 1302.55) = (2905.93, 8116.13)$. The $\widehat{m.s.e.}(\widehat{Y}_{s_0,b})$, $\widehat{m.s.e.}(\widehat{Y}_{s_0,a})$, and $\widehat{m.s.e.}(\widehat{Y}_{s_0,u})$ for plot 1, are respectively 4124.04, 3853.11, and 3929.26. The actual m.s. prediction errors of $\widehat{m.s.e.}(\widehat{Y}_{s_0,b})$ and $\widehat{m.s.e.}(\widehat{Y}_{s_0,u})$ are inside the m.s. prediction error interval, but that of $\widehat{Y}_{s_0,a}$ is outside the interval. Thus, the estimated m.s. prediction error of $\widehat{Y}_{s_0,a}$ does not work well when σ^2_y is not small. Among the six selected non-sampled plots, all of the $\widehat{m.s.e.}(\widehat{Y}_{s_0,b})$ and $\widehat{m.s.e.}(\widehat{Y}_{s_0,u})$ are inside the m.s. prediction error interval, but none of the $\widehat{m.s.e.}(\widehat{Y}_{s_0,a})$ is inside the interval. In this case, $\hat{\sigma}^2_{Y,p,a}$ is the worst estimator.

To mimic the Tba observations in simulations, the values of parameters we chose were $\mu_z = 0$, $\sigma_z = 1$, $\theta = 1.03$, $k = 0.13$, $\gamma = -34.41$, $\delta = 6.76$, and $\theta' = -120.65$. The back-transformed $\mathbf{Y}$ process is a generalized lognormal with mean $\mu_y = 43.57$ and variance $\sigma_y^2 = 596.62$. The results are summarized in Tables 5.2 and 5.3. In reality, we can only obtain the predictions $\widehat{\widehat{Y}}_{s_0,b}$, $\widehat{\widetilde{Y}}_{s_0,a}$, and $\widehat{\widehat{Y}}_{s_0}$, since we don't know what actual values of Θ and $\boldsymbol{a}$ are. As shown in Table 5.2(b), the means of $\widehat{\widehat{Y}}_{s_0,b}$ at each plot are slightly off the actual mean $\mu_y = 43.57$, but the means of $\widehat{\widehat{Y}}_{s_0,a}$ and $\widehat{\widehat{Y}}_{s_0}$ are close to the actual mean. The means of $\hat{\sigma}^2_{\widehat{\widehat{Y}},p,b}$, $\hat{\sigma}^2_{\widehat{\widehat{Y}},p,a}$, and $\hat{\sigma}^2_{\widehat{\widehat{Y}},p}$ are about 570, 454, and 567, respectively, which are lower than the $\widehat{m.s.e}(Y_{p,N+i})$ in Table 5.3. The largest estimated m.s.e. belongs to plot 5 and so it is harder to predict than the others.

From Table 5.2 we note that the means of 1,000 $\widehat{Y}_{s_0,a}$, $\widetilde{Y}_{s_0,a}$, $\widehat{\widetilde{Y}}_{s_0,a}$, $\widehat{\widehat{Y}}_{s_0,a}$, $\widehat{Y}_{s_0}$ $\widetilde{Y}_{s_0}$, $\widehat{\widetilde{Y}}_{s_0}$, and $\widehat{\widehat{Y}}_{s_0}$ are very close, and they are approximately equal to the actual mean $\mu_y = 43.57$; the "plain" back-transformation predictors, $\widehat{Y}_{s_0,b}$, $\widetilde{Y}_{s_0,b}$, $\widehat{\widetilde{Y}}_{s_0,b}$, and $\widehat{\widehat{Y}}_{s_0,b}$, as shown in the 3rd column of Table 5.2, are only about 4% off the actual mean. Unlike the previous simulation results for simulated M realizations, the estimated m.s. prediction errors for three predictors, shown in Table 5.3, are about the same for "plain", "approximately", and "unbiased" back-transform predictors. From Eqns. 5.39 and 5.50, we note that the m.s. prediction errors are equivalent through the first order when σ_y^2 is small. This is noted in the simulation results for Tba with small variance.

The histograms of the generated values and the predictions at six selected plots using $\widehat{Y}_{s_0,b}$, $\widetilde{Y}_{s_0,b}$, $\widehat{\widetilde{Y}}_{s_0,b}$, and $\widehat{\widehat{Y}}_{s_0,b}$ are depicted in Figure 5.5. From Figure 5.5(a), the histogram of the generated values at plots 1 and 2 have normality and the histograms of the generated values at the other plots have a slightly heavy tail. The histograms of predictions $\widehat{\widehat{Y}}_{s_0,b}$, $\widehat{\widehat{Y}}_{s_0,a}$, and $\widehat{\widehat{Y}}_{s_0}$ don't have the same shape as those of the generated values at each plot.

To see how stable prediction and mean squared prediction error (m.s.p.e) are, we plot the average of b predictions and m.s.p.e.'s from b realizations, $b = 100, 200, \ldots, 1,000$, as shown in Figures 5.6 and 5.7. Also, the estimated mean squared prediction error was calculated by $\frac{1}{b}\sum_{i=1}^{b}(Y^b_{p,N+i} - Y^b_{g,N+i})^2$, $b = 100, 200, \ldots, 1{,}000$, where $Y^b_{p,N+i}$, $i = 1, \ldots, 6$ denote the prediction at non-sampled plot i using those three prediction rules based on the generated values at sampled plots, $1, \ldots, N = 312$ from the b^{th} generated realization of size $(N + 6)$, and $Y^b_{g,N+i}$ denotes the b^{th} generated values at non-sampled plot i. The results are depicted in Figure 5.8. From Figure 5.6, we note that the average of the predictions using those prediction rules are pretty stable as b increases, but the averages of asymptotic back-transform are smaller than "plain" and unbiased" back-transform and a little bit off from the actual mean $\mu_y = 43.57$.

5.5.2 Real data

The six non-sampled plots, shown by black numbers in Figure 5.4, on the 0.85 mile grid were randomly selected for making predictions. The 1,000 predictions at those non-sampled plots were obtained by the simple bootstrap method and the semiparametric bootstrap; the simple bootstrap method resamples N values from the N observations with replacement and the semiparametric bootstrap method resamples N values from the independent components with replacement (as described in Chapter 4). Figures 5.5 display the histograms of approximately unbiased predictions $\widehat{Y}_{s_0,a}$ and unbiased predictions $\widehat{Y}_{s_0,u}$ at the selected non-sampled plots using the simple bootstrap method. The histograms of $\widehat{Y}_{s_0,a}$ and $\widehat{Y}_{s_0}$ at selected non-sampled plots have a similar shape. The histograms of the predictions calculated by the two prediction rules based on the semiparametric bootstrapped samples are portrayed in Figure 5.6. The histograms of predictions

based on the semiparametric bootstrapped samples are quite different from those based on the simple bootstrapped sample shown in Figure 5.5.

The results of 1,000 M predictions at each non-sampled plots from bootstrapped samples are summarized in Table 5.6. The averages of unbiased M predictions based on the simple bootstrapped samples ,$\overline{\hat{Y}}_{s_0,(.)}$, are larger than those of the approximately unbiased M predictions, as shown in the top panel of Table 5.6. But the estimated standard deviations of $\overline{\widehat{Y}}_{s_0,(.)}$ from the simple bootstrapped samples are smaller than those of $\widehat{Y}_{s_0,(.)}$ from the semiparametric bootstrapped samples. For both simple bootstrapped and semiparametric bootstrapped samples, $\overline{\hat{\sigma}}^2_{Y,p,a}$ is much smaller than the other standard deviations $\overline{\hat{\sigma}}^2_{Y,p,b}$ and $\overline{\hat{\sigma}}^2_{Y,p}$. As we expected, $\overline{\hat{\sigma}}^2_{Y,p,a}$ is not a good estimator when σ^2_y is large (see Eqns. 5.39 and 5.50).

Mean squared (m.s.) prediction errors play a key role in measuring the reliability of predictions. Table 5.1 also lists the plug-in estimates from Eqns. (5.19) and (5.28). To investigate the unbiasedness of m.s prediction error estimates, the mean and standard deviation of 1,000 $\hat{\sigma}^2_{Y,p}$ and $\hat{\sigma}^2_{Y,p,a}$ are calculated. For example, the plug-in estimate of m.s. prediction error for $\widehat{Y}_{s_0,a}$ at plot 1 is 1665.62. Using the semiparametric bootstrap method, the average of 1,000 estimated m.s. prediction error is 1665.76 which is quite close to the plug-in estimates, but the standard deviation of 1,000 estimated m.s. prediction error is 273.19. So the m.s. prediction error of the M value at plot 1 has the interval estimate given by $(1665.52 - 1.96 \times 273.19, 1665.52 + 1.96 \times 273.19) = (1119.14, 2211.9)$. For unbiased prediction, $\widehat{Y}_{s_0,u}$, the actual m.s. prediction error has interval estimate given by (4271.71, 12547.67). The difference in m.s. prediction error of the predictions, $\widehat{Y}_{s_0,a}$ and $\widehat{Y}_{s_0,u}$, is very large. Is it worth while to get an unbiased estimate and pay such a huge price on m.s. prediction error? To examine this, the leave-one-out

simple bootstrap method and leave-one-out semiparametric bootstrap method are used to predict the one withheld for each of the actual N observations.

The intervals discussed earlier are approximate normal prediction intervals (PIs). We will use the bootstrap-t method to calculate better approximate prediction intervals. As $N \to \infty$, the bootstrap and normal approximate intervals converge to each other, but in general the bootstrap-t prediction intervals provide better approximations than the normal approximate prediction intervals. The bootstrap-t method is built by generating b bootstrapped samples and then computing the bootstrap prediction of Y_{s_0} for each samples. For example, if $b = 1,000$, the estimate of the 2.5% point is the 25th largest of the the $Y_{s_0}^b$s and the estimate of the 97.5% point is the 975th largest of the $Y_{s_0}^b$s. Finally, the "bootstrap-t" prediction interval with level $\alpha = 0.05$ is

$$\left(\hat{Y}_{s_0} - \hat{t}_{(1-\alpha)}\,\widehat{mspe}, \hat{Y}_{s_0} + \hat{t}_{(\alpha)}\,\widehat{mspe}\right).$$

Also, we investigate the stability of prediction intervals for Y_{s_0} as the bootstrap size b grows large.

Table 5.16 summarizes the results for normal and bootstrap-t prediction intervals for Tba at 6 selected non-sampled plot locations. The bootstrapped samples from $b = 1,000, 2,000, \ldots, 10,000$, were conducted, but we only summarize the results for $b = 1,000$ and $b = 10,000$ bootstrapped samples because there is no dramatical change as b increases. The lower and upper endpoints of bootstrap-t prediction intervals for Tba at non-sampled plots greatly differ from the those of normal prediction intervals. For bootstrap-t prediction intervals, the endpoints of the first column in Table 5.16, $\tilde{Y}_{s_0}$, are much greater than the semiparametric bootstrap method. The columns of $\tilde{Y}_{s_0,l}$, $\tilde{Y}_{s_0,b}$, $\tilde{Y}_{s_0,a}$, and $\tilde{Y}_{s_0,u}$ in Table 5.16, greatly differ from those based on semiparametric bootstrapped samples. This

is because those prediction equations are based on the indirect correlation fitting and the semiparametric method assumes the errors are correlated.

5.5.3 Prediction evaluation

In this section, cross-validated mean squared error (CVMSE) is used to evaluate the performance of predictions for plot based data and subplot based data.

5.5.3.1 Plot based data

The total number of plots used is N=313. Each time we leave out a plot and use the prediction rules based on the remaining 312 plot observations to predict it. Finally, we calculate the MSE by

$$MSE = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 / N$$

where y_i is the actual value and $\hat{y}_i$ is the estimated value at plot i using the prediction rules based on the remaining $(N-1)$ observations. The estimated cross-validated MSE (CVMSE) for M, Lt, and Tba data are summarized in Table 5.17. We note that there is not much improvement in MSE for different prediction rules. For M data, the reduction in MSE is very slight. The improvement for Lt data is better than that for M data, but the best is only a 2% reduction. The predictions do not work any better than using the sample mean for Tba. This is because the coefficients in each prediction rules are about the same as those in the sample mean, $1/(N-1) = 0.0032$. As mentioned in Chapter 3, the correlation among plot based data is not large enough to reduce the MSE. We next examine whether the prediction rules using subplot data show some improvement over using just the sample mean.

5.5.3.2 Subplot based data

The total number of subplots we used is $N = 1470$. The procedure for leave-one-out cross validation is very time-consuming, so we only randomly selected 150 subplots (about 10%) to estimate the cross-validated MSE. The results are summarized in Table 5.18.

For Lt data, we note that exponential correlation fitting based on the Y scale with direct and indirect correlation fitting, seems better no matter which predictor equations we chose. Under direct fitting of sample correlations there is no big difference between Eqn. (5.20) with correlation fittings (I)(1) and (I)(2). For indirect fitting Eqn (5.22) with indirect fitting (II)(1) and (II)(2), (II)(1) is much better than (II)(2). The back-Transformed Predictions based on the Z scale assume that there is not only a specific correlation function between the data but also that there is a 3-parameter lognormal transformation between the untransformed Y scale and the transformed Z scale. For all three back-transformed predictions based on the Z scale, the predictions with (III)(2) are much worse than those with (III)(1). The "unbiased" transformed back-transformed prediction has the estimated MSE of 2953.69 (versus the estimated MSE for Eqn. (5.20) with (I)(1), 2982.51). We didn't reduce the MSE with the stronger assumption under (III). All predictions with both direct and indirect exponential fitting function are much better than the predictions using just the sample mean to predict under the leave-one-subplot-out procedure,

For Tba data, no matter what correlation functions we fit, the estimated CV MSEs for all prediction rules are not much different; they ranged from 23.42 to 25.87. The estimated MSEs for correlation fitting (II) are slightly better than those for the other correlation fittings. The estimated CVMSE for predictions using just the sample mean of the $(N-1)$ observation as the prediction at the leave-out subplot is 45.90. In general, all prediction rules have an approximate

44% reduction in MSE relative to using just the sample mean. The assumptions for prediction rules (I)(1) and (I)(2) are weaker than the other prediction rules, but the estimated CVMSEs are as good as the others. Thus predictions rules (I) are suggested to predict for Tba at non-sampled subplots.

As to M data, we note that prediction rules (I) only improve estimation slightly. The prediction rules (II)-(III) do not improve estimation at all compared to the sample mean predictions. The "unbiased" back-transformed predictions have large MSE. It's because the subplot based M data does not follow a 3-parameter lognormal. When we treat it as a 3-parameter lognormal, the "bias" corrections force the predictions to be very different.

The coefficient a_i's in each prediction rules range from 0.6 through 0.2, which are much larger than the weights in the sample mean, $1/(N-1)$ =0.0007. Also, the a_i coefficients allow neighboring subplot observations to have negative weights for the predictions. If the model assumptions hold, the prediction rules (I)-(III) are better than the sample mean. For Lt data the reduction in MSE is about 33% and for Tba data it is approximately 44%.

Among the predictors, direct, indirect, "plain" back-transformed, "approximately" unbiased back-transformed, and "unbiased" back-transformed predictors, the smallest prediction error belonged to the direct predictor, $\widetilde{Y}_{s_0}$. The direct predictor $\widetilde{Y}_{s_0}$ doesn't assume that $Y(.)$ follows a Gaussian process as a kriging predictor does; and doesn't assume there is a lognormal relationship between the untransformed Y's and the transformed Z's. The predictors described in this chapter have a closed form for the prediction error of an individual prediction. To see how reliable they are, the simple bootstrap and semiparameter bootstrap methods are suggested for the case as follows:

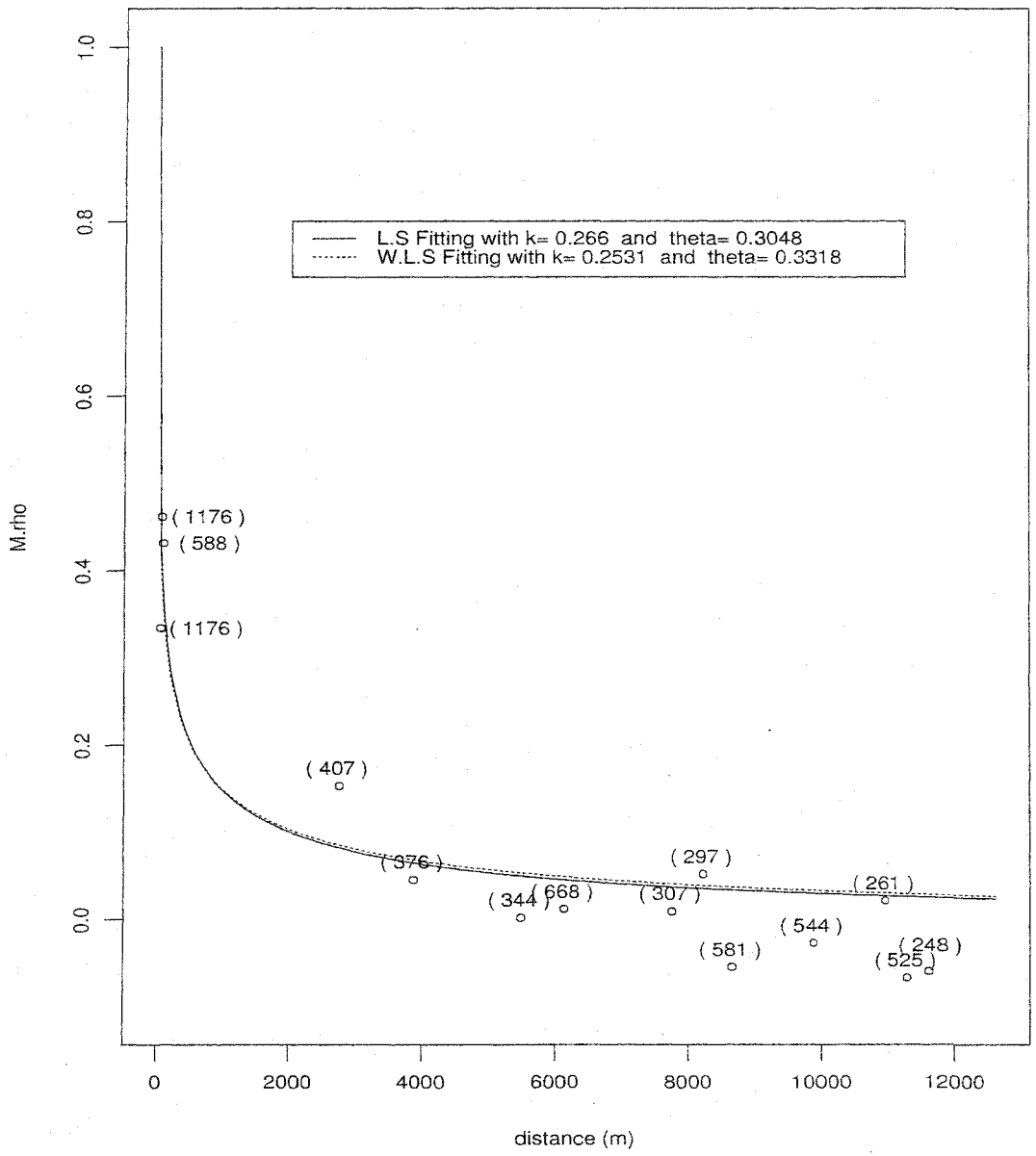
(a) if there exists strong spatial dependence among the data, the semiparametric bootstrap method is suggested, or,

(b) if not, the simple bootstrap method will be ok and it is not as time-consuming as the semiparametric bootstrap method.

The prediction errors of the M predictions at 6 non-sampled plot locations on the 0.85 mile are about the same value of $\sqrt{5320} \approx 73$. And the averages of 1,000 bootstrapped prediction errors using simple bootstrap and semiparametric bootstrap methods are pretty close to the estimates of prediction error by substituting the N observations and the estimated a_i based on the N observations.

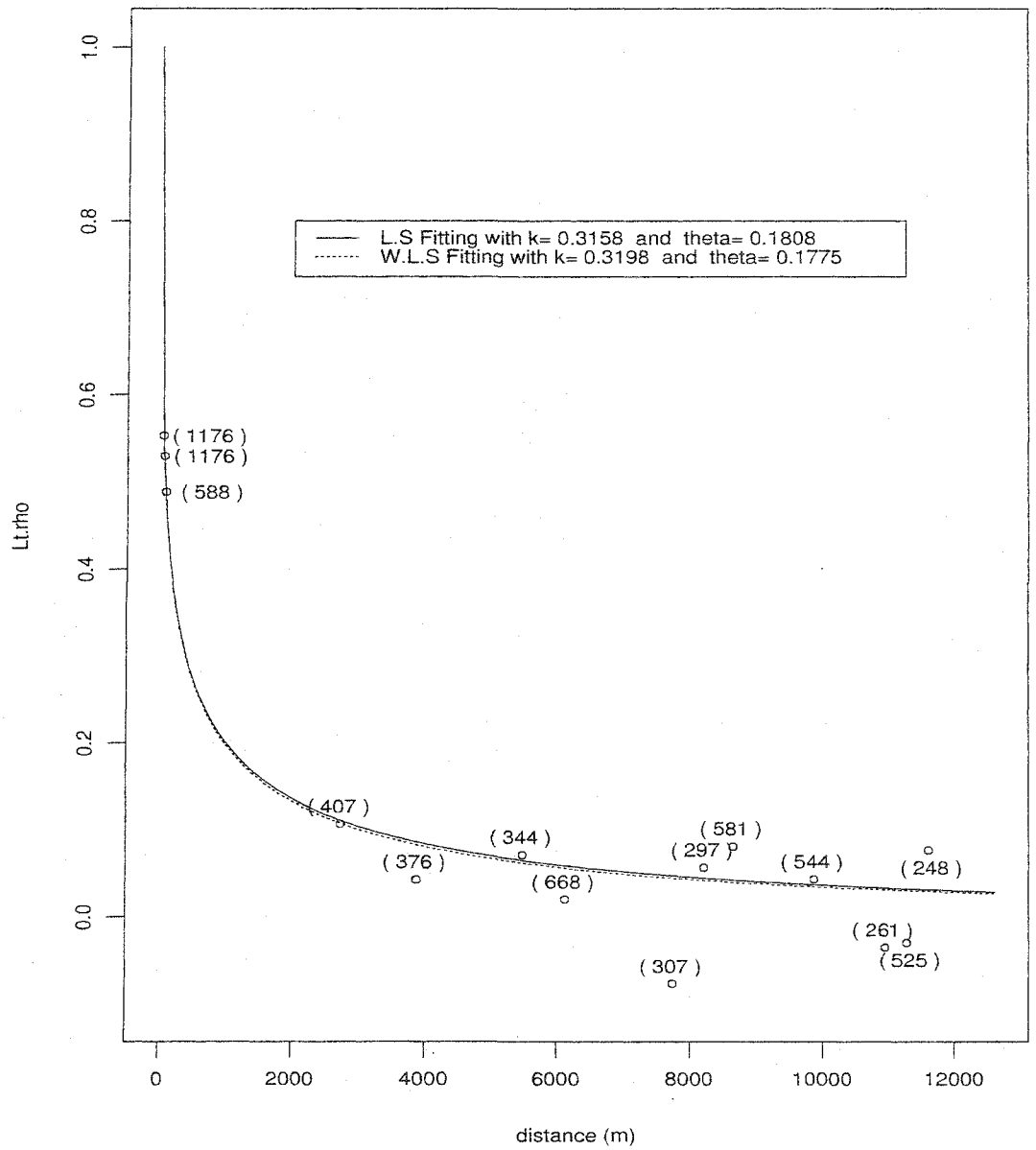
In summary, for subplot based data the prediction rules work well if the model is correctly specified, but for plot based data the prediction rules do not work well even if the model is correctly specified. The predictions based on plot data don't work well because adjacent plots contain little or no information (i.e. there is essentially no spatial dependence at the plot scale of the data) about each other. From correlation fittings, we can see that there is a strong spatial dependence at the lag distance less than 0.85 mile and then the spatial correlation drops quickly beyond that. Even the correlation between 1.7-mile neighbors is only about 0.2. Plots need to be a lot closer together to have useful spatial information or maybe subplots should be used as primary sampling units instead of plots. Also, it is not necessary to sample all five of the subplots for each plot to have useful spatial dependence for predictions.

Figure 5.1: Sample correlations and the fitted exponential function for M data[1]



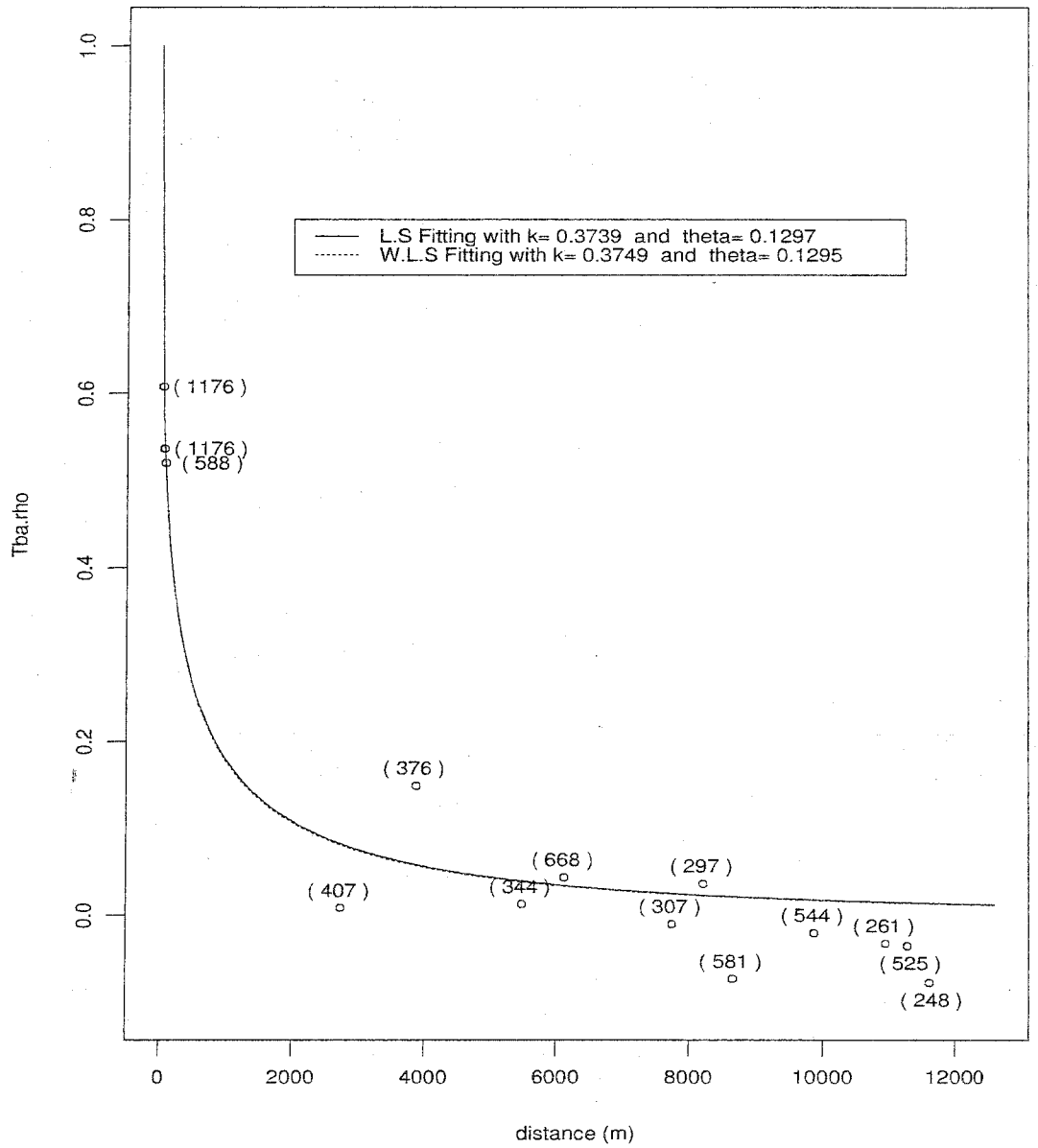
[1] the correlation function is $\rho_{ij} = exp(-\theta d_{ij}^{k})$

Figure 5.2: Sample correlations and the fitted exponential function for Lt data[1]



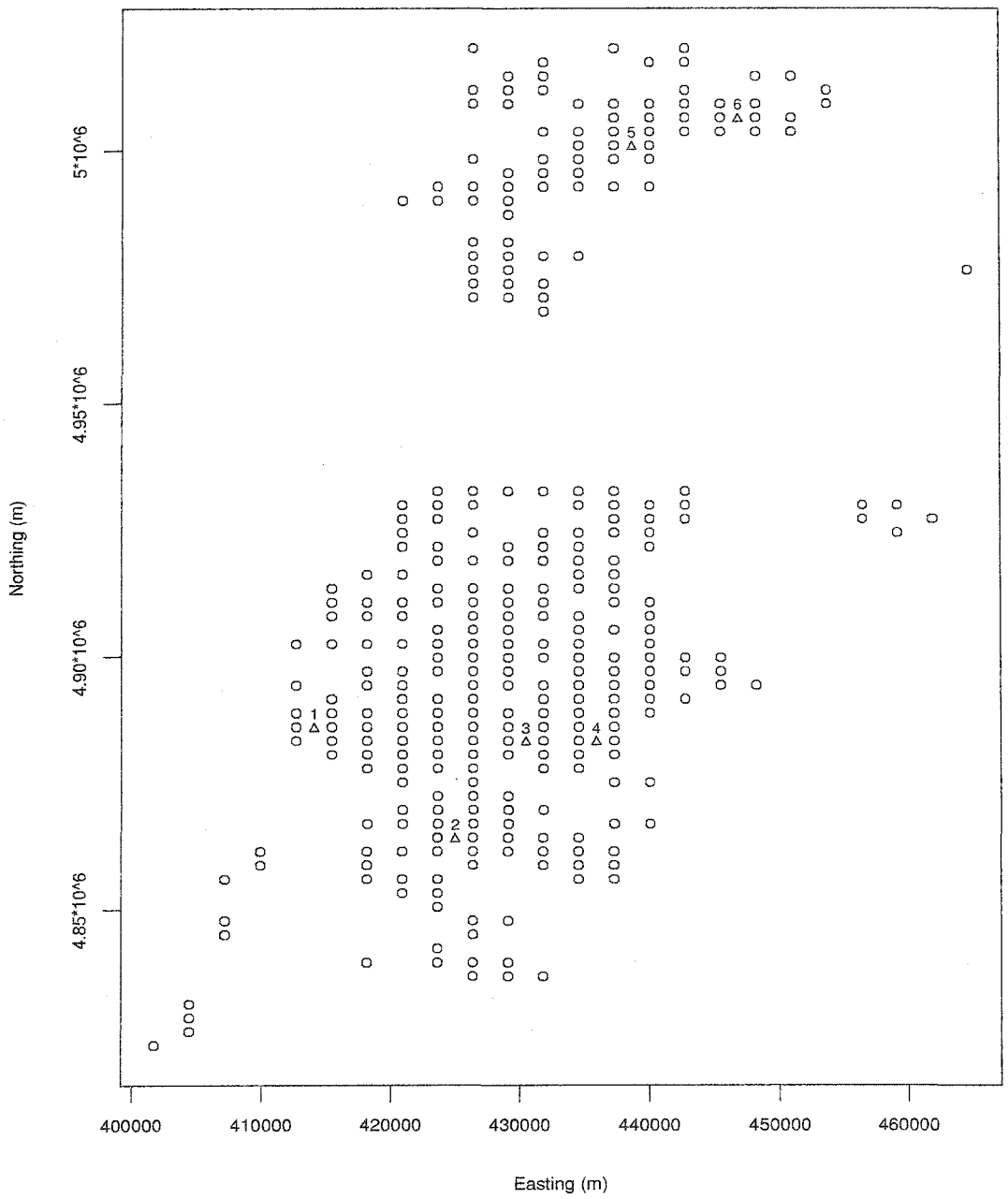
[1] the correlation function is $\rho_{ij} = exp(-\theta d_{ij}^{k})$

Figure 5.3: Sample correlations and the fitted exponential function for Tba data[1]



[1] the correlation function is $\rho_{ij} = exp(-\theta d_{ij}^{k})$

Figure 5.4: Locations of all sample plots and six non-sample plots [1]



[1] o indicates the sampled plots and △ indicates the selected non-sampled plots

Table 5.1: The three predictions for 1,000 simulated M realizations

plot	$\overline{\hat{Y}}_{s_0,b}$	$\overline{\hat{Y}}_{s_0,a}$	$\overline{\hat{Y}}_{s_0}$	$\overline{\hat{\sigma}}^2_{Y,p,b}$	$\overline{\hat{\sigma}}^2_{Y,p,a}$	$\overline{\hat{\sigma}}^2_{Y,p}$
1	41.45	58.64	62.42	5994.98	1659.15	5511.03
	(11.44)	(12.05)	(16.27)	(1426.76)	(283.17)	(1302.55)
2	41.76	58.92	62.78	5990.20	1657.52	5507.38
	(10.04)	(10.55)	(14.16)	(1426.28)	(282.99)	(1302.13)
3	42.13	59.27	63.25	5987.83	1657.08	5506.34
	(10.63)	(10.98)	(14.61)	(1426.05)	(283.02)	(1302.14)
4	41.74	58.90	62.80	5992.62	1658.60	5509.79
	(8.94)	(9.52)	(12.78)	(1426.41)	(283.07)	(1302.36)
5	42.23	59.40	63.39	5991.94	1657.65	5507.66
	(11.97)	(12.20)	(15.85)	(1426.50)	(282.93)	(1302.06)
6	41.68	58.86	62.69	5992.42	1657.45	5507.19
	(9.37)	(9.86)	(13.08)	(1426.83)	(283.03)	(1302.17)

(a). $\hat{\mathcal{E}}[\overline{\hat{Y}}_{s_0,(.)}]$, $\hat{\mathcal{E}}[\hat{\sigma}^2_{Y,p,(.)}(s_0)]$, and their estimated s.e., $\widehat{s.e.}(\hat{Y}_{s_0,(.)})$ and $\widehat{s.e.}(\hat{\sigma}^2_{Y,p,(.)})$ (in parenthesis) based on 1,000 M realizations

plot	$\widehat{m.s.e.}(\hat{Y}_{s_0,b})$	$\widehat{m.s.e.}(\hat{Y}_{s_0,a})$	$\widehat{m.s.e.}(\hat{Y}_{s_0})$
1	4124.04	3853.11	3929.26
2	4583.31	4232.80	4269.08
3	5553.45	5141.50	5154.93
4	7299.96	7033.04	7068.82
5	7073.57	6617.54	6617.70
6	4967.53	4564.53	4594.44

(b). $\widehat{m.s.e.}(\hat{Y}_{s_0,(.)})$ based on 1,000 M realizations

Table 5.2: The sample means of three predictions for 1,000 simulated Tba realizations

plot		$\widehat{Y}_{s_0,b}$	$\widehat{Y}_{s_0,a}$	$\widehat{Y}_{s_0}$	$\hat{\sigma}^2_{\widehat{Y},p,b}$	$\hat{\sigma}^2_{\widehat{Y},p,a}$	$\hat{\sigma}^2_{\widehat{Y},p}$
1	mean	41.48	43.11	43.12	562.84	541.83	560.23
	s.e.	8.03	8.04	8.10	62.93	60.81	62.64
2	mean	41.78	43.41	43.41	562.00	541.03	559.41
	s.e.	6.43	6.43	6.48	63.93	61.77	63.63
3	mean	42.08	43.71	43.72	561.23	540.29	558.64
	s.e.	6.62	6.60	6.65	66.23	63.98	65.93
4	mean	41.83	43.46	43.47	562.38	541.39	559.78
	s.e.	5.99	6.00	6.05	63.78	61.63	63.49
5	mean	42.04	43.67	43.68	562.70	541.70	560.10
	s.e.	6.51	6.50	6.56	61.81	59.73	61.52
6	mean	41.90	43.53	43.54	562.70	541.69	560.09
	s.e.	6.16	6.16	6.21	61.91	59.83	61.62
plot		$\widetilde{Y}_{s_0,b}$	$\widetilde{Y}_{s_0,a}$	$\widetilde{Y}_{s_0}$	$\hat{\sigma}^2_{\widetilde{Y},p,b}$	$\hat{\sigma}^2_{\widetilde{Y},p,a}$	$\hat{\sigma}^2_{\widetilde{Y},p}$
1	mean	41.54	43.25	43.26	574.26	551.92	571.24
	s.e.	6.24	6.28	6.32	61.21	57.43	60.64
2	mean	41.76	43.47	43.48	573.81	551.49	570.80
	s.e.	5.07	5.09	5.12	61.27	57.49	60.70
3	mean	41.94	43.65	43.66	573.70	551.39	570.69
	s.e.	4.97	4.98	5.01	61.32	57.53	60.74
4	mean	41.74	43.45	43.46	574.10	551.77	571.08
	s.e.	4.52	4.55	4.58	61.23	57.45	60.66
5	mean	41.96	43.67	43.68	573.84	551.51	570.81
	s.e.	5.06	5.07	5.10	61.24	57.45	60.66
6	mean	41.80	43.51	43.52	573.81	551.47	570.78
	s.e.	4.81	4.83	4.86	61.26	57.48	60.68

(a). $\widehat{Y}_{s_0,b}$ in Eqn. (5.25) $\widehat{Y}_{s_0,a}$ in Eqn. (5.32) $\widehat{Y}_{s_0,}$ in Eqn. (5.40)
$\widetilde{Y}_{s_0,b}$ in Eqn. (5.27) $\widetilde{Y}_{s_0,a}$ in Eqn. (5.36) $\widetilde{Y}_{s_0,}$ in Eqn. (5.41)

The rows of mean are calculated by $\frac{1}{1,000}\sum_{b=1}^{1,000}\hat{Y}^b_{s_0,(.)}$ for $\hat{Y}^b_{s_0,(.)}$, and $\frac{1}{1,000}\sum_{b=1}^{1,000}\hat{\sigma}^2_{\hat{Y}^b_{s_0,(.)}}$ for $\hat{\sigma}^2_{\hat{Y}^b_{s_0,(.)}}$, i.e. $\frac{1}{1,000}\sum_{b=1}^{1,000} b^{th} estimate$. The rows of s.e. are calculated by $\frac{1}{1,000}\sum_{b=1}^{1,000}(b^{th}\, estimate - mean)^2$.

(Continued)

plot		$\widehat{\widehat{Y}}_{s_0,b}$	$\widehat{\widehat{Y}}_{s_0,a}$	$\widehat{\widehat{Y}}_{s_0}$	$\hat{\sigma}^2_{\widehat{\widehat{Y}},p,b}$	$\hat{\sigma}^2_{\widehat{\widehat{Y}},p,a}$	$\hat{\sigma}^2_{\widehat{\widehat{Y}},p}$
1	mean	41.49	43.10	43.12	553.76	530.64	550.90
	s.e.	10.36	10.42	10.51	81.89	76.64	81.11
2	mean	41.74	43.35	43.37	552.94	529.85	550.10
	s.e.	8.61	8.64	8.71	82.58	77.32	81.80
3	mean	42.02	43.63	43.65	552.17	529.12	549.34
	s.e.	8.52	8.53	8.60	84.28	79.00	83.51
4	mean	41.76	43.37	43.38	553.30	530.20	550.46
	s.e.	7.78	7.82	7.87	82.50	77.24	81.72
5	mean	41.95	43.57	43.58	553.62	530.50	550.77
	s.e.	8.19	8.20	8.26	81.05	75.80	80.26
6	mean	41.83	43.44	43.46	553.62	530.50	550.76
	s.e.	7.89	7.90	7.96	81.13	75.87	80.33

plot		$\widehat{\widehat{Y}}_{s_0,b}$	$\widehat{\widehat{Y}}_{s_0,a}$	$\widehat{\widehat{Y}}_{s_0}$	$\hat{\sigma}^2_{\widehat{\widehat{Y}},p,b}$	$\hat{\sigma}^2_{\widehat{\widehat{Y}},p,a}$	$\hat{\sigma}^2_{\widehat{\widehat{Y}},p}$
1	mean	41.52	43.48	43.25	570.86	545.47	567.44
	s.e.	8.72	8.86	8.88	114.53	104.38	113.00
2	mean	41.71	43.66	43.43	570.42	545.05	567.01
	s.e.	7.46	7.64	7.57	114.52	104.37	112.99
3	mean	41.88	43.83	43.60	570.31	544.95	566.90
	s.e.	7.28	7.44	7.38	114.53	104.39	113.01
4	mean	41.67	43.62	43.40	570.70	545.33	567.29
	s.e.	6.77	6.94	6.87	114.53	104.38	113.00
5	mean	41.88	43.83	43.60	570.45	545.07	567.02
	s.e.	7.18	7.38	7.26	114.50	104.35	112.97
6	mean	41.72	43.67	43.44	570.42	545.03	566.99
	s.e.	6.96	7.16	7.03	114.50	104.35	112.97

(b). $\widehat{\widehat{Y}}_{s_0,b}$ in Eqn. (5.30) $\widehat{\widehat{Y}}_{s_0,a}$ in Eqn. (5.37) $\widehat{\widehat{Y}}_{s_0,}$ in Eqn. (5.42)
$\widehat{\widehat{Y}}_{s_0,b}$ in Eqn. (5.31) $\widehat{\widehat{Y}}_{s_0,a}$ in Eqn. (5.38) $\widehat{\widehat{Y}}_{s_0,}$ in Eqn. (5.43)

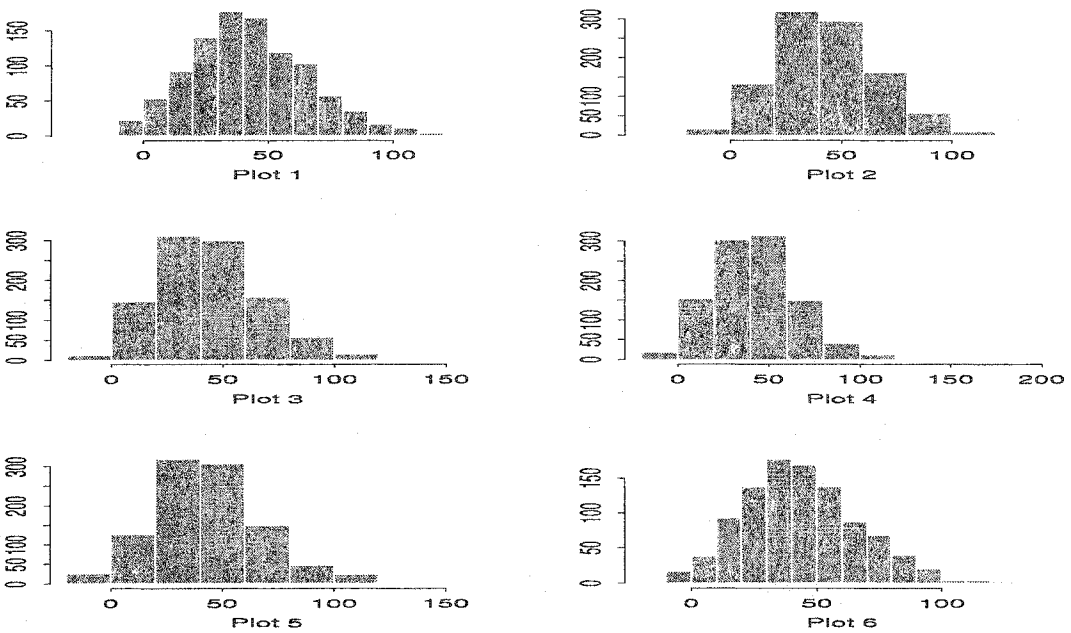
The rows of mean are calculated by $\frac{1}{1,000}\sum_{b=1}^{1,000}\hat{Y}^b_{s_0,(.)}$ for $\hat{Y}^b_{s_0,(.)}$, and $\frac{1}{1,000}\sum_{b=1}^{1,000}\hat{\sigma}^2_{\hat{Y}^b_{s_0,(.)}}$ for $\hat{\sigma}^2_{\hat{Y}^b_{s_0,(.)}}$, i.e. $\frac{1}{1,000}\sum_{b=1}^{1,000} b^{th} estimate$. The rows of s.e. are calculated by $\frac{1}{1,000}\sum_{b=1}^{1,000}(b^{th}\, estimate - mean)^2$.

Table 5.3: The estimated mean squared errors (m.s.e.s) of three predictions for 1,000 simulated Tba realizations

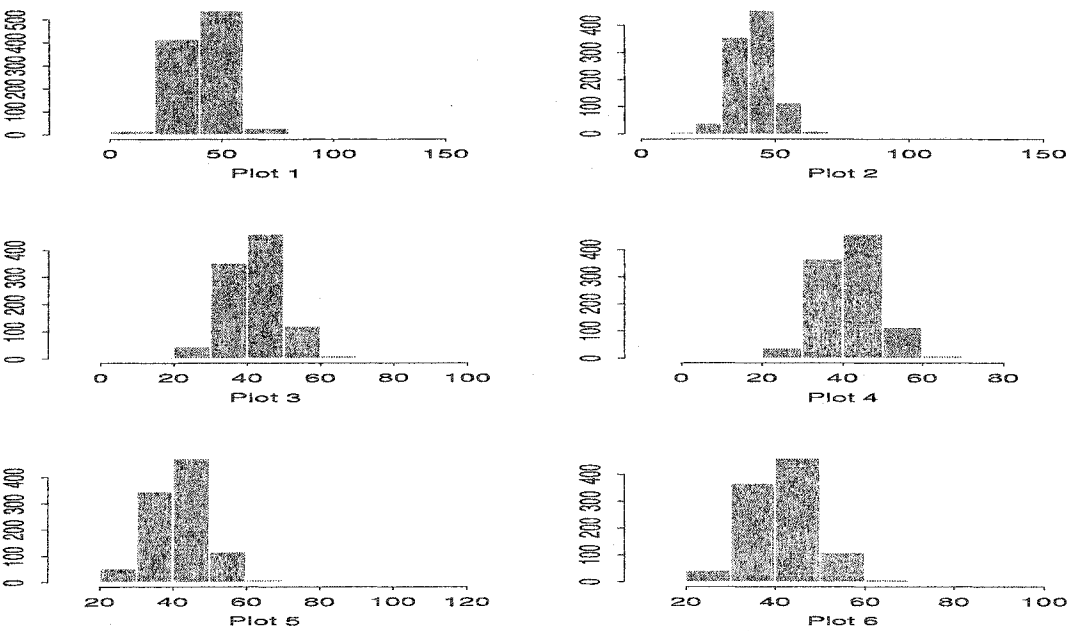
plot	$\widehat{m.s.e.}(\widehat{Y}_{s_0,b})$	$\widehat{m.s.e.}(\widehat{Y}_{s_0,a})$	$\widehat{m.s.e.}(\widehat{Y}_{s_0})$	$\widehat{m.s.e.}(\widetilde{Y}_{s_0,b})$	$\widehat{m.s.e.}(\widetilde{Y}_{s_0,a})$	$\widehat{m.s.e.}(\widetilde{Y}_{s_0})$
1	590.94	590.48	591.25	570.62	570.75	571.11
2	584.84	581.75	582.28	572.19	569.21	569.45
3	588.80	584.94	585.31	571.53	567.20	567.36
4	567.80	569.01	569.45	557.78	559.80	559.98
5	635.11	631.36	631.84	623.82	620.01	620.18
6	577.37	573.25	573.73	564.58	559.76	559.97
plot	$\widehat{m.s.e.}(\widehat{\widehat{Y}}_{s_0,b})$	$\widehat{m.s.e.}(\widehat{\widehat{Y}}_{s_0,a})$	$\widehat{m.s.e.}(\widehat{\widehat{Y}}_{s_0})$	$\widehat{m.s.e.}(\widehat{\widetilde{Y}}_{s_0,b})$	$\widehat{m.s.e.}(\widehat{\widetilde{Y}}_{s_0,a})$	$\widehat{m.s.e.}(\widehat{\widetilde{Y}}_{s_0})$
1	626.72	627.75	629.27	600.76	604.18	603.35
2	612.83	608.73	609.81	597.80	600.28	594.24
3	602.92	599.85	600.43	585.11	580.47	581.91
4	588.56	589.37	590.02	579.19	582.79	581.26
5	658.56	654.40	655.08	648.55	648.26	644.77
6	593.11	588.00	588.61	581.27	580.69	575.87

$\widehat{m.s.e.}(Y_p^i) = \frac{1}{1,000}\sum_{b=1}^{1,000}(Y_{p,N+i}^b - Y_{g,N+i}^b)^2$ where $Y_{p,N+i}^b$, $i = 1,\ldots,6$ denote the prediction at non-sampled plot i using those three prediction rules based on the generated values at sampled plots, $1,\ldots,N = 312$ from the b^{th} generated realization of size $(N+6)$, and $Y_{g,N+i}^b$ denotes the b^{th} generated values at non-sampled plot i.

Figure 5.5: 1,000 simulated Tba values at 6 non-sampled plots and 1,000 Tba predicitons at 6 non-sampled plots using prediction rules based on $N = 312$ generated simulated Tba values at sampled plots. [1]



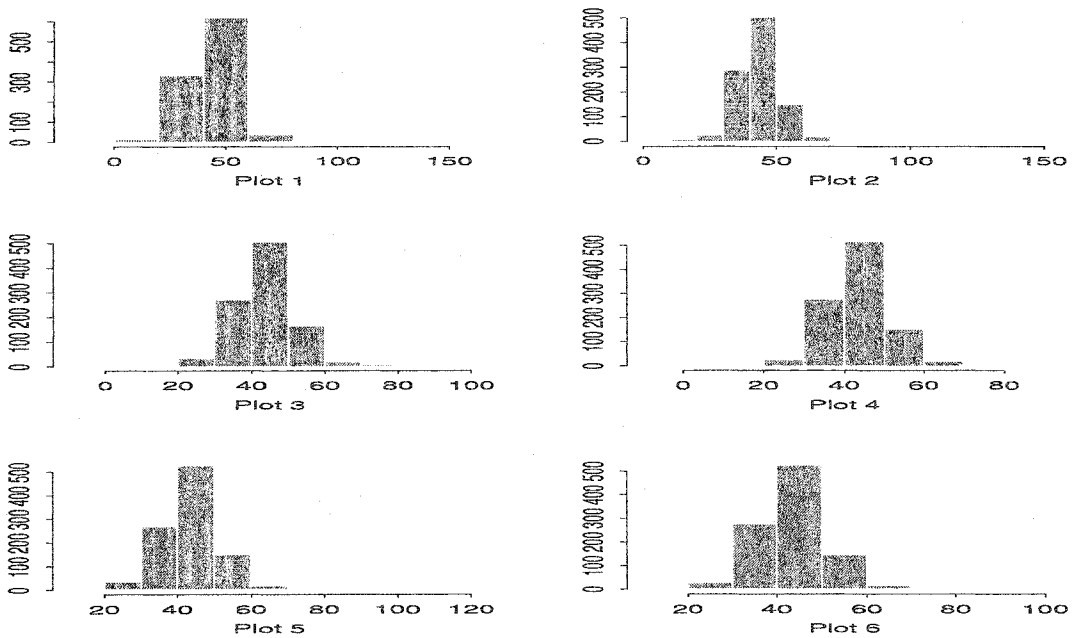
(a). generated Y values: $Y_{g,N+i}$, $i = 1, \ldots, 6$



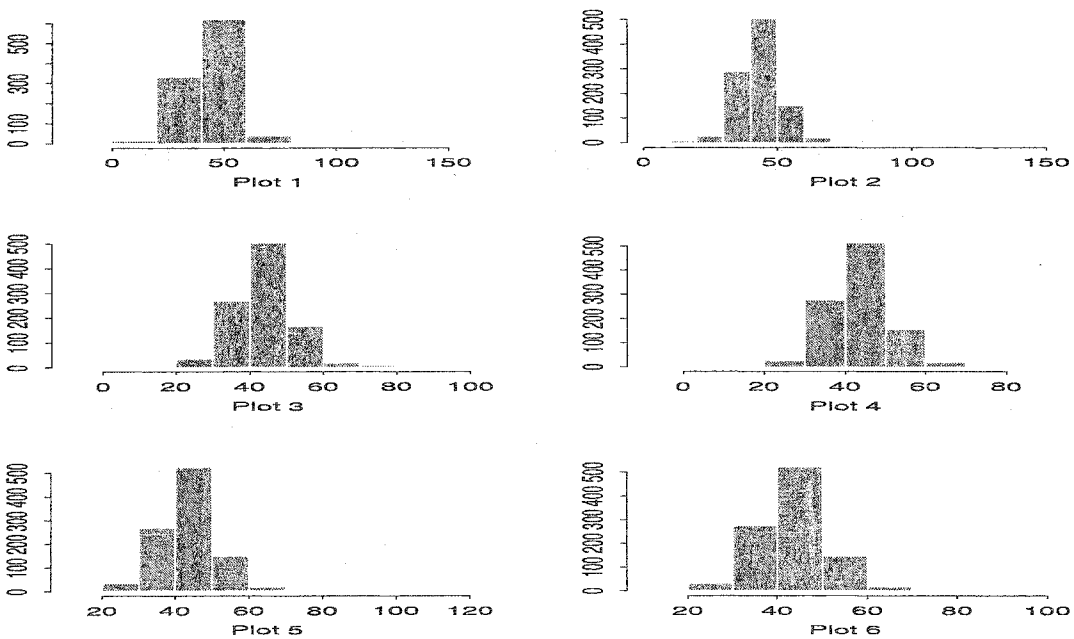
(b). predictions: $\hat{\tilde{Y}}_{s_0,b}$

[1] Six non-sampled plots are depicted in Figure 5.4

(Continued)



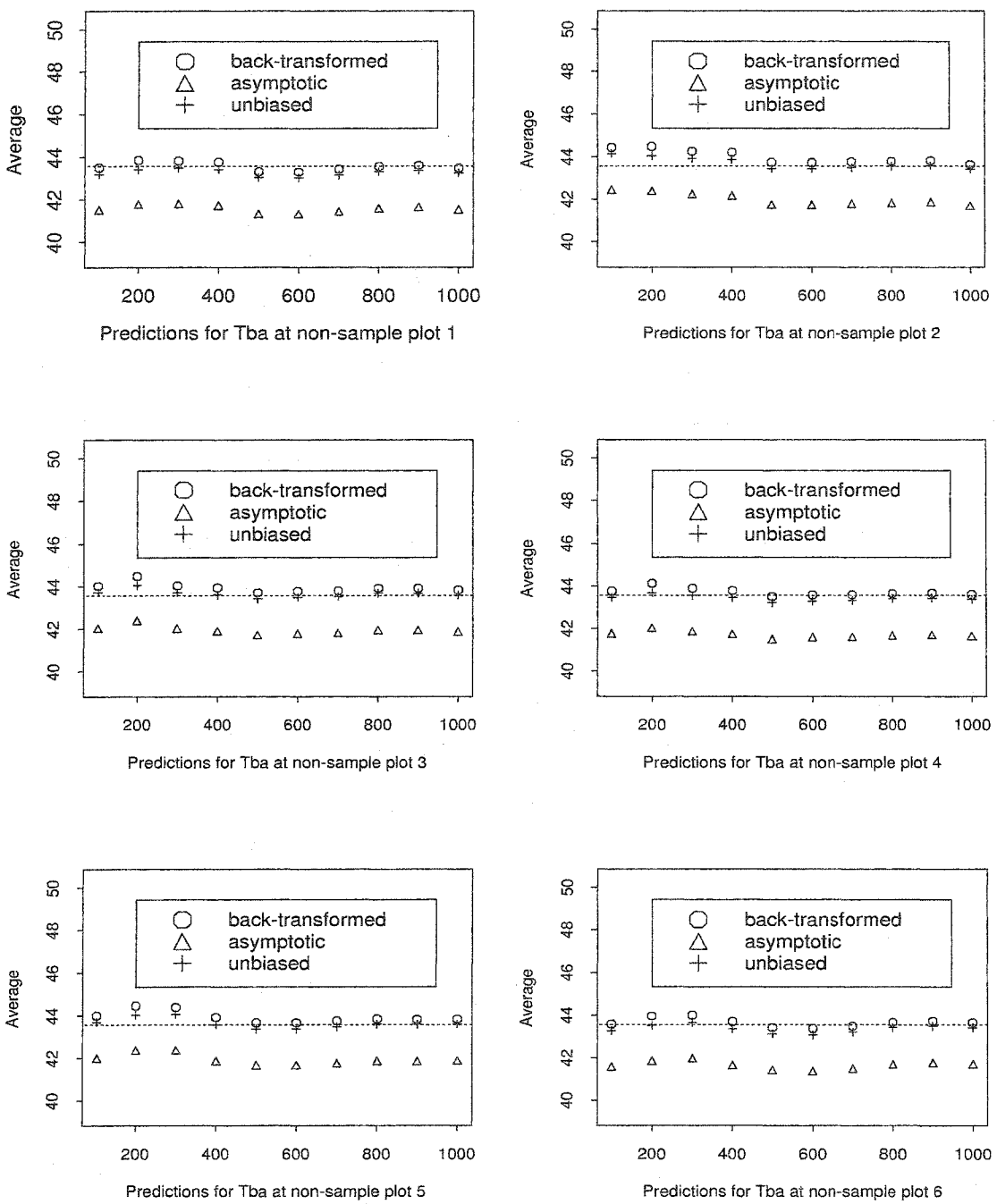
(c). predictions: $\hat{\hat{Y}}_{s_0,a}$



(d). predictions: $\hat{\hat{Y}}_{s_0,}$

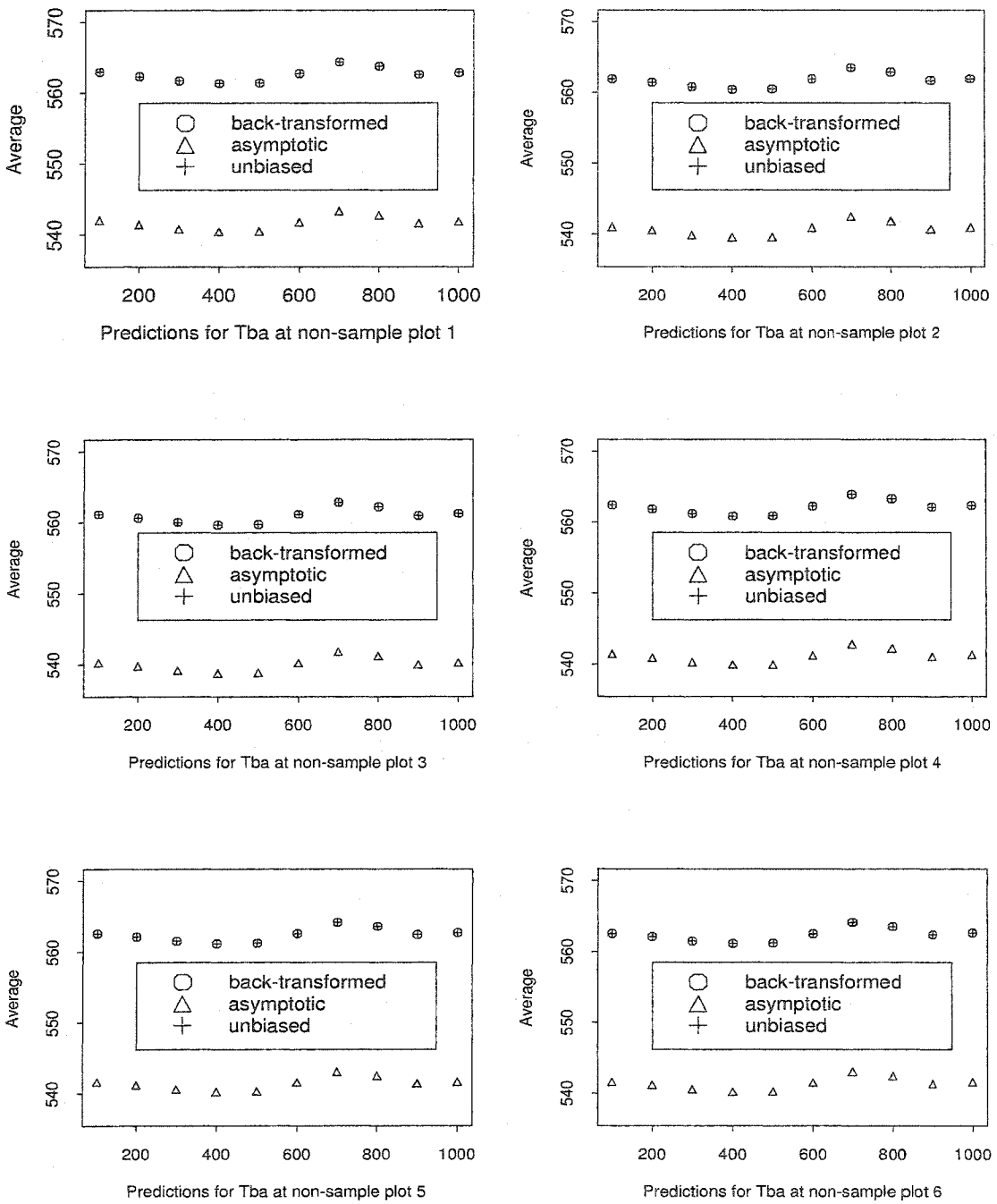
[1] Six non-sampled plots are depicted in Figure 5.4

Figure 5.6: The averages of predictions of Tba at 6 non-sampled plots based on b realizations [1]



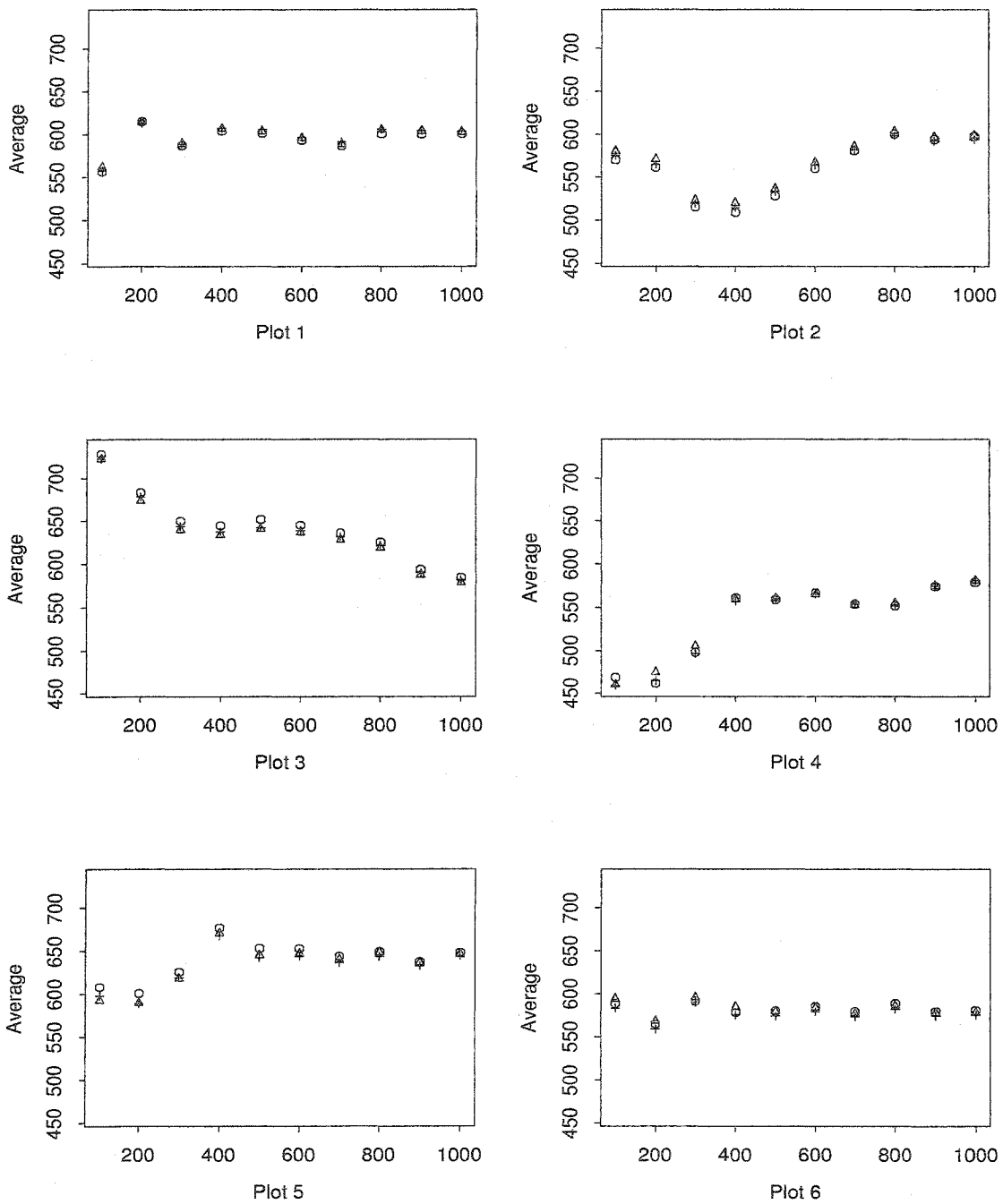
[1] b=100, 200, ..., 1000.

Figure 5.7: The averages of m.s. prediction errors of Tba at 6 non-sampled plots based on b realizations [1]



[1] b=100, 200, ..., 1000.

Figure 5.8: The estimated m.s. prediction errors of Tba at 6 non-sampled plots based on b realizations [1]



[1] b=100, 200, ..., 1000.

Table 5.4: The sample means of three predictions for 1,000 simulated Lt realizations

plot		$\widehat{Y}_{s_0,b}$	$\widehat{Y}_{s_0,a}$	$\widehat{Y}_{s_0}$	$\hat{\sigma}^2_{\widehat{Y},p,b}$	$\hat{\sigma}^2_{\widehat{Y},p,a}$	$\hat{\sigma}^2_{\widehat{Y},p}$
1	mean	326.65	386.14	393.43	78029.13	44950.54	73047.77
	s.e.	269.69	268.76	294.67	16179.68	9182.64	15797.99
2	mean	313.99	373.09	378.59	77478.82	44622.34	72515.31
	s.e.	138.36	137.16	149.43	17428.37	9840.14	17212.86
3	mean	313.00	371.78	377.25	77201.84	44486.38	72275.49
	s.e.	85.35	84.24	93.30	18205.02	10243.84	18151.96
4	mean	313.20	372.44	377.86	77773.47	44808.08	72821.06
	s.e.	84.25	82.33	87.73	16629.55	9422.52	16299.57
5	mean	317.48	376.74	382.81	77507.75	44598.28	72510.77
	s.e.	107.09	105.53	116.33	16910.40	9592.73	16582.98
6	mean	313.58	372.98	378.69	77644.60	44676.59	72622.34
	s.e.	91.19	90.24	100.14	16939.75	9590.60	16632.10
plot		$\widetilde{Y}_{s_0,b}$	$\widetilde{Y}_{s_0,a}$	$\widetilde{Y}_{s_0}$	$\hat{\sigma}^2_{\widetilde{Y},p,b}$	$\hat{\sigma}^2_{\widetilde{Y},p,a}$	$\hat{\sigma}^2_{\widetilde{Y},p}$
1	mean	308.98	369.25	374.65	84856.95	49228.27	80325.58
	s.e.	89.16	95.06	104.46	23578.76	12173.23	22124.56
2	mean	307.04	367.27	372.32	84805.06	49197.40	80282.35
	s.e.	55.23	61.03	64.80	23560.53	12163.67	22110.38
3	mean	306.75	366.93	371.91	84783.57	49189.29	80270.90
	s.e.	47.29	52.63	55.09	23554.82	12162.34	22108.09
4	mean	307.27	367.51	372.56	84833.37	49217.39	80310.44
	s.e.	45.42	50.08	52.64	23570.77	12169.95	22120.14
5	mean	308.39	368.66	373.87	84817.83	49198.62	80284.19
	s.e.	47.49	51.82	54.82	23563.90	12162.99	22109.83
6	mean	306.03	366.32	371.23	84820.89	49196.06	80280.33
	s.e.	41.34	46.52	48.69	23563.88	12161.62	22106.55

(a). $\widehat{Y}_{s_0,b}$ in Eqn. (5.25) $\widehat{Y}_{s_0,a}$ in Eqn. (5.32) $\widehat{Y}_{s_0}$ in Eqn. (5.40)
$\widetilde{Y}_{s_0,b}$ in Eqn. (5.27) $\widetilde{Y}_{s_0,a}$ in Eqn. (5.36) $\widetilde{Y}_{s_0}$ in Eqn. (5.41)

The rows of mean are calculated by $\frac{1}{1,000}\sum_{b=1}^{1,000}\hat{Y}^b_{s_0,(.)}$ for $\hat{Y}^b_{s_0,(.)}$, and $\frac{1}{1,000}\sum_{b=1}^{1,000}\hat{\sigma}^2_{\hat{Y}^b_{s_0,(.)}}$ for $\hat{\sigma}^2_{\hat{Y}^b_{s_0,(.)}}$, i.e. $\frac{1}{1,000}\sum_{b=1}^{1,000} b^{th} estimate$. The rows of s.e. are calculated by $\frac{1}{1,000}\sum_{b=1}^{1,000}(b^{th}\, estimate - mean)^2$.

(Continued)

plot		$\widehat{\widehat{Y}}_{s_0,b}$	$\widehat{\widehat{Y}}_{s_0,a}$	$\widehat{\widehat{Y}}_{s_0}$	$\hat{\sigma}^2_{\widehat{\widehat{Y}},p,b}$	$\hat{\sigma}^2_{\widehat{\widehat{Y}},p,a}$	$\hat{\sigma}^2_{\widehat{\widehat{Y}},p}$
1	mean	359.48	417.58	432.69	76971.53	43889.41	72058.78
	s.e.	776.63	778.60	910.00	27094.87	13015.20	25296.49
2	mean	326.94	384.65	393.08	76411.08	43566.81	71526.18
	s.e.	333.86	336.09	382.99	27384.14	13321.48	25676.91
3	mean	317.37	374.77	381.40	76132.88	43435.94	71294.61
	s.e.	120.35	124.91	138.04	27658.64	13544.89	26041.26
4	mean	315.48	373.34	379.50	76707.87	43748.43	71829.08
	s.e.	102.62	106.57	114.94	27139.61	13112.97	25387.70
5	mean	316.75	374.61	380.52	76427.11	43536.88	71503.01
	s.e.	105.45	107.10	116.84	27073.93	13142.12	25307.60
6	mean	312.99	371.01	376.60	76575.95	43616.78	71624.78
	s.e.	91.88	94.74	104.01	27209.12	13169.34	25427.57

plot		$\widehat{\widehat{Y}}_{s_0,b}$	$\widehat{\widehat{Y}}_{s_0,a}$	$\widehat{\widehat{Y}}_{s_0}$	$\hat{\sigma}^2_{\widehat{\widehat{Y}},p,b}$	$\hat{\sigma}^2_{\widehat{\widehat{Y}},p,a}$	$\hat{\sigma}^2_{\widehat{\widehat{Y}},p}$
1	mean	317.08	377.78	385.07	92798.91	51250.31	87602.69
	s.e.	141.43	152.28	180.87	79279.73	29398.90	73370.07
2	mean	311.49	372.14	377.66	92738.83	51217.06	87553.97
	s.e.	97.03	108.47	122.89	79189.60	29366.22	73308.31
3	mean	310.03	370.64	375.55	92714.22	51208.45	87541.16
	s.e.	82.49	94.05	100.83	79154.84	29358.38	73290.41
4	mean	309.90	370.56	375.29	92771.72	51238.73	87586.01
	s.e.	77.55	88.85	94.02	79242.60	29388.73	73353.46
5	mean	310.92	371.61	376.45	92753.92	51218.29	87556.04
	s.e.	78.78	89.86	95.16	79214.79	29367.48	73313.42
6	mean	308.43	369.14	373.59	92756.85	51215.02	87550.38
	s.e.	74.09	85.85	90.23	79207.74	29359.33	73291.06

(b). $\widehat{\widehat{Y}}_{s_0,b}$ in Eqn. (5.30) $\widehat{\widehat{Y}}_{s_0,a}$ in Eqn. (5.37) $\widehat{\widehat{Y}}_{s_0,}$ in Eqn. (5.42)
$\widehat{\widehat{Y}}_{s_0,b}$ in Eqn. (5.31) $\widehat{\widehat{Y}}_{s_0,a}$ in Eqn. (5.38) $\widehat{\widehat{Y}}_{s_0,}$ in Eqn. (5.43)

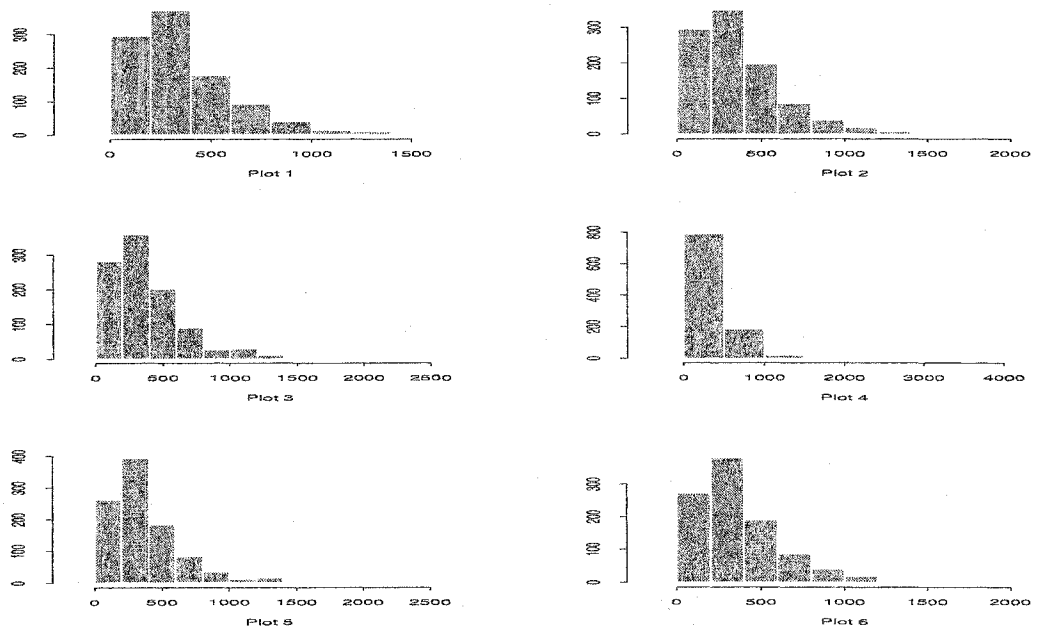
The rows of mean are calculated by $\frac{1}{1,000}\sum_{b=1}^{1,000}\hat{Y}^b_{s_0,(.)}$ for $\hat{Y}^b_{s_0,(.)}$, and $\frac{1}{1,000}\sum_{b=1}^{1,000}\hat{\sigma}^2_{\hat{Y}^b_{s_0,(.)}}$ for $\hat{\sigma}^2_{\hat{Y}^b_{s_0,(.)}}$, i.e. $\frac{1}{1,000}\sum_{b=1}^{1,000} b^{th} estimate$. The rows of s.e. are calculated by $\frac{1}{1,000}\sum_{b=1}^{1,000}(b^{th}\, estimate - mean)^2$.

Table 5.5: The estimated mean squared errors (m.s.e.s) of three predictions for 1,000 simulated Lt realizations

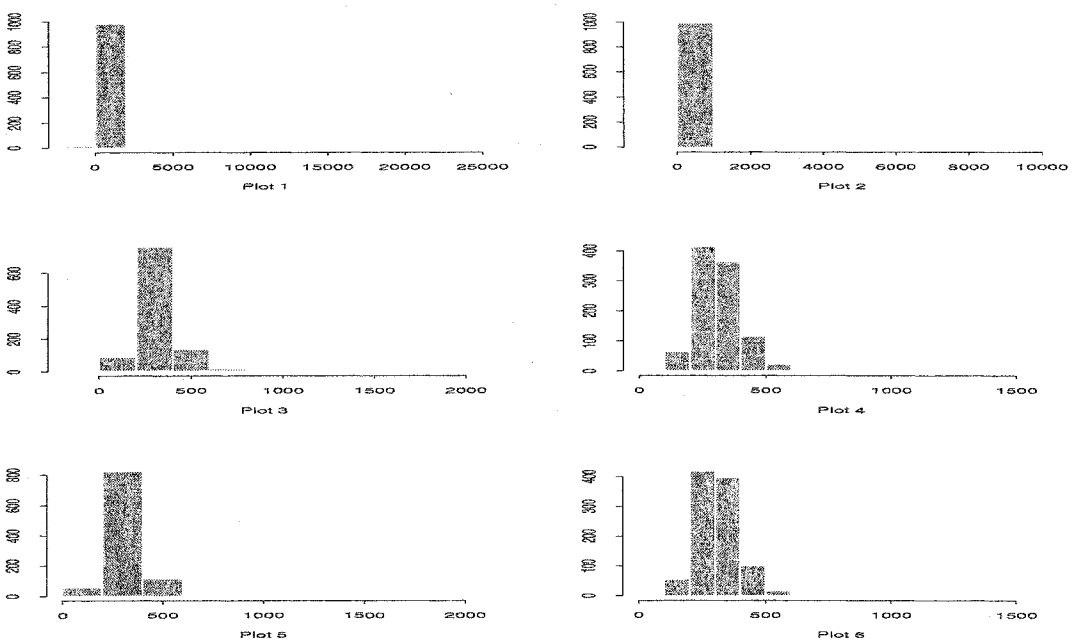
plot	$\widehat{m.s.e.}$ $(\widehat{Y}_{s_0,b})$	$\widehat{m.s.e.}$ $(\widehat{Y}_{s_0,a})$	$\widehat{m.s.e.}$ $(\widehat{Y}_{s_0})$	$\widehat{m.s.e.}$ $(\widetilde{Y}_{s_0,b})$	$\widehat{m.s.e.}$ $(\widetilde{Y}_{s_0,a})$	$\widehat{m.s.e.}$ $(\widetilde{Y}_{s_0})$
1	136516.64	136104.61	150453.49	74334.88	73072.45	74847.27
2	91086.77	87691.69	91205.56	76642.17	73484.53	73960.59
3	87409.71	83643.42	85024.20	83083.56	78817.41	78976.41
4	88016.31	85951.49	86937.81	83255.54	81628.07	82054.64
5	102449.84	98527.68	100654.85	95684.83	91380.36	91558.32
6	85668.90	82139.32	84269.69	77983.36	73731.84	74021.62
plot	$\widehat{m.s.e.}$ $(\widehat{\widehat{Y}}_{s_0,b})$	$\widehat{m.s.e.}$ $(\widehat{\widehat{Y}}_{s_0,a})$	$\widehat{m.s.e.}$ $(\widehat{\widehat{Y}}_{s_0})$	$\widehat{m.s.e.}$ $(\widetilde{\widetilde{Y}}_{s_0,b})$	$\widehat{m.s.e.}$ $(\widetilde{\widetilde{Y}}_{s_0,a})$	$\widehat{m.s.e.}$ $(\widetilde{\widetilde{Y}}_{s_0})$
1.00	664433.89	671152.29	894327.18	84225.36	85841.56	95514.36
2.00	178954.31	178707.85	212130.77	81551.49	80655.10	83431.71
3.00	92412.14	90147.98	93269.51	85489.73	82860.87	83879.39
4.00	90786.52	90025.33	91996.35	86366.76	86493.45	87555.78
5.00	101986.42	98482.36	100424.78	98846.57	96224.07	96880.58
6.00	84557.50	81262.28	83202.99	80058.34	77376.78	77728.57

$\widehat{m.s.e.}(Y_p^i) = \frac{1}{1,000}\sum_{b=1}^{1,000}(Y_{p,N+i}^b - Y_{g,N+i}^b)^2$ where $Y_{p,N+i}^b$, $i = 1,\ldots,6$ denote the prediction at non-sampled plot i using those three prediction rules based on the generated values at sampled plots, $1,\ldots,N = 312$ from the b^{th} generated realization of size $(N+6)$, and $Y_{g,N+i}^b$ denotes the b^{th} generated values at non-sampled plot i.

Figure 5.9: 1,000 simulated Lt values at 6 non-sampled plots and 1,000 Lt predicitons at 6 non-sampled plots using prediction rules based on $N = 312$ generated simulated Lt values at sampled plots. [1]



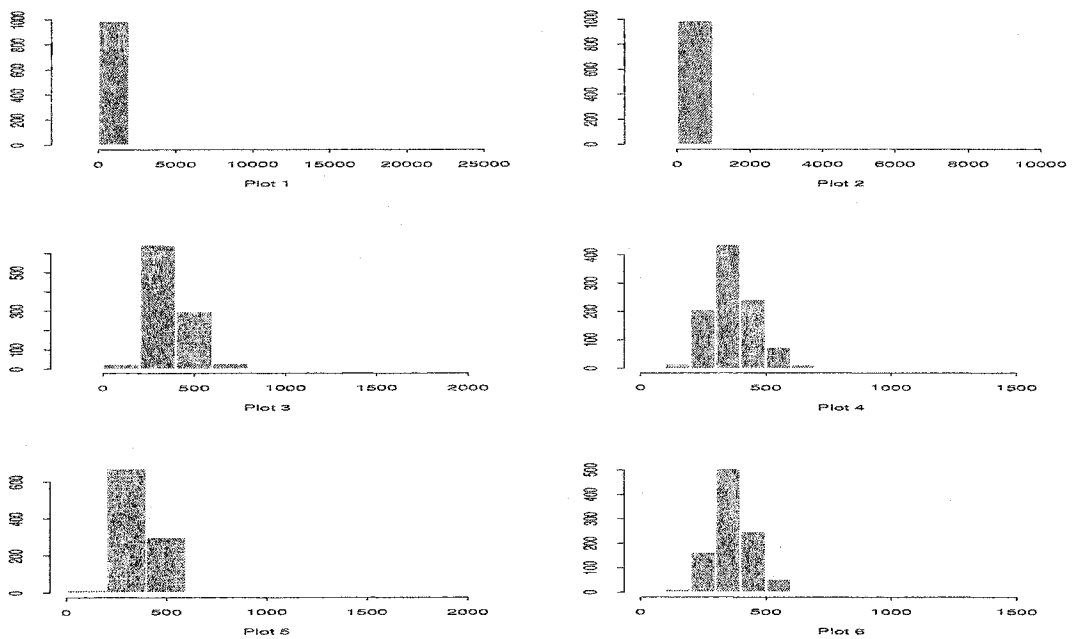
(a). generated Y values: $Y_{g,N+i}$, $i = 1, \ldots, 6$



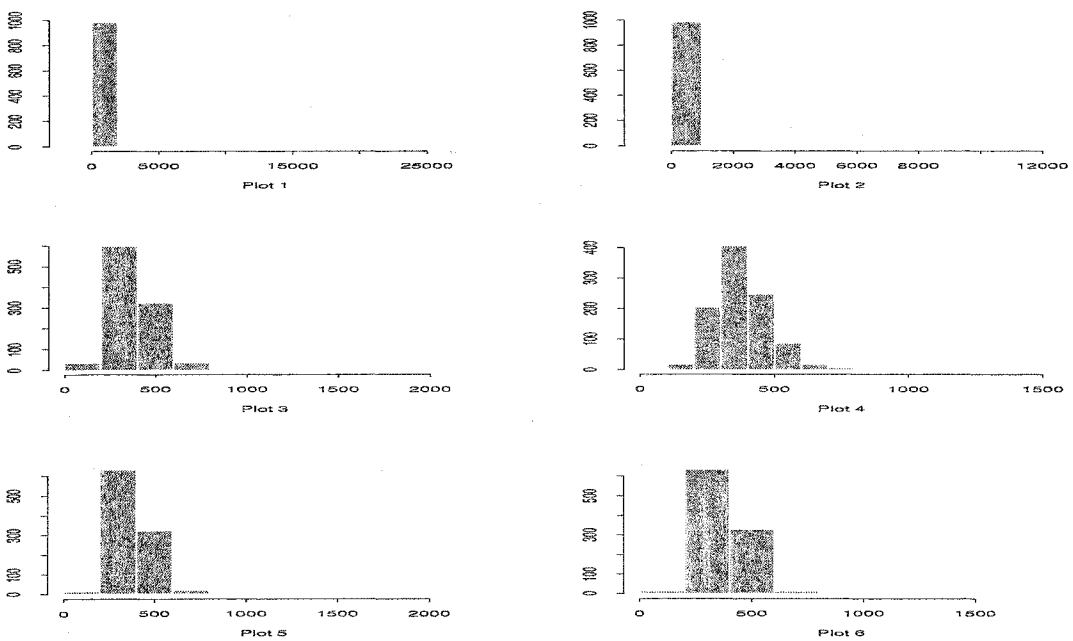
(b). predictions: $\widehat{\widehat{Y}}_{s_0,b}$

[1] Six non-sampled plots are depicted in Figure 5.4

(Continued)



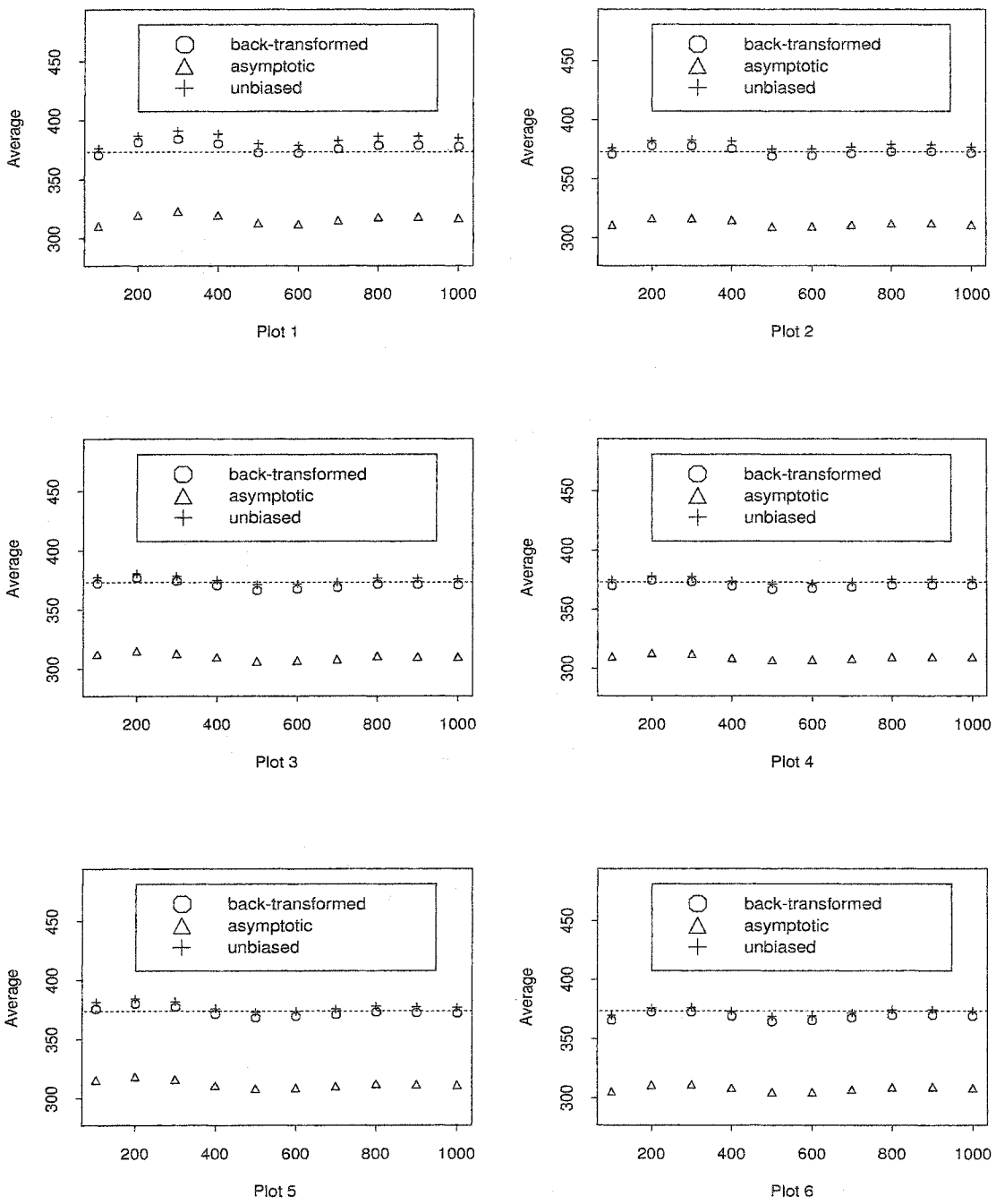
(c). predictions: $\hat{\tilde{Y}}_{s_0,a}$



(d). predictions: $\hat{\tilde{Y}}_{s_0,}$

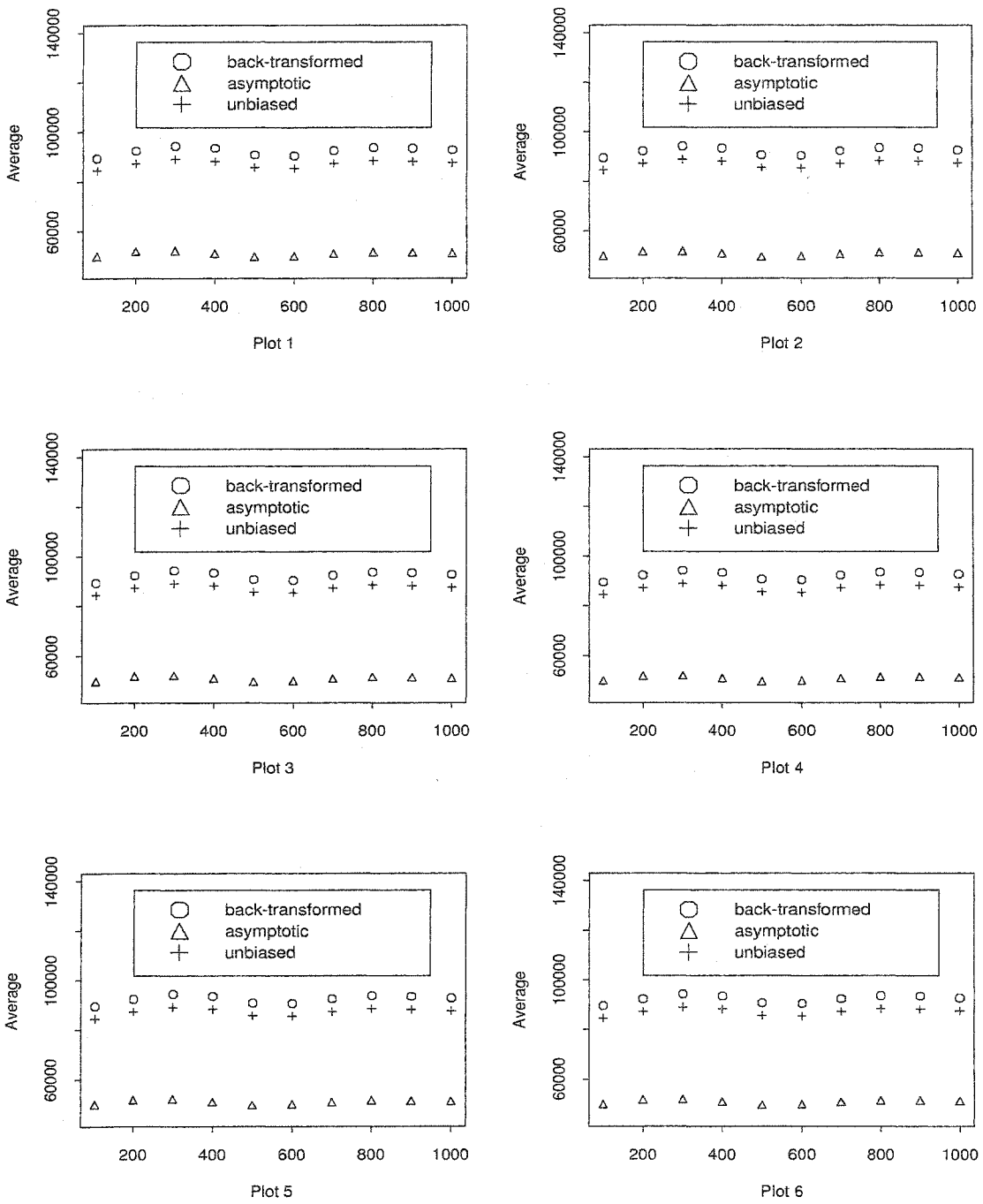
[1] Six non-sampled plots are depicted in Figure 5.4

Figure 5.10: The averages of predictions of Lt at 6 non-sampled plots based on b realizations [1]



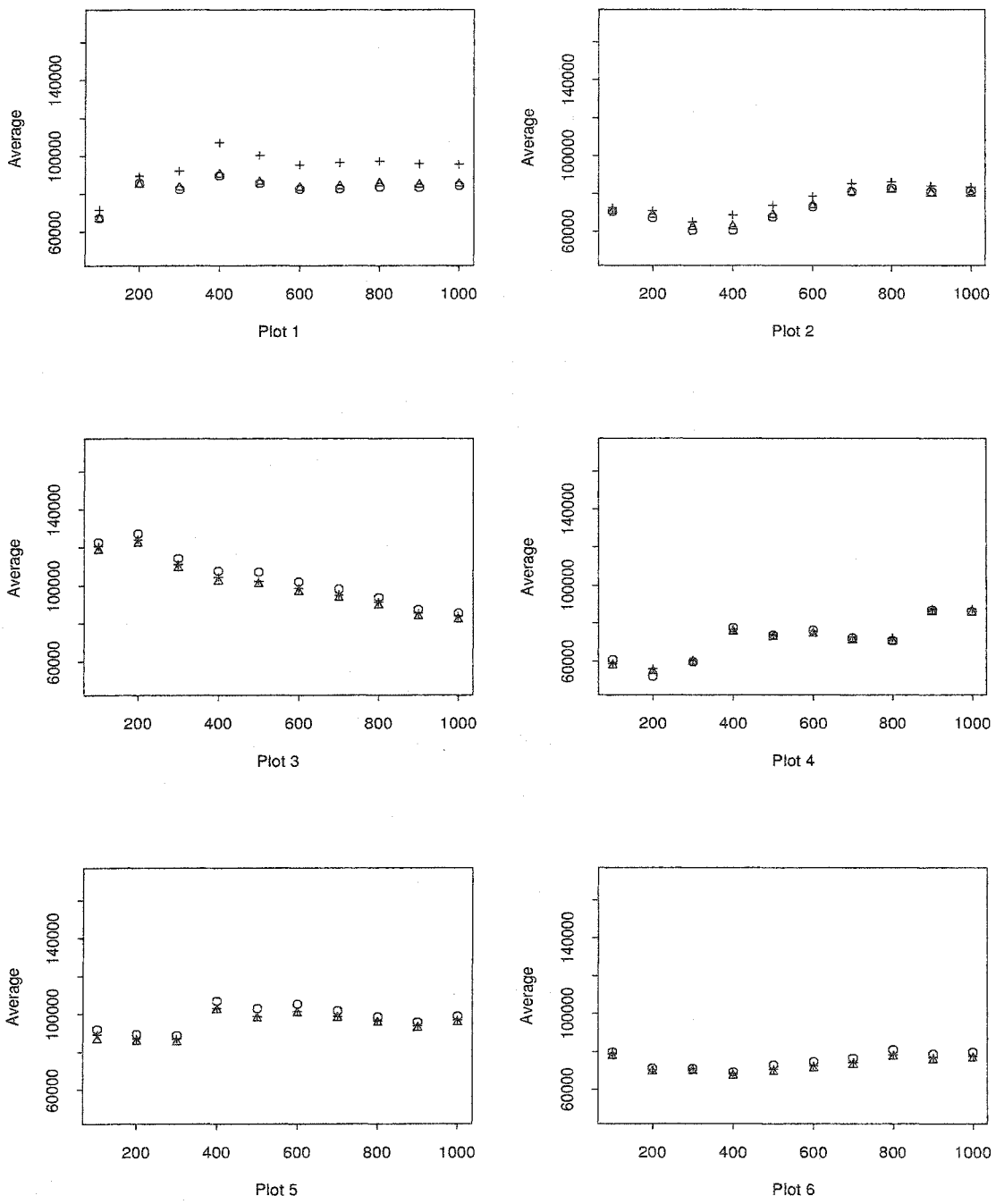
[1] b=100, 200, ..., 1000.

Figure 5.11: The averages of m.s. predictions of Lt at 6 non-sampled plots based on b realizations [1]



[1] b=100, 200, ..., 1000.

Figure 5.12: The estimated m.s. predictions of Lt at 6 non-sampled plots based on b realizations [1]



[1] b=100, 200, . . . , 1000.

Table 5.6: The M predictions at the selected 6 non-sampled plot locations from bootstrapped samples

				Average of 1,000 Bootstrapped Estimates					
	Plug-in Estimates[1]			Simple Bootstrap			Semiparametric Bootstrap		
plot	$\hat{Y}_{s_0,b}$	$\hat{Y}_{s_0,a}$	$\hat{Y}_{s_0}$	$\overline{\hat{Y}}_{s_0,b}$	$\overline{\hat{Y}}_{s_0,a}$	$\overline{\hat{Y}}_{s_0}$	$\overline{\hat{Y}}_{s_0,b}$	$\overline{\hat{Y}}_{s_0,a}$	$\overline{\hat{Y}}_{s_0}$
1	29.27	46.57	45.08	36.95	54.78	56.79	34.16	51.38	52.22
				(8.40)	(9.76)	(12.96)	(7.85)	(8.99)	(12.10)
2	37.87	55.14	57.33	39.63	57.45	60.48	38.83	56.03	58.74
				(3.89)	(5.41)	(7.10)	(3.26)	(4.62)	(6.03)
3	46.82	64.08	70.10	43.90	61.71	66.29	45.03	62.21	67.34
				(4.84)	(3.87)	(4.96)	(4.04)	(3.46)	(4.43)
4	31.87	49.16	48.78	37.93	55.76	58.12	35.64	52.85	54.26
				(6.23)	(7.61)	(10.13)	(5.84)	(6.99)	(9.41)
5	56.04	73.33	83.33	47.62	65.45	71.58	51.84	69.05	77.05
				(8.68)	(7.79)	(10.54)	(8.11)	(7.52)	(10.34)
6	57.13	74.41	84.88	49.34	67.18	73.80	53.12	70.35	78.72
				(14.83)	(13.63)	(17.63)	(12.24)	(11.33)	(14.89)
	Plug-in Estimates[1]			Simple Bootstrap			Semiparametric Bootstrap		
plot	$\hat{\sigma}^2_{Y,p,b}$	$\hat{\sigma}^2_{Y,p,a}$	$\hat{\sigma}^2_{Y,p}$	$\overline{\hat{\sigma}}^2_{Y,p,b}$	$\overline{\hat{\sigma}}^2_{Y,p,a}$	$\overline{\hat{\sigma}}^2_{Y,p}$	$\overline{\hat{\sigma}}^2_{Y,p,b}$	$\overline{\hat{\sigma}}^2_{Y,p,a}$	$\overline{\hat{\sigma}}^2_{Y,p}$
1	5791.1	1665.5	5323.3	5995.0	1659.2	5511.0	6021.4	1665.8	5535.7
				(1426.8)	(283.2)	(1302.5)	(1307.1)	(273.2)	(1194.4)
2	5786.5	1664.1	5320.2	5990.2	1657.1	5507.4	6016.4	1664.1	5531.9
				(1426.3)	(283.0)	(1302.1)	(1306.3)	(273.0)	(1193.7)
3	5784.0	1663.5	5318.8	5987.8	1657.1	5506.3	6014.0	1663.6	5530.8
				(1426.1)	(283.0)	(1302.1)	(1306.0)	(273.0)	(1193.6)
4	5789.4	1665.2	5322.6	5992.6	1658.6	5509.8	6019.0	1665.2	5534.4
				(1426.4)	(283.1)	(1302.4)	(1306.6)	(273.1)	(1194.1)
5	5788.6	1664.6	5321.3	5991.9	1657.6	5507.7	6018.2	1664.2	5532.3
				(1426.5)	(282.9)	(1302.1)	(1306.6)	(273.0)	(1193.7)
6	5788.0	1664.2	5320.3	(5992.4)	1657.4	5507.2	6018.7	1664.0	5531.7
				(1426.8)	(283.0)	(1302.2)	(1306.8)	(273.1)	(1193.8)

[1] The plug-in estimates for $\hat{\sigma}^2_{Y,p,a}$ and $\hat{\sigma}^2_{Y,p}$ are calculated by plugging in the N observations and the estimated a_i based on the N observatoins. Also, the sample standard deviations of $\hat{Y}_{s_0,a}$, $\hat{Y}_{s_0}$, $\hat{\sigma}^2_{Y,p,a}$ and $\hat{\sigma}^2_{Y,p}$ are listed in parentheses.

Table 5.7: The Plug-in estimates for M at the selected 6 non-sampled plot locations

	Plug-in Estimates[1]				
plot	$\widetilde{Y}_{s_0}$	$\widetilde{Y}_{s_0,I}$	$\widetilde{Y}_{s_0,b}$	$\widetilde{Y}_{s_0,a}$	$\widetilde{Y}_{s_0,u}$
1	14.16	54.71	29.15	46.66	44.87
2	43.73	60.01	37.79	55.26	57.17
3	70.20	62.24	46.87	64.33	70.12
4	34.58	53.23	31.81	49.30	48.65
5	79.14	73.80	56.08	73.57	83.32
6	136.13	77.26	57.26	74.75	85.01
	Plug-in Estimates[1]				
plot	$\hat{\sigma}^2_{Y,p}$	$\hat{\sigma}^2_{Y,p,I}$	$\hat{\sigma}^2_{Y,p,b}$	$\hat{\sigma}^2_{Y,p,a}$	$\hat{\sigma}^2_{Y,p,u}$
1	2783.20	4355.37	5930.36	1708.26	5451.60
2	2781.68	4352.69	5925.70	1706.83	5448.46
3	2778.71	4351.56	5923.16	1706.20	5447.08
4	2783.03	4354.70	5928.64	1707.95	5450.92
5	2782.49	4353.52	5927.74	1707.30	5449.50
6	2779.34	4352.94	5927.17	1706.83	5448.47

[1] The plug-in estimates for $\widetilde{Y}_{s_0}$, $\widetilde{Y}_{s_0,I}$, $\widetilde{Y}_{s_0,b}$, $\widetilde{Y}_{s_0,a}$, $\widetilde{Y}_{s_0,u}$, $\hat{\sigma}^2_{Y,p}$, $\hat{\sigma}^2_{Y,p,I}$, $\hat{\sigma}^2_{Y,p,b}$, $\hat{\sigma}^2_{Y,p,a}$, and $\hat{\sigma}^2_{Y,p,u}$ are calculated by plugging the N observations and the estimated a_i based on the N observatoins into Eqns. (5.20), (5.22), (5.31), (5.38), (5.43), (5.21), (5.23), (5.29), (5.39), and (5.45), repsectively.

Table 5.8: The averages of M predictions at the selected 6 non-sampled plot locations from bootstrapped samples

	1,000 bootstrapped samples				
Plot	$\widetilde{\overline{Y}}_{s_0}$	$\widetilde{\overline{Y}}_{s_0,I}$	$\widetilde{\overline{Y}}_{s_0,b}$	$\widetilde{\overline{Y}}_{s_0,a}$	$\widetilde{\overline{Y}}_{s_0,u}$
	Simple Bootstrap Method				
1	62.84	62.41	41.61	59.72	62.75
	(9.74)	(7.32)	(6.69)	(10.59)	(8.71)
2	62.30	62.36	41.32	59.42	62.31
	(10.57)	(8.12)	(7.15)	(10.99)	(9.41)
3	62.66	62.92	41.77	59.86	62.87
	(9.62)	(8.86)	(7.05)	(10.55)	(8.41)
4	62.90	62.59	41.70	59.80	62.86
	(9.84)	(7.75)	(6.69)	(10.53)	(8.57)
5	62.38	62.66	41.72	59.83	62.82
	(10.08)	(7.19)	(6.44)	(10.67)	(8.06)
6	62.92	62.67	41.77	59.89	62.92
	(11.08)	(7.95)	(6.73)	(10.81)	(8.62)
	Semiparametric Bootstrap Method				
1	62.03	61.34	50.03	59.76	60.59
	(27.51)	(31.54)	(28.17)	(28.16)	(31.03)
2	62.96	63.34	52.07	61.79	62.77
	(27.20)	(31.05)	(28.28)	(28.12)	(30.98)
3	63.50	63.59	52.42	62.12	63.22
	(25.74)	(30.78)	(28.16)	(28.16)	(30.93)
4	64.86	65.46	53.57	63.30	64.46
	(30.28)	(35.29)	(31.81)	(31.67)	(34.54)
5	62.71	63.17	51.53	61.25	62.28
	(26.88)	(32.74)	(29.60)	(29.59)	(32.65)
6	63.16	63.35	51.68	61.40	62.44
	(27.17)	(31.88)	(28.21)	(28.24)	(31.20)

[1] The bootstrapped estimates for $\widetilde{Y}_{s_0}$, $\widetilde{Y}_{s_0,I}$, $\widetilde{Y}_{s_0,b}$, $\widetilde{Y}_{s_0,a}$, $\widetilde{Y}_{s_0,u}$, $\hat{\sigma}^2_{Y,p}$, $\hat{\sigma}^2_{Y,p,I}$, $\hat{\sigma}^2_{Y,p,b}$, $\hat{\sigma}^2_{Y,p,a}$, and $\hat{\sigma}^2_{Y,p,u}$ are calculated by plugging the bootstrapped samples and the estimated a_i based on the bootstrapped samples into Eqns. (5.20), (5.22), (5.31), (5.38), (5.43), (5.21), (5.23), (5.29), (5.39), and (5.45), repsectively. Also, the averages of the bootstrapped estimates are listed and their sample standard deviations are listed in parentheses.

Table 5.9: The normal and bootstrap-t prediction interval (PI) with level $\alpha = 0.05$ estimates for M at the selected 6 non-sampled plot locations

Plot	Y_{s_0}	$Y_{s_0,I}$	$Y_{s_0,b}$	$Y_{s_0,a}$	$Y_{s_0,u}$
	Normal Interval Estimates				
1	(-89.25, 117.56)	(-74.64, 184.06)	(-121.78, 180.09)	(-34.35, 127.66)	(-99.85, 189.58)
2	(-59.64, 147.11)	(-69.30, 189.32)	(-113.09, 188.67)	(-25.71, 136.24)	(-87.50, 201.85)
3	(-33.11, 173.52)	(-67.05, 191.53)	(-103.97, 197.72)	(-16.63, 145.29)	(-74.54, 214.77)
4	(-68.82, 137.98)	(-76.11, 182.57)	(-119.10, 182.73)	(-31.70, 130.30)	(-96.05, 193.36)
5	(-24.24, 182.53)	(-55.52, 203.13)	(-94.82, 206.99)	(-7.42, 154.56)	(-61.37, 228.01)
6	(32.80, 239.46)	(-52.05, 206.58)	(-93.63, 208.16)	(-6.22, 155.73)	(-59.67, 229.68)
	Bootstrap-t Interval Estimates				
Plot	Simple Bootstrap Method:1,000 bootstrapped samples				
1	(41.98, 72.21)	(37.22, 64.04)	(-0.46, 28.64)	(38.88, 55.28)	(30.69, 69.04)
2	(71.22, 99.34)	(88.01, 118.12)	(36.02, 67.14)	(77.11, 94.03)	(76.88, 120.09)
3	(45.35, 76.27)	(47.02, 77.50)	(4.02, 34.44)	(43.91, 60.95)	(36.00, 78.34)
4	(44.01, 74.51)	(37.78, 63.77)	(0.57, 28.48)	(39.94, 55.93)	(33.35, 69.71)
5	(36.57, 65.83)	(34.55, 62.38)	(-8.87, 21.02)	(30.23, 46.69)	(19.98, 60.20)
6	(41.88, 77.88)	(45.48, 72.45)	(12.88, 40.92)	(52.64, 68.84)	(48.09, 85.80)
Plot	Semiparametric Bootstrap Method:1,000 bootstrapped samples				
1	(9.85, 93.01)	(-2.52, 118.82)	(-28.35 97.54)	(15.59, 82.35)	(-12.04, 118.62)
2	(77.59, 161.45)	(61.61, 185.29)	(35.39 160.09)	(77.15, 145.15)	(57.89, 191.76)
3	(20.50, 103.15)	(12.13, 127.13)	(-20.66 103.08)	(21.76, 88.13)	(-3.48, 128.60)
4	(24.76, 118.24)	(14.56, 155.56)	(-10.10 135.39)	(32.90, 109.94)	(7.08, 157.86)
5	(11.54, 94.69)	(-1.55, 119.78)	(-26.34 99.95)	(16.52, 85.51)	(-10.59, 125.03)
6	(24.23, 110.49)	(12.57, 137.67)	(-11.20 114.04)	(32.65, 99.71)	(6.87, 139.83)

Normal PIs for M at non-sampled plots with level $\alpha = 0.05$ are $\left(\hat{Y}_{s_0} - 1.96\,\widehat{mspe}, \tilde{Y}_{s_0} + 1.96\,\widehat{mspe}\right)$. Bootstrap-t PIs for Tba at non-sampled plots with level $\alpha = 0.05$ are $\left(\tilde{Y}_{s_0} - \hat{t}_{(1-\alpha)}\,\widehat{mspe}, \tilde{Y}_{s_0} + \hat{t}_{(\alpha)}\,\widehat{mspe}\right)$, where the empirical α and $1-\alpha$ quantiles is defined by the k largest and $(b+1-k)$ largest values of $Y^b_{(.),s_0}$, respectively, and $k = [(b+1)\alpha]$, the largest integer $\leq (b+1)\alpha$.

Table 5.10: The Plug-in estimates for Lt at the selected 6 non-sampled plot locations

	Plug-in Estimates[1]				
plot	$\widetilde{Y}_{s_0}$	$\widetilde{Y}_{s_0,I}$	$\widetilde{Y}_{s_0,b}$	$\widetilde{Y}_{s_0,a}$	$\widetilde{Y}_{s_0,u}$
1	366.99	367.64	301.14	360.21	364.78
2	382.81	367.64	301.14	360.21	364.78
3	357.75	367.64	301.14	360.21	364.78
4	364.88	367.64	301.14	360.21	364.78
5	353.88	367.64	301.14	360.21	364.78
6	332.11	367.64	301.14	360.21	364.78
	Plug-in Estimates[1]				
plot	$\hat{\sigma}^2_{Y,p}$	$\hat{\sigma}^2_{Y,p,I}$	$\hat{\sigma}^2_{Y,p,b}$	$\hat{\sigma}^2_{Y,p,a}$	$\hat{\sigma}^2_{Y,p,u}$
1	70124.14	70076.19	80519.00	47181.89	76268.33
2	69990.37	70076.19	80519.00	47181.89	76268.33
3	69971.56	70076.19	80519.00	47181.89	76268.33
4	70066.70	70076.19	80519.00	47181.89	76268.33
5	69976.19	70076.19	80519.00	47181.89	76268.33
6	69986.57	70076.19	80519.00	47181.89	76268.33

[1] The plug-in estimates for $\widetilde{Y}_{s_0}$, $\widetilde{Y}_{s_0,I}$, $\widetilde{Y}_{s_0,b}$, $\widetilde{Y}_{s_0,a}$, $\widetilde{Y}_{s_0,u}$, $\hat{\sigma}^2_{Y,p}$, $\hat{\sigma}^2_{Y,p,I}$, $\hat{\sigma}^2_{Y,p,b}$, $\hat{\sigma}^2_{Y,p,a}$, and $\hat{\sigma}^2_{Y,p,u}$ are calculated by plugging the N observations and the estimated a_i based on the N observatoins into Eqns. (5.20), (5.22), (5.31), (5.38), (5.43), (5.21), (5.23), (5.29), (5.39), and (5.45), repsectively.

Table 5.11: The averages of Lt predictions at the selected 6 non-sampled plot locations from bootstrapped samples

	1,000 bootstrapped samples				
Plot	$\widetilde{\overline{Y}}_{s_0}$	$\widetilde{\overline{Y}}_{s_0,I}$	$\widetilde{\overline{Y}}_{s_0,b}$	$\widetilde{\overline{Y}}_{s_0,a}$	$\widetilde{\overline{Y}}_{s_0,u}$
	Simple Bootstrap Method				
1	366.65	368.39	302.83	362.03	366.84
	(39.72)	(38.98)	(37.31)	(37.11)	(40.84)
2	369.06	367.25	301.64	360.81	365.56
	(40.67)	(35.77)	(32.29)	(33.58)	(37.09)
3	369.03	369.83	304.12	363.25	368.26
	(43.12)	(39.98)	(38.87)	(38.65)	(42.25)
4	367.40	368.21	302.40	361.57	366.34
	(44.44)	(37.97)	(36.03)	(36.27)	(39.73)
5	366.60	366.27	301.45	360.64	365.34
	(42.59)	(38.36)	(34.45)	(35.45)	(39.12)
6	367.82	366.85	301.80	361.02	365.85
	(41.58)	(39.13)	(35.91)	(37.47)	(41.36)
	Semiparametric Bootstrap Method				
1	370.84	370.80	312.95	366.07	369.50
	(64.94)	(66.48)	(64.10)	(64.59)	(67.62)
2	371.00	370.03	312.23	365.27	368.67
	(66.03)	(65.64)	(64.17)	(65.26)	(68.38)
3	370.73	370.78	313.07	366.06	369.56
	(68.32)	(68.54)	(67.23)	(68.42)	(71.42)
4	370.83	371.26	313.20	366.26	369.79
	(66.46)	(66.65)	(64.60)	(66.01)	68.92)
5	367.38	367.94	310.13	363.22	366.42
	(68.10)	(69.35)	(67.34)	(68.79)	(72.57)
6	368.42	368.11	310.25	363.38	366.66
	(68.17)	(69.11)	(67.03)	(68.82)	(72.55)

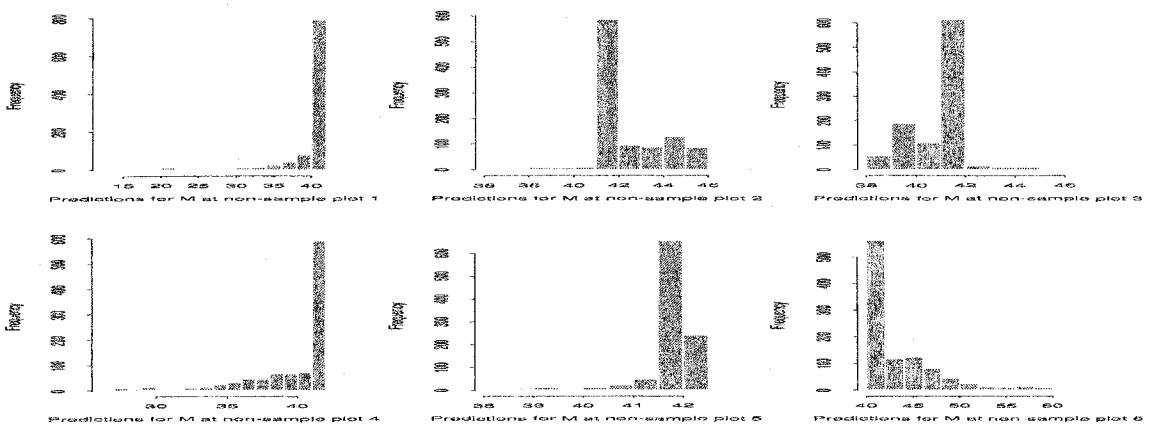
[1] The bootstrapped estimates for $\widetilde{Y}_{s_0}$, $\widetilde{Y}_{s_0,I}$, $\widetilde{Y}_{s_0,b}$, $\widetilde{Y}_{s_0,a}$, $\widetilde{Y}_{s_0,u}$, $\hat{\sigma}^2_{Y,p}$, $\hat{\sigma}^2_{Y,p,I}$, $\hat{\sigma}^2_{Y,p,b}$, $\hat{\sigma}^2_{Y,p,a}$, and $\hat{\sigma}^2_{Y,p,u}$ are calculated by plugging the bootstrapped samples and the estimated a_i based on the bootstrapped samples into Eqns. (5.20), (5.22), (5.31), (5.38), (5.43), (5.21), (5.23), (5.29), (5.39), and (5.45), repsectively. Also, the averages of the bootstrapped estimates are listed and their sample standard deviations are listed in parentheses.

Table 5.12: The normal and bootstrap-t prediction interval (PI) with level $\alpha = 0.05$ estimates for Lt at the selected 6 non-sampled plot locations

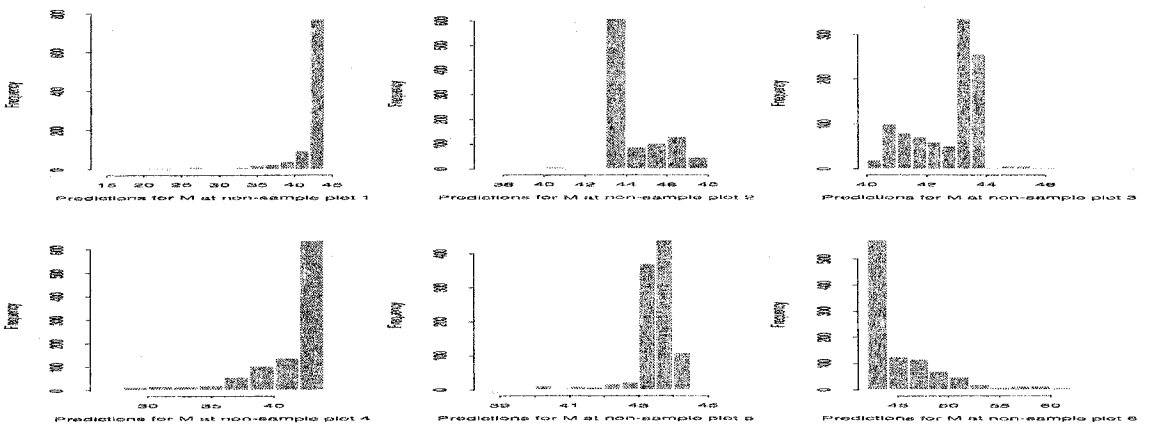
Plot	$\tilde{Y}_{s_0}$	$\tilde{Y}_{s_0,I}$	$\tilde{Y}_{s_0,b}$	$\tilde{Y}_{s_0,a}$	$\tilde{Y}_{s_0,u}$
	Normal Interval Estimates				
1	(-152.03, 886.02)	(-151.21, 886.49)	(-255.03, 857.30)	(-65.53, 785.95)	(-176.51, 906.06)
2	(-135.72, 901.35)	(-151.21, 886.49)	(-255.03, 857.30)	(-65.53, 785.95)	(-176.51, 906.06)
3	(-160.71, 876.21)	(-151.21, 886.49)	(-255.03, 857.30)	(-65.53, 785.95)	(-176.51, 906.06)
4	(-153.93, 883.70)	(-151.21, 886.49)	(-255.03, 857.30)	(-65.53, 785.95)	(-176.51, 906.06)
5	(-164.60, 872.36)	(-151.21, 886.49)	(-255.03, 857.30)	(-65.53, 785.95)	(-176.51, 906.06)
6	(-186.41, 850.63)	(-151.21, 886.49)	(-255.03, 857.30)	(-65.53, 785.95)	(-176.51, 906.06)
	Bootstrap-t Interval Estimates				
Plot	Simple Bootstrap Method:1,000 bootstrapped samples				
1	(514.97, 641.06)	(563.48, 704.04)	(459.80, 623.84)	(466.42, 554.90)	(569.37, 754.65)
2	(518.84, 643.61)	(564.22, 715.59)	(470.98, 625.87)	(470.88, 558.85)	(582.00, 758.81)
3	(514.34, 653.66)	(570.52, 726.38)	(478.52, 637.50)	(475.07, 565.85)	(586.27, 769.57)
4	(511.89, 645.16)	(571.31, 713.74)	(475.22, 634.26)	(473.90, 557.23)	(586.76, 751.72)
5	(515.56, 645.14)	(558.48, 705.80)	(462.54, 617.17)	(466.12, 556.16)	(570.35, 749.66)
6	(512.29, 640.76)	(562.69, 719.29)	(472.00, 634.52)	(469.82, 561.11)	(576.65, 766.37)
Plot	Semiparametric Bootstrap Method:1,000 bootstrapped samples				
1	(435.03, 641.82)	(474.57, 726.97)	(383.46, 672.45)	(393.78, 550.58)	(483.79, 775.28)
2	(434.94, 639.46)	(471.47, 726.77)	(380.55, 668.74)	(393.00, 549.71)	(478.96, 773.91)
3	(433.49, 640.55)	(467.53, 726.94)	(380.25, 673.03)	(390.62, 549.86)	(478.89, 777.44)
4	(433.52, 637.05)	(467.44, 723.88)	(379.58, 668.09)	(390.25, 550.84)	(477.24, 776.45)
5	(423.41, 633.26)	(453.51, 719.50)	(362.34, 660.63)	(382.68, 546.14)	(455.72, 768.07)
6	(429.98, 638.48)	(463.94, 722.66)	(372.72, 663.41)	(387.56, 546.52)	(466.78, 770.66)

Normal PIs for Lt at non-sampled plots with level $\alpha = 0.05$ are $\left(\tilde{Y}_{s_0} - 1.96\,\widehat{mspe}, \tilde{Y}_{s_0} + 1.96\,\widehat{mspe}\right)$. Bootstrap-t PIs for Tba at non-sampled plots with level $\alpha = 0.05$ are $\left(\tilde{Y}_{s_0} - \hat{t}_{(1-\alpha)}\,\widehat{mspe}, \tilde{Y}_{s_0} + \hat{t}_{(\alpha)}\,\widehat{mspe}\right)$, where the empirical α and $1-\alpha$ quantiles is defined by the k largest and $(b+1-k)$ largest values of $Y^b_{(.),s_0}$, respectively, and $k = [(b+1)\alpha]$, the largest integer $\leq (b+1)\alpha$.

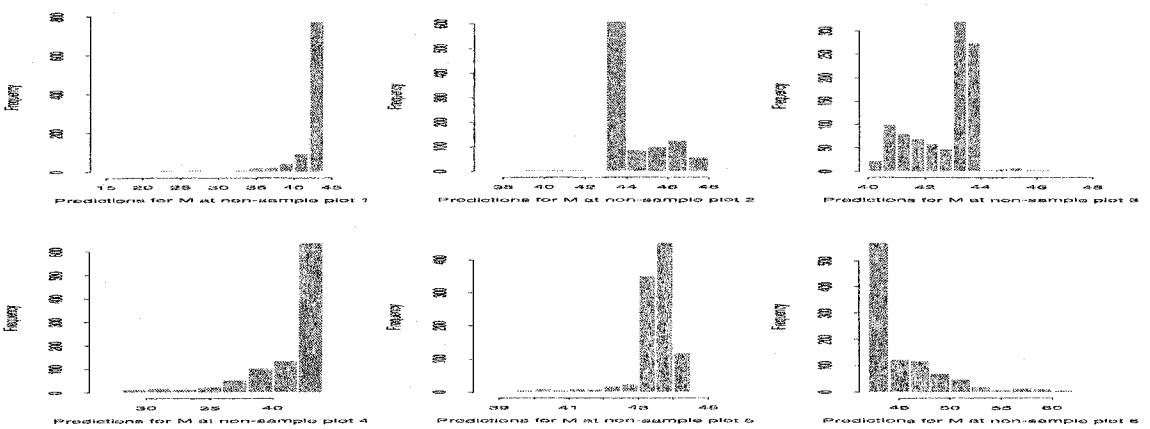
Figure 5.13: 1000 bootstrapping predictions of Tba at 6 non-sampled plots by using the simple bootstrap method [1]



(a). $\hat{Y}_{s_0,b}$



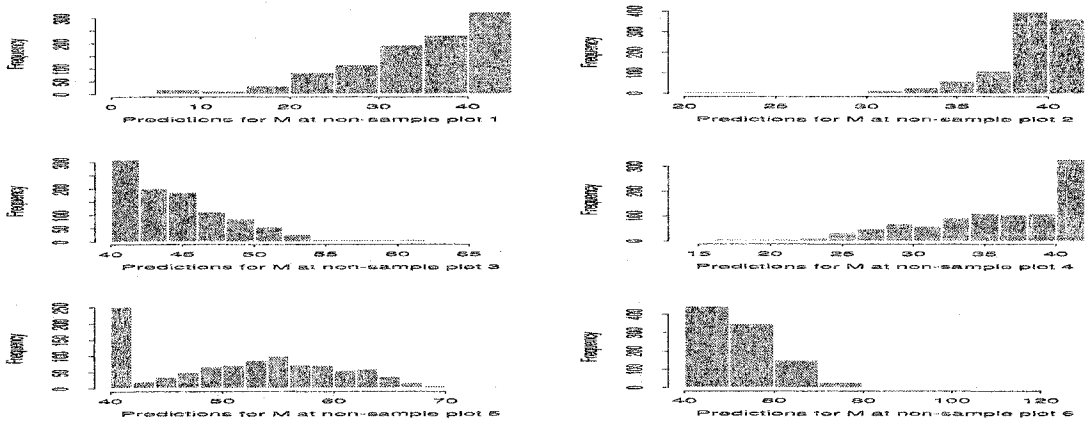
(b). $\hat{Y}_{s_0,a}$



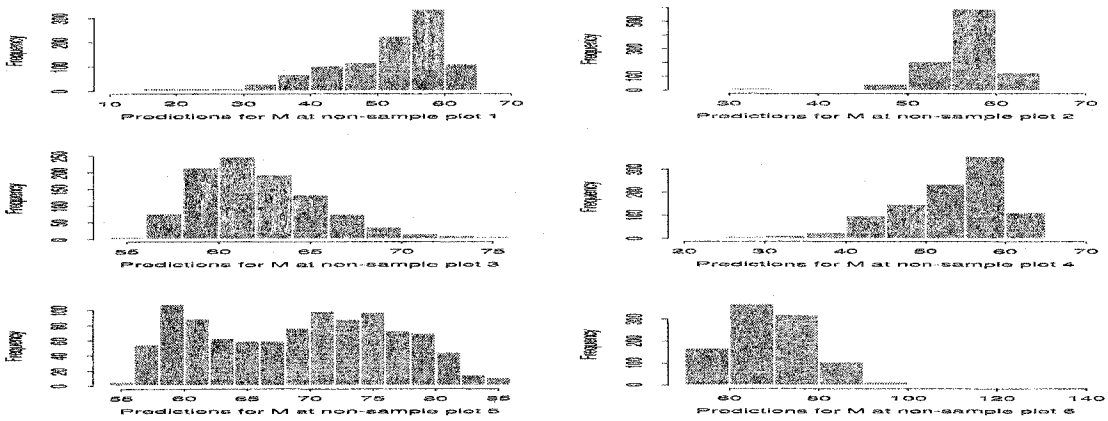
(c). $\hat{Y}_{s_0}$

[1] Six non-sampled plots are depicted in Figure 5.4

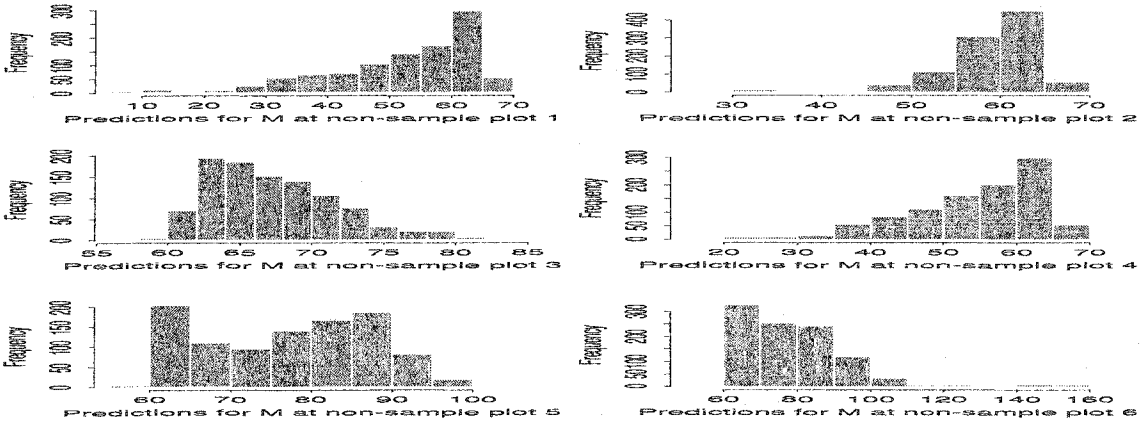
Figure 5.14: 1000 bootstrapping predictions of M at 6 non-sampled plots by using the semiparametric bootstrap method [1]



(a). $\hat{Y}_{s_0,b}$



(b). $\hat{Y}_{s_0,a}$



(c). $\hat{Y}_{s_0}$

[1] Six non-sampled plots are depicted in Figure 5.4

Table 5.13: The Plug-in estimates for Tba at the selected 6 non-sampled plot locations

	Plug-in Estimates[1]				
plot	$\widetilde{Y}_{s_0}$	$\widetilde{Y}_{s_0,I}$	$\widetilde{Y}_{s_0,b}$	$\widetilde{Y}_{s_0,a}$	$\widetilde{Y}_{s_0,u}$
1	43.06	42.62	40.76	42.48	42.47
2	44.91	45.76	43.91	45.63	45.66
3	41.99	41.37	39.52	41.24	41.22
4	41.74	40.75	38.87	40.59	40.57
5	43.76	43.57	42.28	44.00	44.01
6	45.15	46.06	44.88	46.60	46.64
	Plug-in Estimates[1]				
plot	$\hat{\sigma}^2_{Y,p}$	$\hat{\sigma}^2_{Y,p,I}$	$\hat{\sigma}^2_{Y,p,b}$	$\hat{\sigma}^2_{Y,p,a}$	$\hat{\sigma}^2_{Y,p,u}$
1	587.01	590.50	577.61	555.80	574.59
2	586.37	589.59	576.71	554.93	573.70
3	586.33	589.44	576.55	554.79	573.56
4	586.73	590.11	577.22	555.43	574.21
5	586.20	589.48	576.63	554.83	573.60
6	586.32	589.58	576.74	554.93	573.69

[1] The plug-in estimates for $\widetilde{Y}_{s_0}$, $\widetilde{Y}_{s_0,I}$, $\widetilde{Y}_{s_0,b}$, $\widetilde{Y}_{s_0,a}$, $\widetilde{Y}_{s_0,u}$, $\hat{\sigma}^2_{Y,p}$, $\hat{\sigma}^2_{Y,p,I}$, $\hat{\sigma}^2_{Y,p,b}$, $\hat{\sigma}^2_{Y,p,a}$, and $\hat{\sigma}^2_{Y,p,u}$ are calculated by plugging the N observations and the estimated a_i based on the N observatoins into Eqns. (5.20), (5.22), (5.31), (5.38), (5.43), (5.21), (5.23), (5.29), (5.39), and (5.45), repsectively.

Table 5.14: The averages of Tba predictions at the selected 6 non-sampled plot locations from bootstrapped samples

	1,000 bootstrapped samples				
Plot	$\overline{\widetilde{Y}}_{s_0}$	$\overline{\widetilde{Y}}_{s_0,I}$	$\overline{\widetilde{Y}}_{s_0,b}$	$\overline{\widetilde{Y}}_{s_0,a}$	$\overline{\widetilde{Y}}_{s_0,u}$
	Simple Bootstrap Method				
1	43.35	43.40	41.65	43.32	43.33
	(3.72)	(3.96)	(4.03)	(3.94)	(3.96)
2	43.48	43.48	41.74	43.41	43.42
	(3.64)	(3.53)	(3.59)	(3.52)	(3.55)
3	43.34	43.41	41.66	43.33	43.34
	(3.43)	(3.56)	(3.64)	(3.56)	(3.59)
4	43.58	43.49	41.75	43.42	43.43
	(4.10)	(3.89)	(3.93)	(3.87)	(3.90)
5	43.39	43.37	41.64	43.31	43.32
	(3.71)	(3.62)	(3.70)	(3.65)	(3.67)
6	43.59	43.57	41.82	43.49	43.50
	(3.85)	(4.05)	(4.11)	(4.01)	(4.03)
	Semiparametric Bootstrap Method				
1	43.87	43.87	42.10	43.78	43.80
	(5.87)	(5.93)	(5.99)	(5.88)	(5.90)
2	43.55	43.57	41.85	43.52	43.53
	(5.94)	(6.10)	(6.20)	(6.10)	(6.12)
3	43.42	43.46	41.71	43.39	43.40
	(6.13)	(6.11)	(6.19)	(6.10)	(6.12)
4	43.42	43.43	41.70	43.38	43.39
	(6.08)	(6.01)	(6.09)	(6.01)	(6.03)
5	43.76	43.67	41.93	43.61	43.62
	(5.86)	(5.77)	(5.88)	(5.78)	(5.80)
6	43.75	43.72	41.97	43.65	43.67
	(5.66)	(5.58)	(5.67)	(5.58)	(5.60)

[1] The bootstrapped estimates for $\widetilde{Y}_{s_0}$, $\widetilde{Y}_{s_0,I}$, $\widetilde{Y}_{s_0,b}$, $\widetilde{Y}_{s_0,a}$, $\widetilde{Y}_{s_0,u}$, $\hat{\sigma}^2_{Y,p}$, $\hat{\sigma}^2_{Y,p,I}$, $\hat{\sigma}^2_{Y,p,b}$, $\hat{\sigma}^2_{Y,p,a}$, and $\hat{\sigma}^2_{Y,p,u}$ are calculated by plugging the bootstrapped samples and the estimated a_i based on the bootstrapped samples into Eqns. (5.20), (5.22), (5.31), (5.38), (5.43), (5.21), (5.23), (5.29), (5.39), and (5.45), repsectively. Also, the averages of the bootstrapped estimates are listed and their sample standard deviations are listed in parentheses.

Table 5.15: The averages of m.s.p.e. for Tba predictions at the selected 6 non-sampled plot locations from bootstrapped samples

Plot	$\hat{\sigma}^2_{Y,p}$	$\hat{\sigma}^2_{Y,p,I}$	$\hat{\sigma}^2_{Y,p,b}$	$\hat{\sigma}^2_{Y,p,a}$	$\hat{\sigma}^2_{Y,p,u}$
	1,000 bootstrapped samples				
	Simple Bootstrap Method				
1	576.39	576.41	562.23	538.98	559.07
	(57.72)	(62.94)	(73.03)	(67.64)	(72.00)
2	576.11	576.13	561.95	538.72	558.80
	(57.77)	(62.98)	(73.05)	(67.66)	(72.02)
3	576.04	576.06	561.88	538.65	558.73
	(57.81)	(63.01)	(73.07)	(67.69)	(72.04)
4	576.29	576.31	562.13	538.88	558.97
	(57.73)	(62.95)	(73.03)	(67.64)	(71.99)
5	576.13	576.14	561.97	538.72	558.81
	(57.73)	(62.94)	(73.02)	(67.63)	(71.98)
6	576.10	576.12	561.95	538.71	558.79
	(57.79)	(62.98)	(73.06)	(67.67)	(72.02)
	Semiparametric Bootstrap Method				
1	579.24	579.21	567.53	544.21	564.36
	(49.37)	(52.49)	(56.65)	(51.10)	(55.66)
2	578.88	578.85	567.17	543.87	564.01
	(49.39)	(52.50)	(56.64)	(51.09)	(55.64)
3	578.79	578.76	567.08	543.79	563.92
	(49.41)	(52.52)	(56.65)	(51.10)	(55.66)
4	579.11	579.08	567.40	544.09	564.23
	(49.36)	(52.48)	(56.64)	(51.09)	(55.64)
5	578.89	578.85	567.19	543.87	564.01
	(49.36)	(52.48)	(56.62)	(51.07)	(55.62)
6	578.87	578.84	567.17	543.86	563.99
	(49.39)	(52.51)	(56.65)	(51.10)	(55.65)

[1] The bootstrapped estimates for $\widetilde{Y}_{s_0}$, $\widetilde{Y}_{s_0,I}$, $\widetilde{Y}_{s_0,b}$, $\widetilde{Y}_{s_0,a}$, $\widetilde{Y}_{s_0,u}$, $\hat{\sigma}^2_{Y,p}$, $\hat{\sigma}^2_{Y,p,I}$, $\hat{\sigma}^2_{Y,p,b}$, $\hat{\sigma}^2_{Y,p,a}$, and $\hat{\sigma}^2_{Y,p,u}$ are calculated by plugging the bootstrapped samples and the estimated a_i based on the bootstrapped samples into Eqns. (5.20), (5.22), (5.31), (5.38), (5.43), (5.21), (5.23), (5.29), (5.39), and (5.45), repsectively. Also, the averages of the bootstrapped estimates are listed and their sample standard deviations are listed in parentheses.

Table 5.16: The normal and bootstrap-t prediction interval (PI) with level $\alpha = 0.05$ estimates for Tba at the selected 6 non-sampled plot locations

Plot	$\tilde{Y}_{s_0}$	$\tilde{Y}_{s_0,l}$	$\tilde{Y}_{s_0,b}$	$\tilde{Y}_{s_0,a}$	$\tilde{Y}_{s_0,u}$
	Normal Interval Estimates				
1	(-4.42, 90.55)	(-5.00, 90.25)	(-6.35, 87.86)	(-3.73, 88.68)	(-4.51, 89.46)
2	(-2.56, 92.37)	(-1.83, 93.35)	(-3.15, 90.98)	(-0.54, 91.80)	(-1.28, 92.61)
3	(-5.47, 89.45)	(-6.22, 88.95)	(-7.54, 86.59)	(-4.93, 87.40)	(-5.72, 88.16)
4	(-5.74, 89.21)	(-6.86, 88.36)	(-8.22, 85.96)	(-5.61, 86.78)	(-6.40, 87.53)
5	(-3.70, 91.21)	(-4.01, 91.16)	(-4.79, 89.34)	(-2.17, 90.17)	(-2.93, 90.96)
6	(-2.31, 92.61)	(-1.53, 93.65)	(-2.19, 91.95)	(0.43, 92.77)	(-0.30, 93.59)
	Bootstrap-t Interval Estimates				
Plot	Simple Bootstrap Method:1,000 bootstrapped samples				
1	(58.84, 72.23)	(54.95, 69.08)	(40.47, 54.59)	(57.70, 70.92)	(60.33, 73.87)
2	(51.59, 66.74)	(30.07, 44.79)	(6.90, 20.99)	(23.69, 37.70)	(16.14, 30.43)
3	(54.72, 67.96)	(45.37, 58.73)	(29.54, 42.34)	(46.51, 59.36)	(45.59, 58.72)
4	(57.75, 74.14)	(60.06, 76.05)	(48.22, 63.90)	(64.73, 80.57)	(69.58, 85.75)
5	(58.77, 73.30)	(36.22, 51.06)	(10.53, 25.07)	(27.69, 41.52)	(21.34, 35.57)
6	(55.72, 70.79)	(34.47, 50.11)	(15.20, 30.41)	(31.87, 47.16)	(26.89, 42.57)
Plot	Semiparametric Bootstrap Method:1,000 bootstrapped samples				
1	(37.98, 59.58)	(37.43, 59.39)	(33.69, 56.03)	(37.24, 58.74)	(37.00, 58.94)
2	(34.78, 58.27)	(34.38, 58.70)	(30.78, 55.26)	(34.44, 58.06)	(34.22, 58.26)
3	(36.27, 59.68)	(36.72, 59.66)	(33.30, 56.85)	(37.14, 59.39)	(36.94, 59.61)
4	(37.72, 60.91)	(37.99, 61.20)	(34.70, 58.15)	(38.17, 60.96)	(37.97, 61.16)
5	(39.72, 61.82)	(39.87, 61.99)	(36.11, 58.86)	(39.78, 61.43)	(39.57, 61.75)
6	(38.07, 59.91)	(38.36, 59.99)	(35.04, 57.02)	(38.60, 59.78)	(38.42, 60.04)
Plot	Simple Bootstrap Method:10,000 bootstrapped samples				
1	(58.53, 73.98)	(54.86, 70.02)	(40.50, 55.55)	(57.40, 71.95)	(60.04, 74.95)
2	(51.97, 67.44)	(30.49, 45.53)	(6.98, 21.86)	(23.95, 38.30)	(16.38, 31.15)
3	(54.73, 69.35)	(44.83, 59.54)	(29.07, 43.48)	(45.95, 59.99)	(44.99, 59.40)
4	(57.65, 72.72)	(60.48, 75.75)	(48.27, 63.46)	(65.19, 79.83)	(70.00, 85.04)
5	(59.21, 73.91)	(36.96, 51.55)	(11.14, 25.89)	(28.10, 42.27)	(21.82, 36.39)
6	(54.97, 70.82)	(34.71, 50.32)	(15.03, 30.71)	(31.91, 47.10)	(26.91, 42.50)
Plot	Semiparametric Bootstrap Method:10,000 bootstrapped samples				
1	(38.07, 59.91)	(38.36, 59.99)	(35.04, 57.02)	(38.60, 59.78)	(38.42, 60.04)
2	(37.47, 60.80)	(37.40, 61.01)	(33.90, 57.73)	(37.72, 60.44)	(37.48, 60.75)
3	(38.07, 59.91)	(38.36, 59.99)	(35.04, 57.02)	(38.60, 59.78)	(38.42, 60.04)
4	(37.47, 60.80)	(37.40, 61.01)	(33.90, 57.73)	(37.72, 60.44)	(37.48, 60.75)
5	(38.07, 59.91)	(38.36, 59.99)	(35.04, 57.02)	(38.60, 59.78)	(38.42, 60.04)
6	(37.47, 60.80)	(37.40, 61.01)	(33.90, 57.73)	(37.72, 60.44)	(37.48, 60.75)

Normal PIs for Tba at non-sampled plots with level $\alpha = 0.05$ are $\left(\tilde{Y}_{s_0} - 1.96\,\widehat{mspe}, \tilde{Y}_{s_0} + 1.96\,\widehat{mspe}\right)$. Bootstrap-t PIs for Tba at non-sampled plots with level $\alpha = 0.05$ are $\left(\tilde{Y}_{s_0} - \hat{t}_{(1-\alpha)}\,\widehat{mspe}, \tilde{Y}_{s_0} + \hat{t}_{(\alpha)}\,\widehat{mspe}\right)$, where the empirical α and $1-\alpha$ quantiles is defined by the k largest and $(b+1-k)$ largest values of $Y^b_{(.),s_0}$, respectively, and $k = [(b+1)\alpha]$, the largest integer $\leq (b+1)\alpha$.

Table 5.17: Model selection criteria, cross-validation (CV) MSE, for all competing prediction rules based on plot data:

	Estimated CV MSE		
	M	Lt	Tba
mean	4565.89	72554.39	594.19
(I) (1)	4460.45	72427.10	578.65
(I) (2)	4538.72	71076.21	593.15
(II) (1)	4479.64	72604.46	593.24
(II) (2)	4452.72	72083.02	593.61
(III)(1)			
"plain"	4997.42	76491.22	597.11
"approx. unbiased"	4570.10	72191.29	593.59
"unbiased"	4566.68	72135.66	593.59
(III)(2)			
"plain"	4968.46	76914.79	596.38
"approx. unbiased"	4550.76	72601.75	593.17
"unbiased"	4551.08	72603.81	593.22

The prediction rules are listed in Section 5.4 and the correlation function fittings (I)(1) - (III)(2) are listed in Section 5.3.

Table 5.18: Model selection criteria, , cross-validation (CV) MSE, for all competing prediction rules based on subplot data:

	Estimated CV MSE		
	M	Lt	Tba
mean	259.13	4496.99	45.90
(I) (1)	221.58	2982.51	25.11
(I) (2)	222.54	3083.69	23.68
(II) (1)	259.13	3051.68	24.80
(II) (2)	259.13	5043.43	23.42
(III)(1)			
"plain"	410.63	2838.51	25.86
"approx. unbiased"	400.63	2846.88	25.46
"unbiased"	$1.8 \times e^{70}$	2952.69	25.37
(III)(2)			
"plain"	389.95	3305.26	24.32
"approx. unbiased"	382.07	3333.25	24.08
"unbiased"	$3.6 \times e^{61}$	3484.45	25.87

The prediction rules are listed in Section 5.4 and the correlation function fittings (I)(1) - (III)(2) are listed in Section 5.3.

6. SPATIAL ZERO-INFLATED MODELS

A variety of zero-inflated models have been studied in the literature. We list several of these below (Sections 6.1, 6.2, 6.3), but do not utilize them. Instead a simple spatial zero-inflated Poisson model is introduced in Section 6.4 and applied to our data. It did not fit our data, and so the Poisson distribution was replaced by a negative binomial distribution described in Section 6.5. For continuous data, we study a spatial zero-inflated exponential (SZIE) model in Section 6.6 and a spatial zero-inflated gamma (SZIG) model in Section 6.7.

6.1 Poisson Regression Model

The Poisson regression model is the bench mark model for count data in much the same way as the normal linear model is the bench mark for continuous data. In the previous chapter, we ignored the discrete nature of some of the dependent variables. Under the normal linear model, the probability of the zero value of the outcome is zero whereas the probability of the zero values in the discrete case is not zero in general. The standard Poisson regression model makes the following three assumptions:

(1) $Y_i \mid \boldsymbol{x}_i \sim Poisson[\lambda(\boldsymbol{x}_i)]$ where $Y_i \mid \boldsymbol{x}_i$ stands for the conditional distribution of Y_i given the realization $\boldsymbol{x}_i$, which represents the explanatory variables, and $\lambda(\boldsymbol{x}_i) = \mathcal{E}(Y_i \mid \boldsymbol{x}_i)$.

(2) $\lambda(\boldsymbol{x}_i) = exp(\boldsymbol{x}_i'\boldsymbol{\beta})$, $i = 1, \ldots, N$, where $\boldsymbol{\beta}$ is a $k \times 1$ parameter vector.

(3) $Y_i \mid \boldsymbol{x}_i$, $i = 1, \ldots, N$ are independently distributed.

Assumption 3 doesn't allow for spatial dependence of the data. Hence, this Poisson regression model won't be able to interpret the possible spatial dependence in our data.

6.2 Auto-Poisson Model

When spatial data are counts, it is natural to fit a model based on the Poisson distribution. The dependence structure of this type of data can be modeled by a conditional probability approach (Besag, 1974; Gilks et al., 1996). Besag (1974) introduced conditionally specified auto-Poisson models for spatial count data, which link an observation of a Poisson process at a given location with those in its spatial neighborhoods. Suppose that Y_i has a conditional Poisson distribution with mean λ_i dependent upon the neighboring plot (site) values. Such a model is called "auto-Poisson" because it has a conditional "Poisson" distribution and there is "auto" correlation among neighboring observations.

The auto-Poisson conditional specification (assuming pairwise-only dependence between plots) is for $y_i = 0, 1, \ldots,$

$$Pr(Y_i = y_i \mid \{y_j : j \neq i\}) = \frac{exp[-\lambda_i(\{y_j : j \neq i\})][\lambda_i(\{y_j : j \neq i\})]^{y_i}}{y_i!} \tag{6.1}$$

where $\lambda_i(\{y_j : j \neq i\}) = exp\{\alpha_i + \sum_{j=1}^{N} \theta_{ij} y_j\}$, N denotes the number of sample sites and $\theta_{ij} = 0$ unless plots i and j are neighbors of each other under some chosen neighborhood system.

The joint distribution for $\boldsymbol{y}$=$(y_1, y_2, \ldots, y_n)$ is

$$Pr(\boldsymbol{y}) = \frac{exp(\mathrm{Q}(\boldsymbol{y}))}{\sum_{\boldsymbol{z}} exp(\mathrm{Q}(\boldsymbol{z}))} \tag{6.2}$$

where $\sum_{\boldsymbol{z}} exp((\mathrm{Q}(\boldsymbol{z}))$ is the normalizing constant and $\mathrm{Q}(\boldsymbol{y})$ is given by

$$\mathrm{Q}(\boldsymbol{y}) = \sum_{i=1}^{N} \alpha_i y_i + \sum\sum_{1 \leq i \leq j \leq N} \theta_{ij} y_i y_j - \sum_{i=1}^{N} log(y_i!). \tag{6.3}$$

But the auto-Poisson model proposed by Besag (1974) has restrictions on the parameters making it applicable only to spatial data in which the interaction coefficients are non-positive. (i.e. the parameter space is $\{ \boldsymbol{\alpha}, (\theta_{ij})$: $\boldsymbol{\alpha} \in R^N; \theta_{ij} \leq 0$, for all $i, j = 1, \ldots, N \}$). The explanatory variables might be used in modeling by $exp(\alpha_i) = exp(\boldsymbol{x}_i'\boldsymbol{\beta})$ where $\boldsymbol{x}_i'$ is a matrix of covariates at plot i, and $\boldsymbol{\beta}$ is a vector corresponding to the coefficients of $\boldsymbol{x}_i'$.

6.3 Zero-Inflated Count Data Model

For discrete distributions, it is relatively straightforward to select one or more specific outcomes and increase the probability of that outcome or outcomes relative to the probability of some underlying model. The only two restrictions are the fundamental requirements for probabilities, namely that they are non-negative and sum up to one. Such a modeling strategy certainly can improve the ability of the probability model to describe actual discrete data.

Real data can display overdispersion through an excess of zeros. In most such situations, one-parameter distribution fitting for outcomes is not valid in as much as it forces a relationship between the mean and variance (e.g. mean=variance for a Poisson distribution). As mentioned in the previous chapter, transformations to normality can't change the property of excessive zeros for seedling variables M_s, Lt_s and Tba_s. Neither the Poisson regression model nor the Auto-Poisson model has been modified to accommodate an excess of zeros. Also, under the normality assumption of the model in the previous chapter, the probability of any particular outcome is zero whereas the probability of a discrete outcome is not zero in general. Also, the normal model allows for negative outcomes whereas counts are non-negative. We address this phenomenon with a two-component mixture (or compound) model where one component is taken to be a degenerate distribution having mass 1 at $y_i = 0$ and the other is a Poisson distribution. A

mixture model of this type with a degenerate component is sometimes referred to as a nonstandard mixture model (McLachlan and Peel, 2000).

6.3.1 Zero-inflated Poisson (SZIP) model without covariates

The idea of adjusting the probability of a zero outcome for the Poisson distribution is explained by Johnson and Kotz (1969, pp. 204-206) and termed the Poisson with zeros (PWZ) distribution. Other names for these distributions are "inflated" (Singh, 1966; Pandey, 1965) and "pseudo-contagious" Poisson distributions (Cohen, 1960). We refer to this model as a standard "zero-inflated" Poisson (ZIP) model. If Y_i follows a ZIP model, then Y_i can be expressed as

$$Y_i = D_0\,(1 - B_i) + P_i'\,B_i = P_i'\,B_i \tag{6.4}$$

where $D_0 \equiv 0$, the B_i's are independent Bernoulli random variables with mean p, $0 < p < 1$, and the P_i''s are independent and their distribution could be either a Poisson or a "shifted" Poisson with mean parameter λ. The zero-inflated models with a standard Poisson P_i' were discussed by Yip (1988) for zero-inflated count data. He pointed out that explicit expressions for the maximum likelihood estimators (MLEs) of λ and p are very unlikely. Let's take a closer look at the likelihood function of $\boldsymbol{y}$=$(y_1, y_2, \ldots, y_N)$,

$$L(\lambda, p; \boldsymbol{y}) = \Pi_{i=1}^{N} \left\{ \left[(1-p) + pe^{-\lambda}\right] I_{\{0\}}(y_i) + p\,\frac{e^{-\lambda}\lambda^{y_i}}{y_i!}\, I_{\{1,2,\ldots\}}(y_i) \right\}. \tag{6.5}$$

Thus the log-likelihood function is

$$\begin{aligned} & l(\lambda, p; \boldsymbol{y}) \\ = \;& ln\left(L(\lambda, p; \boldsymbol{y})\right) \\ = \;& \sum_{i=1}^{N} I_{\{0\}}(y_i) log\left((1-p) + p\,e^{-\lambda}\right) + \left(N - \sum_{i=1}^{N} I_{\{0\}}(y_i)\right) log(p) \\ & -\lambda\left(N - \sum_{i=1}^{N} I_{\{0\}}(y_i)\right) + \sum_{i=1}^{N} y_i\, log(\lambda) - \sum_{i=1}^{N} log(y_i!) \end{aligned} \tag{6.6}$$

Setting $N_0 = \sum_{i=1}^{N} I_{\{0\}}(Y_i)$ and differentiating equation (6.6) with respect to λ and p, respectively, yields

$$\frac{\partial l}{\partial p} = \frac{-1 + e^{-\lambda}}{1 - p + p\,e^{-\lambda}} N_0 + \frac{N - N_0}{p}, \tag{6.7}$$

and

$$\frac{\partial l}{\partial \lambda} = \frac{-p\,e^{-\lambda}}{1 - p + p\,e^{-\lambda}} N_0 - (N - N_0) + \sum_{i=1}^{N} y_i/\lambda. \tag{6.8}$$

For any given $\lambda \geq log(N/N_0)$, the function $l(\lambda, p; \boldsymbol{y})$ is maximized when

$$p = (1 - e^{-\lambda})^{-1}(1 - \frac{N_0}{N}) \tag{6.9}$$

$$\lambda = \sum_{i=1}^{N} y_i [N - \frac{1-p}{1 - p + pe^{-\lambda}} N_0]^{-1}. \tag{6.10}$$

It's unlikely to get a closed form for the maximum likelihood estimators of λ and p. In Fong and Yip (1993), an EM-algorithm is proposed to estimate the parameters for a mixture model of two discrete distribution components including the Binomial, Negative Binomial and Poisson.

6.4 Spatial Zero-Inflated Poisson Model Without Covariates

6.4.1 Non-spatial motivation

We note that in Eqns. (6.4) and (6.6) there are two types of zeros: one type is obtained as B_i=0; the other as B_i=1 and $y_i = 0$. Next we separate observations with zeros and observations with one or more counts. This can be achieved using a "one-shifted" Poisson distribution instead of a standard Poisson distribution. That is, the P_i' in Eqn. (6.4) takes the form of $P_i + 1$ where $P_i \sim Poisson(\lambda)$. In this paper, we will only consider the zero-inflated Poisson models with a "one-shifted" Poisson distribution and then it is feasible to find a closed form for the maximum likelihood estimators of p and λ (shown later).

If Y_i follows a zero-inflated Poisson with a "one-shifted" Poisson distribution, then by the law of total probability (which is $P(A) = \sum_j P(B_j)P(A|B_j)$, where $\{B_j\}$ is a countable collection of events that partition the sample space), the probability mass function of Y_i is given below:

$$\begin{aligned} & Pr(Y_i = y_i) \\ = \; & Pr(Y_i = 0 \mid B_i = 0)Pr(B_i = 0) + Pr(Y_i = y_i - 1 \mid B_i = 1)Pr(B_i = 1) \\ = \; & (1-p)\, I_{\{0\}}(y_i) + p\, \frac{e^{-\lambda}\lambda^{(y_i-1)}}{(y_i-1)!}\, I_{\{1,2,\ldots\}}(y_i). \end{aligned} \tag{6.11}$$

Observing i.i.d. $Y_1, \ldots, Y_N$, the full likelihood is given by

$$L(\lambda; \boldsymbol{y}) = \Pi_{i=1}^{N} \left[(1-p)\, I_{\{0\}}(y_i) + p\, \frac{e^{-\lambda}\lambda^{(y_i-1)}}{(y_i-1)!}\, I_{\{1,2,\ldots\}}(y_i) \right] \tag{6.12}$$

and the corresponding log-likelihood function is

$$\begin{aligned} l(\lambda, p; \boldsymbol{y}) &= ln\left(L(\lambda, p; \boldsymbol{y})\right) \\ &= \sum_{i=1}^{N} I_{\{0\}}(y_i) log(1-p) + (N - \sum_{i=1}^{N} I_{\{0\}}(y_i)) log(p) - \lambda(N - \sum_{i=1}^{N} I_{\{0\}}(y_i)) \\ &\quad + \sum_{i=1}^{N} (y_i - 1) log(\lambda)\, I_{\{1,2,\ldots\}}(y_i) - \sum_{i=1}^{N} log(y_i - 1)!\, I_{\{1,2,\ldots\}}(y_i). \end{aligned} \tag{6.13}$$

Setting $N_0 = \sum_{i=1}^{N} I_{\{0\}}(Y_i)$ and differentiating with respect to λ and p, respectively, yields

$$\frac{\partial l}{\partial p} = \frac{-N_0}{1-p} + \frac{N - N_0}{p}, \tag{6.14}$$

and

$$\frac{\partial l}{\partial \lambda} = -(N - N_0) + \frac{\sum_{i=1}^{N} Y_i - (N - N_0)}{\lambda}. \tag{6.15}$$

Hence, the maximum likelihood estimates of λ and p are

$$\hat{\lambda} = \frac{\sum_{i=1}^{N} Y_i}{N - N_0} - 1 \text{ and } \hat{p} = \frac{N - N_0}{N}.$$

As mentioned earlier, unlike in Eqns. (6.9) and (6.10) we have an explicit form of the MLE for ZIP models with a "one-shifted" Poisson distribution.

So far we have discussed a 2-component Poisson mixture distribution with one component degenerate at zero and the other having a Poisson distribution. The subsequent section will introduce a Poisson mixture which embeds spatial dependence.

6.4.2 Spatial zero-inflated Poisson model without covariates

The variable Y_i which follows a 9-summand ZIP model can be expressed as

$$\begin{aligned} Y_i &= D_0\,(1-B_i) + (P_i^{(s)} + P_i^{(hr)} + P_i^{(hl)} + P_i^{(vu)} + P_i^{(vd)} \\ &\quad + P_i^{(c_1)} + P_i^{(c_2)} + P_i^{(c_3)} + P_i^{(c_4)} + 1)\,B_i \equiv (P_i+1)B_i \end{aligned} \tag{6.16}$$

where $P_i^{(s)}$ denotes a Poisson distribution associated with the site i and $P_i^{(hr)}$ denotes a Poisson distribution associated with the horizontal-right direction at site (plot) i. The remaining $P^{(\cdot)}$'s are similarly defined; where hl, vu, vd, c_1, c_2, c_3, and c_4 stand, respectively, for horizontal-left, vertical-up, vertical-down, south-west corner, north-east corner, south-east corner, and north-west corner. $P_1^{(\cdot)}, \ldots, P_N^{(\cdot)}$ (N denoting the number of sites) are assumed identically Poisson distributed with mean $\lambda_{(\cdot)}$, where "." ranges over s, hr, hl, vu, vd, c_1, c_2, c_3, and c_4. Furthermore, all the $P_i^{(\cdot)}$'s and B_i's are assumed mutually independent. Such "full" models have ten parameters. Since horizontal-left neighbors and horizontal-right neighbors are equi-distance to plot i, it makes sense to assume the directional means are the same for both left and right horizontal neighbors. Similarly, we can assume there are the same directional means between vertical neighbors and also corner neighbors. To mimic similar correlation schemes in Chapter 4 and reduce the number of parameters, we assume $\lambda_{hl} = \lambda_{hr} = \lambda_h$, $\lambda_{vu} = \lambda_{vd} = \lambda_v$, $\lambda_{c_1} = \lambda_{c_2} = \lambda_{c(1)}$, and $\lambda_{c_3} = \lambda_{c_4} = \lambda_{c(2)}$, then the model has the 6 parameters, p, λ_s, λ_h, λ_v, $\lambda_{c(1)}$, and $\lambda_{c(2)}$. The parameter space is $\Theta = \{\ \theta=(p,\ \lambda_s,\ \lambda_h,\ \lambda_v,$

$\lambda_{c^{(1)}}$, $\lambda_{c^{(2)}}$): $0 \leq p \leq 1$, all λ's ≥ 0 }. The mean schemes for directional means of SZIP models are depicted in Figure 6.1. To distinguish from the standard ZIP model, we call model (6.16) the spatial zero-inflated Poisson (SZIP) model. The SZIP model can accommodate both an excessive number of zeros and spatial dependence in the data.

6.4.3 Parameter estimation

P_i in Eqn. (6.16) has a Poisson distribution with mean λ since the $P^{(\cdot)}$'s are independently Poisson distributed, where $\lambda := \lambda_s + 2(\lambda_h + \lambda_v + \lambda_{c^{(1)}} + \lambda_{c^{(2)}})$. Thus

$$\begin{aligned}
\mathcal{E}[Y_i] &= \mathcal{E}[P_i + 1]\,\mathcal{E}[B_i] \\
&= (\lambda + 1)p, \qquad (6.17) \\
Var[Y_i] &= \mathcal{E}[(P_i + 1)^2 B_i^2] - \mathcal{E}^2[(P_i + 1)B_i] \\
&= \mathcal{E}[B_i^2]\left(Var[P_i + 1] + \mathcal{E}^2[P_i + 1]\right) - \mathcal{E}^2[(P_i + 1)]\mathcal{E}^2[(B_i] \\
&= p\left\{\lambda + (\lambda + 1)^2\right\} - p^2(\lambda + 1)^2 \\
&= p\lambda + p(1 - p)(\lambda + 1)^2. \qquad (6.18)
\end{aligned}$$

Since two adjacent horizontal non-zero-value plots have one horizontal Poisson distribution in common, say $P^{(h)}$, write $Y_i = P^{(h)} + R_i + 1$ and $Y_j = P^{(h)} + R_j + 1$, and then the covariance of Y_i and Y_j is

$$\begin{aligned}
Cov[Y_i, Y_j] &= Cov[B_i\,(R_i + P^{(h)} + 1), B_j\,(R_j + P^{(h)} + 1)] \\
&= Cov[B_i P^{(h)}, B_j P^{(h)}] \\
&= \mathcal{E}[B_i]\mathcal{E}[B_j]\mathcal{E}[(P^{(h)})^2] - (\mathcal{E}[B_i]\mathcal{E}[P^{(h)}])(\mathcal{E}[B_j]\mathcal{E}[P^{(h)}]) \\
&= p^2(\lambda_h + \lambda_h^2) - (p\lambda_h)^2 \\
&= p^2\lambda_h \qquad (6.19)
\end{aligned}$$

and hence the correlation between Y_i and Y_j is

$$\begin{aligned} \rho_{Y_i,Y_j} &= \frac{Cov[Y_i, Y_j]}{Var[Y_i]Var[Y_j]} \\ &= \frac{p\lambda_h}{\lambda + (1-p)(\lambda+1)^2} \end{aligned} \tag{6.20}$$

where $\lambda = \lambda_s + 2(\lambda_h + \lambda_v + \lambda_{c(1)} + \lambda_{c(2)})$.

Similarly, for adjacent vertical plots and for the corner plots, the covariances are respectively $p^2\lambda_v$, $p^2\lambda_{c(1)}$ and $p^2\lambda_{c(2)}$. Since the λ's ≥ 0, the model allows only non-negative correlations for directional neighbors. As λ_s, λ_v, $\lambda_{c(1)}$ and $\lambda_{c(2)}$ tend to zero and p tends to one,

$$\rho_{Y_i,Y_j} \to \frac{1}{2}.$$

Moment estimators for the parameters are

$$\hat{p} = \overline{B} \tag{6.21}$$

$$\hat{\lambda}_h = max(0, S_h/\overline{B}^2) \tag{6.22}$$

$$\hat{\lambda}_v = max(0, S_v/\overline{B}^2) \tag{6.23}$$

$$\hat{\lambda}_{c(1)} = max(0, S_{c(1)}/\overline{B}^2) \tag{6.24}$$

$$\hat{\lambda}_{c(2)} = max(0, S_{c(2)}/\overline{B}^2) \tag{6.25}$$

$$\hat{\lambda}_s = max\left(0, \frac{\overline{Y}}{\overline{B}} - 2(\hat{\lambda}_h + \hat{\lambda}_v + \hat{\lambda}_{c(1)} + \hat{\lambda}_{c(2)}) - 1\right), \tag{6.26}$$

where $\overline{B} = 1 - N_0/N$ is the sample proportion of non-zeros in the sample, N_0 is the number of observed zeros, $\overline{Y}$ is the sample mean, S_h is the sample covariance of adjacent horizontal plots, S_v is the sample covariance of adjacent vertical plots, $S_{c(1)}$ is the sample covariance of adjacent north-east corner and south-west corner plots, and $S_{c(2)}$ is the sample covariance of adjacent south-east corner and north-west corner plots. In the cross products for S_h, S_v, $S_{c(1)}$, and $S_{c(2)}$, we modify the sample covariances using the following different means to get the cross products: (a) mean of non-zeros, (b) overall mean, and (c) individual means for each of the directional neighbors.

6.5 Further Distribution for Count Data

A two-parameter generalization of the one-parameter Poisson distribution is the negative binomial distribution. We introduce a spatial zero-inflated negative binomial (SZINB) mixture model which is similarly constructed as the spatial zero-inflated Poisson models.

6.5.1 Spatial zero-inflated negative binomial (SZINB) model

A random variable U has a negative binomial distribution with parameters $\alpha \geq 0$ and $\zeta \geq 0$, written $U \sim Negbin(\alpha, \zeta)$, if its probability function is given by

$$P(U = k) = \begin{pmatrix} \alpha + k - 1 \\ k \end{pmatrix} \left(\frac{1}{1+\zeta}\right)^k \left(\frac{\zeta}{1+\zeta}\right)^\alpha \quad k = 0, 1, 2, \ldots$$

The mean and variance are given by

$$\mathcal{E}[U] = \alpha\zeta$$

and

$$Var[U] = \alpha\zeta(1+\zeta) = \mathcal{E}[U](1+\zeta).$$

Since $\zeta \geq 0$, the variance of the negative binomial distribution generally exceeds its mean, i.e. there is overdispersion. The overdispersion vanishes in the limit for $\zeta \to 0$; if $\zeta \to 0$ and $\alpha \to \infty$ such that $\alpha\zeta = \lambda$, a constant, then the negative binomial distribution converges to the Poisson distribution with parameter λ.

We will reparameterize from (α, ζ) above to (λ, ζ), where $\lambda = \alpha\zeta = \mathcal{E}[U_i]$. Thus a spatial zero-inflated negative binomial model is given as:

$$Y_i = D_0 (1 - B_i) + (U_i + 1) B_i = (U_i + 1) B_i \tag{6.27}$$

where $D_0 = 0$, the B_i's are independent Bernoulli random variables with mean p, $0 < p < 1$, and the U_i's are defined by

$$\begin{aligned} U_i &= U_i^{(s)} + U_i^{(hr)} + U_i^{(hl)} + U_i^{(vu)} + U_i^{(vd)} + U_i^{(c_1)} + U_i^{(c_2)} \\ &\quad + U_i^{(c_3)} + U_i^{(c_4)}) \end{aligned} \tag{6.28}$$

where $U_i^{(s)}$ denotes a negative binomial distribution, $Negbin(\lambda_s, \zeta)$, associated with the site i and $U_i^{(hr)}$ denotes a negative binomial distribution, $Negbin(\lambda_{hr}, \zeta)$, associated with the horizontal-right direction at site (plot) i. The remaining U's are similarly defined; where hl, vu, vd, c_1, c_2, c_3, and c_4 stand, respectively, for horizontal-left, vertical-up, vertical-down, south-west corner, north-east corner, south-east corner, and north-west corner. $U_1^{(.)}, \ldots, U_N^{(.)}$ are assumed identically negative binomial distributed with $\lambda_{(.)}$ and ζ where "(.)" ranges over s, hr, hl, vu, vd, c_1, c_2, c_3, and c_4. Furthermore, all the $U_i^{(.)}$'s and B_i's are assumed mutually independent for each i. This "full" model has eleven parameters. To reduce the number of parameters, we assume $\lambda_{hl} = \lambda_{hr} = \lambda_h$, $\lambda_{vu} = \lambda_{vd} = \lambda_v$, $\lambda_{c_1} = \lambda_{c_2} = \lambda_{c(1)}$, and $\lambda_{c_3} = \lambda_{c_4} = \lambda_{c(2)}$, so that the model has the 7 parameters, p, ζ, λ_s, λ_h, λ_v, $\lambda_{c(1)}$, and $\lambda_{c(2)}$. For simplicity, we also consider the same for correlation schemes as in Chapter 4. Thus for directional mean schemes, we will consider the same structural schemes as for the SZIP models (see Figure 6.1), but with the additional parameter, ζ.

The main advantage of the negative binomial distribution over the Poisson distribution is that the additional parameter ζ introduces substantial flexibility into the modeling of the variance function; it introduces overdispersion relative to the mean-variance equality implied by the Poisson distribution. Hence, after mixing the k-summand of the negative binomial distributions, a spatial zero-inflated negative binomial (SZINB) model is more overdispersed than a spatial zero-inflated Poisson (SZIP) model as shown later.

6.5.2 Parameter estimation

Assume that Y_i follows a spatial zero-inflated negative binomial (SZINB) model. Then

$$\begin{aligned}
\mathcal{E}[Y_i] &= \mathcal{E}[U_i + 1]\,\mathcal{E}[B_i] \\
&= (\lambda + 1)p, \text{ where } \lambda = \lambda_s + 2(\lambda_h + \lambda_v + \lambda_{c^{(1)}} + \lambda_{c^{(2)}}) \qquad (6.29) \\
Var[Y_i] &= \mathcal{E}[(U_i + 1)^2 B_i^2] - \mathcal{E}^2[(U_i + 1)B_i] \\
&= \mathcal{E}[B_i^2]\left(Var[U_i + 1] + \mathcal{E}^2[U_i + 1]\right) - \mathcal{E}^2[U_i + 1]\mathcal{E}^2[(B_i] \\
&= p\left\{\lambda(1 + \zeta) + (\lambda + 1)^2\right\} - p^2(\lambda + 1)^2 \\
&= p\lambda(1 + \zeta) + p(1 - p)(\lambda + 1)^2. \qquad (6.30)
\end{aligned}$$

We note that the difference between Eqns. (6.18) and (6.30) is $p\lambda\zeta$. The values of p, λ, and ζ are positive, so SZINB models possess a larger variance than SZIP models.

Since two adjacent horizontal plots, say plots i and j, have a horizontal negative binomial distribution, say $U^{(h)}$ in common, write $U_i = U^{(h)} + R_i + 1$ and $U_j = U^{(h)} + R_j + 1$ where R indicates the remainder of variables in Eqn. (6.28). Then the covariance of Y_i and Y_j is

$$\begin{aligned}
Cov[Y_i, Y_j] &= Cov[B_i\,(R_i + U^{(h)} + 1), B_j\,(R_j + U^{(h)} + 1)] \\
&= Cov[B_i U^{(h)}, B_j U^{(h)}] \\
&= \mathcal{E}[B_i]\mathcal{E}[B_j]\mathcal{E}[(U^{(h)})^2] - (\mathcal{E}[B_i]\mathcal{E}[U^{(h)}])(\mathcal{E}[B_j]\mathcal{E}[U^{(h)}]) \\
&= p^2(\lambda_h(1 + \zeta) + \lambda_h^2) - (p\lambda_h)^2 \\
&= p^2\lambda_h(1 + \zeta) \qquad (6.31)
\end{aligned}$$

and hence the correlation between Y_i and Y_j is

$$\begin{aligned}
\rho_{Y_i,Y_j} &= \frac{Cov[Y_i, Y_j]}{Var[Y_i]Var[Y_j]} \\
&= \frac{p\lambda_h(1 + \zeta)}{\lambda(1 + \zeta) + (1 - p)(\lambda + 1)^2} \qquad (6.32)
\end{aligned}$$

Similarly, the correlations between vertical neighbors, between $C^{(1)}$-corner neighbors, and $C^{(2)}$-corner neighbors are

$$\rho_i = \frac{p\lambda_i(1+\zeta)}{\lambda(1+\zeta)+(1-p)(\lambda+1)^2} \quad i = v,\ c^{(1)},\ \text{and}\ c^{(2)}. \tag{6.33}$$

Note that the covariances converge to those from a spatial ZIP model as $\zeta \to 0$.

Moment estimators for the parameters are

$$\hat{p} = \overline{B} \tag{6.34}$$

$$\hat{\zeta} = max\left(0, \frac{S^2 + (1-\frac{1}{\overline{B}})\overline{Y}^2}{\overline{Y}-\overline{B}} - 1\right) \tag{6.35}$$

$$\hat{\lambda}_h = max\left(0, \frac{S_h}{(1+\hat{\zeta})\overline{B}^2}\right) \tag{6.36}$$

$$\hat{\lambda}_v = max\left(0, \frac{S_v}{(1+\hat{\zeta})\overline{B}^2}\right) \tag{6.37}$$

$$\hat{\lambda}_{c^{(1)}} = max\left(0, \frac{S_{c^{(1)}}}{(1+\hat{\zeta})\overline{B}^2}\right) \tag{6.38}$$

$$\hat{\lambda}_{c^{(2)}} = max\left(0, \frac{S_{c^{(2)}}}{(1+\hat{\zeta})\overline{B}^2}\right) \tag{6.39}$$

$$\hat{\lambda}_s = max\left(0, \frac{\overline{Y}}{\overline{B}} - 2(\hat{\lambda}_h + \hat{\lambda}_v + \hat{\lambda}_{c^{(1)}} + \hat{\lambda}_{c^{(2)}}) - 1\right), \tag{6.40}$$

where $\overline{B} = 1 - N_0/N$ is the sample proportion of non-zeros in the sample, N_0 is the number of observed zeros, $\overline{Y}$ is the sample mean, S^2 is the sample variance, S_h is the sample covariance of adjacent horizontal plots, S_v is the sample covariance of adjacent vertical plots, $S_{c^{(1)}}$ is the sample covariance of adjacent north-east corner and south-west corner plots, and $S_{c^{(2)}}$ is the sample covariance of adjacent south-east corner and north-west corner plots.

6.6 Spatial Zero-Inflated Exponential (SZIE) Model for Continuous Data

Y_i follows a spatial Zero-Inflated Exponential (SZIE) model if Y_i can be expressed as

$$Y_i = D_0 (1 - B_i) + U_i B_i = U_i B_i \tag{6.41}$$

where the degenerate variable $D_0 = 0$, the B_i's are independent Bernoulli random variables with mean p, $0 < p < 1$, and the U_i's are defined by

$$\begin{aligned} U_i &= U_i^{(s)} + U_i^{(hr)} + U_i^{(hl)} + U_i^{(vu)} + U_i^{(vd)} + U_i^{(c_1)} + U_i^{(c_2)} \\ &\quad + U_i^{(c_3)} + U_i^{(c_4)}) \end{aligned} \tag{6.42}$$

where $U_i^{(s)}$ denotes an exponential distribution, i.e. $\frac{1}{\beta} e^{-x/\beta}, x > 0$, associated with site (plot) i and $U_i^{(hr)}$ denotes an exponential distribution associated with the horizontal-right direction at site i. The remaining $U_i^{(\cdot)}$'s are similarly defined; where hl, vu, vd, c_1, c_2, c_3, and c_4 indicate respectively, horizontal-left, vertical-up, vertical-down, south-west corner, north-east corner, south-east corner, and north-west corner. The variables $U_1^{(\cdot)}, \ldots, U_N^{(\cdot)}$ are assumed identically exponentially distributed with mean β, where "." ranges over s, hr, hl, vu, vd, c_1, c_2, c_3, and c_4. Further, all the $U_i^{(\cdot)}$'s and B_i's are assumed mutually independent for each i. Such a "full" model has ten parameters. To avoid over-parameterization, we immediately assume $\beta_{hl} = \beta_{hr} = \beta_h$, $\beta_{vu} = \beta_{vd} = \beta_v$, $\beta_{c_1} = \beta_{c_2} = \beta_{c(1)}$, and $\beta_{c_3} = \beta_{c_4} = \beta_{c(2)}$, then the model has 6 parameters, p, β_s, β_h, β_v, $\beta_{c(1)}$, and $\beta_{c(2)}$. In such doing, we have the same correlation scheme for the SZIE as the correlation scheme 2(e) for the models in Chapter 4. Also, we can again consider the directional mean parameter schemes as those for SZIP (see Figure 6.1).

Note that U_i has mean $\beta := \beta_s + 2(\beta_h + \beta_v + \beta_{c^{(1)}} + \beta_{c^{(2)}})$. The parameter space is $\{\ \theta=(p, \beta_s, \beta_h, \beta_v, \beta_{c^{(1)}}, \beta_{c^{(2)}})$: $0 < p < 1$, all β's $> 0\ \}$.

$$\begin{aligned}
\mathcal{E}[Y_i] &= \mathcal{E}[U_i]\,\mathcal{E}[B_i] \\
&= \beta p, \qquad (6.43) \\
Var[Y_i] &= \mathcal{E}[U_i^2 B_i^2] - \mathcal{E}^2[U_i B_i] \\
&= \mathcal{E}[B_i^2]\left(Var[U_i] + \mathcal{E}^2[U_i]\right) - p^2\beta^2 \\
&= p[\beta_s^2 + 2(\beta_h^2 + \beta_v^2 + \beta_{c^{(1)}}^2 + \beta_{c^{(2)}}^2)] + p(1-p)\beta^2 \qquad (6.44) \\
&\equiv g(\theta) \qquad (6.45)
\end{aligned}$$

6.6.1 Parameter estimation

Since two adjacent horizontal plots, say plots i and j have one horizontal exponential distribution, say $U^{(h)}$ in common, write $U_i = U^{(h)} + R_i$ and $U_j = U^{(h)} + R_j$ where R indicates the remainder of variables in Eqn. (6.42), the covariance of Y_i and Y_j is

$$\begin{aligned}
Cov[Y_i, Y_j] &= Cov[B_i\,(R_i + U^{(h)}), B_j\,(R_j + U^{(h)})] \\
&= Cov[B_i U^{(h)}, B_j U^{(h)}] \\
&= \mathcal{E}[B_i]\mathcal{E}[B_j]\mathcal{E}[(U^{(h)})^2] - (\mathcal{E}[B_i]\mathcal{E}[U^{(h)}])(\mathcal{E}[B_j]\mathcal{E}[U^{(h)}]) \\
&= p^2(\beta_h^2 + \beta_h^2) - (p\beta_h)^2 \\
&= p^2\beta_h^2 \qquad (6.46)
\end{aligned}$$

and the correlation between Y_i and Y_j is

$$\begin{aligned}
\rho_{Y_i,Y_j} &= \frac{Cov[Y_i, Y_j]}{Var[Y_i]Var[Y_j]} \\
&= \frac{p^2\beta_h^2}{g(\theta)} \qquad (6.47)
\end{aligned}$$

Similarly, the correlations between vertical neighbors, between $C^{(1)}$-corner neighbors, and $C^{(2)}$-corner neighbors are

$$\rho_i = \frac{p^2\beta_i^2}{g(\theta)}, \quad i = v, c^{(1)}, \text{ and } c^{(2)} \tag{6.48}$$

As β_s, β_v, $\beta_{c^{(1)}}$ and $\beta_{c^{(2)}}$ tend to zero and p tends to one,

$$\rho_{Y_i,Y_j} \to \frac{1}{2}.$$

Similarly, for adjacent vertical plots and for adjacent corner plots, the covariances are respectively $p^2\beta_v^2$, $p^2\beta_{c^{(1)}}^2$ and $p^2\beta_{c^{(2)}}^2$.

Moment estimators for the parameters are

$$\hat{p} = \overline{B} \tag{6.49}$$

$$\hat{\beta}_h = \sqrt{max(0, S_h)}/\overline{B} \tag{6.50}$$

$$\hat{\beta}_v = \sqrt{max(0, S_v)}/\overline{B} \tag{6.51}$$

$$\hat{\beta}_{c^{(1)}} = \sqrt{max(0, S_{c^{(1)}})}/\overline{B} \tag{6.52}$$

$$\hat{\beta}_{c^{(2)}} = \sqrt{max(0, S_{c^{(2)}})}/\overline{B} \tag{6.53}$$

$$\hat{\beta}_s = max\left(0, \frac{\overline{Y}}{\overline{B}} - 2(\hat{\beta}_h + \hat{\beta}_v + \hat{\beta}_{c^{(1)}} + \hat{\beta}_{c^{(2)}})\right), \tag{6.54}$$

where $\overline{B} = 1 - N_0/N$ is the sample proportion of non-zeros in the sample, N_0 is the number of observed zeros, $\overline{Y}$ is the sample mean, S_h is the sample covariance of adjacent horizontal plots, S_v is the sample covariance of adjacent vertical plots, $S_{c^{(1)}}$ is the sample covariance of adjacent north-east and south-west corner plots, and $S_{c^{(2)}}$ is the sample covariance of adjacent south-east and north-west corner plots. In the cross products for S_h, S_v, $S_{c^{(1)}}$, and $S_{c^{(2)}}$, we modify the sample covariances using the following different means to get the cross products (a) mean of non-zeros, (b) overall mean, and (c) individual mean for each directional neighbors.

6.7 Spatial Zero-Inflated Gamma (SZIG) Model for Continuous Data

A generalization of the exponential distribution is the gamma distribution. This subsection introduces a spatial zero-inflated gamma (SZIG) mixture model which is similarly constructed as the spatial zero-inflated exponential models.

A random variable U has a gamma distribution with parameters $\alpha \geq 0$ and $\gamma \geq 0$, denoted by $U \sim Gamma(\alpha/\gamma, \gamma)$, where the mean and variance are given by

$$\mathcal{E}[U] = \alpha$$

and

$$Var[U] = \alpha\gamma = \mathcal{E}[U](\gamma).$$

When $\alpha = \beta = \gamma$, the distribution yields the Exponential distribution.

A spatial zero-inflated gamma model is given as:

$$Y_i = D_0\,(1 - B_i) + U_i\,B_i = U_i\,B_i \tag{6.55}$$

where $D_0 = 0$, the B_i's are independent Bernoulli random variables with mean p, $0 < p < 1$, and the U_i's are defined by

$$\begin{aligned} U_i \;=\; & U_i^{(s)} + U_i^{(hr)} + U_i^{(hl)} + U_i^{(vu)} + U_i^{(vd)} + U_i^{(c_1)} + U_i^{(c_2)} \\ & + U_i^{(c_3)} + U_i^{(c_4)}) \end{aligned} \tag{6.56}$$

where $U_i^{(s)}$ denotes a gamma distribution, $Gamma(\frac{\alpha_s}{\gamma}, \gamma)$, associated with the site i and $U_i^{(hr)}$ denotes a gamma distribution, $Gamma(\frac{\alpha_{hr}}{\gamma}, \gamma)$, associated with the horizontal-right direction at site (plot) i. The remaining $U^{(\cdot)}$'s are similarly defined; where hl, vu, vd, c_1, c_2, c_3, and c_4 denote respectively, horizontal-left, vertical-upper, vertical down, south-west corner, north-east corner, south-east corner, and north-west corner. $U_1^{(\cdot)}, \ldots, U_N^{(\cdot)}$ are assumed identically gamma distributed with α and γ where "$\cdot$" ranges over s, hr, hl, vu, vd, c_1, c_2, c_3, and c_4.

Furthermore, all the $U^{(\cdot)}$'s and B's are assumed mutually independent. Such a "full" model has eleven parameters. To avoid over-parameterization, we assume $\alpha_{hl} = \alpha_{hr} = \alpha_h$, $\alpha_{vu} = \alpha_{vd} = \alpha_v$, $\alpha_{c_1} = \alpha_{c_2} = \alpha_{c^{(1)}}$, and $\alpha_{c_3} = \alpha_{c_4} = \alpha_{c^{(2)}}$, then the model has 7 parameters, p, γ, α_s, α_h, α_v, $\alpha_{c^{(1)}}$, and $\alpha_{c^{(2)}}$. Constructing the similar directional mean schemes as SZIP (see Figure 6.1), the same correlation schemes as those for the models in Chapter 4 exist but under different model assumption.

The main advantage of the gamma distribution over the exponential distribution is that the additional parameter γ introduces substantial flexibility into the modeling of the variance function. Hence, after mixing the 9-summands (see Eqn. (6.56)) of the gamma distributions, a spatial zero-inflated gamma (SZIG) model is more overdispersed than a spatial zero-inflated Exponential (SZIE) model as shown later.

6.7.1 Parameter estimation

Assume that Y_i follows a spatial zero-inflated gamma (SZIG) model. Then

$$\begin{aligned} \mathcal{E}[Y_i] &= \mathcal{E}[U_i]\,\mathcal{E}[B_i] \\ &= \alpha p, \end{aligned} \tag{6.57}$$

where $\alpha = \alpha_s + 2(\alpha_h + \alpha_v + \alpha_{c^{(1)}} + \alpha_{c^{(2)}})$, and

$$\begin{aligned} Var[Y_i] &= \mathcal{E}[U_i^2 B_i^2] - \mathcal{E}^2[U_i B_i] \\ &= \mathcal{E}[B_i^2]\left(Var[U_i] + \mathcal{E}^2[U_i]\right) - \mathcal{E}^2[U_i]\mathcal{E}^2[(B_i] \\ &= p\left\{\alpha\gamma + \alpha^2\right\} - p^2\alpha^2 \\ &= p\alpha\gamma + p(1-p)\alpha^2. \end{aligned} \tag{6.58}$$

Since two adjacent horizontal plots, say plots i and j have one horizontal Gamma distribution, say $U^{(h)}$ in common, write $U_i = U^{(h)} + R_i$ and $U_j = U^{(h)} + R_j$ where R indicates the remainder of variables in Eqn. (6.56), and then the

covariance of Y_i and Y_j is

$$\begin{aligned} Cov[Y_i, Y_j] &= Cov[B_i(R_i + U^{(h)}), B_j(R_j + U^{(h)})] \\ &= Cov[B_i U^{(h)}, B_j U^{(h)}] \\ &= \mathcal{E}[B_i]\mathcal{E}[B_j]\mathcal{E}[(U^{(h)})^2] - (\mathcal{E}[B_i]\mathcal{E}[U^{(h)}])(\mathcal{E}[B_j]\mathcal{E}[U^{(h)}]) \\ &= p^2(\alpha_h\gamma + \alpha_h^2) - (p\alpha_h)^2 \\ &= p^2\alpha_h\gamma \end{aligned} \tag{6.59}$$

and hence the correlation between Y_i and Y_j is

$$\begin{aligned} \rho_{Y_i,Y_j} &= \frac{Cov[Y_i, Y_j]}{Var[Y_i]Var[Y_j]} \\ &= \frac{p\alpha_h\gamma}{\alpha\gamma + (1-p)\alpha^2} \end{aligned} \tag{6.60}$$

Similarly, the correlations between vertical neighbors, $\mathbf{C}^{(1)}$-corner neighbors, and $\mathbf{C}^{(2)}$-corner neighbors is

$$\rho_i = \frac{p\alpha_i\gamma}{\alpha\gamma + (1-p)\alpha^2}, \quad i = v,\ c^{(1)},\ \text{and}\ c^{(2)}. \tag{6.61}$$

Moment estimators for the parameters are

$$\hat{p} = \overline{B} \tag{6.62}$$

$$\hat{\gamma} = max\left(0, \frac{S^2 + (1 - \frac{1}{\overline{B}})\overline{Y}^2}{\overline{Y}}\right) \tag{6.63}$$

$$\hat{\alpha}_h = max\left(0, \frac{S_h}{\hat{\gamma}\hat{p}^2}\right) \tag{6.64}$$

$$\hat{\alpha}_v = max\left(0, \frac{S_v}{\hat{\gamma}\hat{p}^2}\right) \tag{6.65}$$

$$\hat{\alpha}_{c^{(1)}} = max\left(0, \frac{S_{c^{(1)}}}{\hat{\gamma}\hat{p}^2}\right) \tag{6.66}$$

$$\hat{\alpha}_{c^{(2)}} = max\left(0, \frac{S_{c^{(2)}}}{\hat{\gamma}\hat{p}^2}\right) \tag{6.67}$$

$$\hat{\alpha}_s = max\left(0, \frac{\overline{Y}}{\overline{B}} - 2(\hat{\alpha}_h + \hat{\alpha}_v + \hat{\alpha}_{c^{(1)}} + \hat{\alpha}_{c^{(2)}})\right), \tag{6.68}$$

where $\overline{B} = 1 - N_0/N$ is the sample proportion of non-zeros in the sample, N_0 is the number of observed zeros, $\overline{Y}$ is the sample mean, S^2 is the sample variance, S_h is the sample covariance of adjacent horizontal plots, S_v is the sample covariance of adjacent vertical plots, $S_{c(1)}$ is the sample covariance of adjacent north-east corner and south-west corner plots, and $S_{c(2)}$ is the sample covariance of adjacent south-east corner and north-west corner plots.

The estimator of γ in Eqn. (6.63) allows for the zero-value of $\hat{\gamma}$. When $\hat{\gamma} = 0$, U_i in Eqn. (6.56) is degenerate at zero, which causes problems. To overcome this difficulty, an alternative estimator for γ is

$$\hat{\gamma}_{alt} = \frac{S^2_{(-0)}}{\overline{Y}_{(-0)}} \tag{6.69}$$

where $\overline{Y}_{(-0)} = \sum_{i=1}^{N} Y_i/(N - N_0)$ is the sample mean of the $(N - N_0)$ non-zeros, and $S^2_{(-0)}$ is the sample variance of the $(N - N_0)$ non-zeros. Then a non-negative estimated γ is ensured. We didn't obtain zero estimates of γ in our data, but it could happen. Both estimators of γ are applied to our data. The results are summarized in Table (6.8).

For M, Lt, and Tba, prediction rule (3) based on a spatial ZIG model, marked by + in Figure 6.9, has the largest 1CV estimated RMSE. The graph of RMSE of Tba fitted by SZIG is similar to the one fitted by SZIE except for the big jump for scheme (2a). An additional scale parameter γ reduces the overestimation of $\mu = \alpha p$ in SZIG model under scheme (2a). The average of estimated means is 43.62 which is close to the sample mean of Tba values. The smallest RMSE of Tba belongs to SZIG under scheme (2b).

Both spatial zero-inflated models for Tba data agree on the same scheme (2b) with the 1CV estimated RMSE of 19.5 which is close to the smallest RMSE for spatial model (3.4) described earlier.

6.8 Prediction

To make predictions is one of our objectives for this study. We discuss the following prediction rules using spatial zero-inflated (SZI) models shown in Sections 6.4, 6.5, 6.6, and 6.7:

1. predict with estimated mean,

$$\hat{Y}_i = \hat{\mu}, \tag{6.70}$$

2. predict with neighboring observations by weighting them with estimated correlations, The predictor under spatial ZI models is

$$\begin{aligned}\hat{Y}_i = \hat{\mu} &+ \sum_{j \in N_h}^{N} \hat{\rho}_h (Y_j - \hat{\mu}) + \sum_{j \in N_v}^{N} \hat{\rho}_v (Y_j - \hat{\mu}) + \sum_{j \in N_{c^{(1)}}}^{N} \hat{\rho}_{c^{(1)}} (Y_j - \hat{\mu}) \\ &+ \sum_{j \in N_{c^{(2)}}}^{N} \hat{\rho}_{c^{(2)}} (Y_j - \hat{\mu}), \quad i = 1, \ldots, N, \, i \neq j \end{aligned} \tag{6.71}$$

where N_h, N_v, $N_{c^{(1)}}$, and $N_{c^{(1)}}$ respectively denote the neighborhood of plot i in the horizontal, vertical, $\mathbf{C}^{(1)}$-corner and $\mathbf{C}^{(2)}$-corner directional neighbors.

3. predict with neighboring observations by weighting them with estimated correlations and rescaled so that the weights sum up to 1,

The prediction of Y_i using spatial zero-inflated models is usually taken to be

$$\mathcal{E}[Y_i \mid Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_N] = \mathcal{E}[Y_i \mid \text{ neighboring } Y_j's, \, j \neq i].$$

Since it is hard to get an explicit form of the mean of Y_i conditioned on the neighboring Y_j's, a linear combination of the neighboring observations is used instead. Then the prediction of Y_i, $i = 1, \ldots, N$, $i \neq j$, can be

expressed as

$$\hat{Y}_i = \sum_{j \in N_h} \hat{a}_h Y_j + \sum_{j \in N_v} \hat{a}_v Y_j + \sum_{j \in N_{c(1)}} \hat{a}_{c(1)} Y_j + \sum_{j \in N_{c(2)}} \hat{a}_{c(2)} Y_j, \tag{6.72}$$

where N_h, N_v, $N_{c(1)}$, and $N_{c(1)}$ respectively denote the neighborhood of plot i in the horizontal, vertical, $C^{(1)}$-corner and $C^{(2)}$-corner directional neighbors. Thus the mean squared error of the prediction in Eqn. (6.72) is

$$\begin{aligned} & \mathcal{E}[(Y_i - \hat{Y}_i)^2] \\ = \; & \mathcal{E}([Y - \hat{a}_h(Y_{lh} + Y_{rh}) - \hat{a}_v(Y_{uv} + Y_{dv}) - \hat{a}_{c(1)}(Y_{c_1} + Y_{c_2}) \\ & -\hat{a}_{c(2)}(Y_{c_3} + Y_{c_4})]^2) \end{aligned} \tag{6.73}$$

Now constraining $\sum_j a_j = 1$, Eqn. (6.73) yields

$$\begin{aligned} & \mathcal{E}[(Y_i - \hat{Y}_i)^2] \\ = \; & \mathcal{E}([(Y - \mu) - a_h(Y_{lh} - \mu + Y_{rh} - \mu) - a_v(Y_{uv} - \mu + Y_{dv} - \mu) \\ & -a_{c(1)}(Y_{c_1} - \mu + Y_{c_2} - \mu) - a_{c(2)}(Y_{c_3} - \mu + Y_{c_4} - \mu)]^2) \\ = \; & \sigma^2(1 + 2a_h^2 + 2a_v^2 + 2a_{c(1)}^2 + 2a_{c(2)}^2) - 2(2a_h\rho_h\sigma^2) - 2(2a_v\rho_v\sigma^2) \\ & -2(2a_{c(1)}\rho_{c(1)}\sigma^2 + 2a_{c(2)}\rho_{c(2)}\sigma^2) + 2[2a_h a_v(\rho_{c(1)} + \rho_{c(2)})\sigma^2] \\ & +2[2a_h(a_{c(1)} + 2a_{c(2)})\rho_v\sigma^2] + 2[2a_v(a_{c(1)} + 2a_{c(2)})\rho_h\sigma^2] \\ \equiv \; & h\,(a_v, a_h, a_{c(1)}, a_{c(2)}) \end{aligned} \tag{6.74}$$

Eqn. (6.74) can be rewritten as a Lagrangian function corresponding to the constraint $\sum_j a_j = 1$ as

$$\begin{aligned} & g\,(a_v, a_h, a_{c(1)}, a_{c(2)}, \lambda) \\ = \; & h\,(a_v, a_h, a_{c(1)}, a_{c(2)}) - \lambda[2(a_h + a_v + a_{c(1)} + a_{c(2)}) - 1] \end{aligned} \tag{6.75}$$

Differentiating the $g(.)$ with respect to a's and λ, respectively, yields

$$\begin{aligned}
\frac{\partial g}{\partial a_h} &= \sigma^2\left[4a_h - 4\rho_h + a_v(\rho_{c^{(1)}} + \rho_{c^{(2)}}) + 2(a_{c^{(1)}} + 2a_{c^{(2)}})\rho_v\right] - 2\lambda \\
\frac{\partial g}{\partial a_v} &= \sigma^2\left[4a_v - 4\rho_v + a_h(\rho_{c^{(1)}} + \rho_{c^{(2)}}) + 2(a_{c^{(1)}} + 2a_{c^{(2)}})\rho_h\right] - 2\lambda \\
\frac{\partial g}{\partial a_{c^{(1)}}} &= \sigma^2\left[4a_{c^{(1)}} - 4\rho_{c^{(1)}} + 4a_h\rho_v + 4a_v\rho_h\right] - 2\lambda \\
\frac{\partial g}{\partial a_{c^{(2)}}} &= \sigma^2\left[4a_{c^{(2)}} - 4\rho_{c^{(2)}} + 4a_h\rho_v + 4a_v\rho_h\right] - 2\lambda \\
\frac{\partial g}{\partial \lambda} &= 1 - 2(a_h + a_v + a_{c^{(1)}} + a_{c^{(2)}})
\end{aligned}$$

Setting $\frac{\partial g}{\partial a's} = 0$ and $\frac{\partial g}{\partial \lambda} = 0$, the solutions are

$$\begin{aligned}
a_h &= -1/4(6\rho_v - 3\rho_{c^{(1)}} - 3\rho_{c^{(2)}} - 4\rho_h + 4\rho_h^2 - 6\rho_h\rho_{c^{(1)}} - 6\rho_h\rho_{c^{(2)}} - 10\rho_h^2 \\
&\quad +4\rho_h\rho_{c^{(2)}} + 4\rho_h\rho_{c^{(1)}} + 6\rho_h\rho_v + 2\rho_{c^{(2)}}^2 + 4\rho_{c^{(1)}}\rho_{c^{(2)}} + 2\rho_{c^{(1)}} + 1)/D \\
a_v &= -1/4(6\rho_h - 3\rho_{c^{(1)}} - 3\rho_{c^{(2)}} - 4\rho_v + 4\rho_v^2 - 6\rho_v\rho_{c^{(1)}} - 6\rho_v\rho_{c^{(2)}} - 10\rho_v^2 \\
&\quad +4\rho_v\rho_{c^{(2)}} + 4\rho_v\rho_{c^{(1)}} + 6\rho_h\rho_v + 2\rho_{c^{(2)}}^2 + 4\rho_{c^{(1)}}\rho_{c^{(2)}} + 2\rho_{c^{(1)}} + 1)/D \\
a_{c^{(1)}} &= -1/4(1 - 3\rho_h + 6\rho_{c^{(1)}} - 2\rho_{c^{(2)}} - 3\rho_v + 4\rho_{c^{(1)}}\rho_h\rho_v + 2\rho_v^2 - \rho_v\rho_{c^{(1)}} \\
&\quad +7\rho_v\rho_{c^{(2)}} + 2\rho_h^2 + 7\rho_h\rho_{c^{(2)}} - \rho_h\rho_{c^{(1)}} - 4\rho_h\rho_v - \rho_{c^{(2)}}^2 - 6\rho_{c^{(2)}}\rho_{c^{(1)}} \\
&\quad -5\rho_{c^{(1)}}^2 - 4\rho_{c^{(2)}}\rho_h\rho_v - 2\rho_{c^{(1)}}^3 + 2\rho_{c^{(2)}}^3 - 2\rho_{c^{(2)}}\rho_{c^{(1)}}^2 + 4\rho_h\rho_{c^{(1)}}^2 \\
&\quad -2\rho_{c^{(1)}}\rho_h^2 - 2\rho_{c^{(1)}}\rho_v^2 + 2\rho_{c^{(1)}}\rho_{c^{(2)}}^2 + 4\rho_v\rho_{c^{(1)}}^2 + 2\rho_{c^{(2)}}\rho_h^2 \\
&\quad -4\rho_h\rho_{c^{(2)}}^2 - 4\rho_v\rho_{c^{(2)}} + 2\rho_{c^{(2)}}\rho_v)/D
\end{aligned}$$

$$\begin{aligned}
a_{c^{(2)}} &= -1/4(1 - 3\rho_h + 6\rho_{c^{(2)}} - 2\rho_{c^{(1)}} - 3\rho_v + 4\rho_{c^{(1)}}\rho_h\rho_v + 2\rho_v^2 - \rho_v\rho_{c^{(2)}} \\
&\quad +7\rho_v\rho_{c^{(1)}} + 2\rho_h^2 + 7\rho_h\rho_{c^{(1)}} - \rho_h\rho_{c^{(2)}} - 4\rho_h\rho_v - \rho_{c^{(1)}}^2 - 6\rho_{c^{(2)}}\rho_{c^{(1)}} \\
&\quad -5\rho_{c^{(2)}}^2 - 4\rho_{c^{(2)}}\rho_h\rho_v - 2\rho_{c^{(1)}}^3 + 2\rho_{c^{(2)}}^3 - 2\rho_{c^{(2)}}\rho_{c^{(1)}}^2 + 4\rho_h\rho_{c^{(1)}}^2 \\
&\quad -2\rho_{c^{(1)}}\rho_h^2 - 2\rho_{c^{(1)}}\rho_v^2 + 2\rho_{c^{(1)}}\rho_{c^{(2)}}^2 + 4\rho_v\rho_{c^{(1)}}^2 + 2\rho_{c^{(2)}}\rho_h^2 \\
&\quad -4\rho_h\rho_{c^{(2)}}^2 - 4\rho_v\rho_{c^{(2)}} + 2\rho_{c^{(2)}}\rho_v)/D
\end{aligned}$$

and

$$\begin{aligned} D &= -2 + 2\rho_h + 2\rho_{c^{(1)}}\rho_{c^{(2)}} + \rho_{c^{(1)}} + \rho_{c^{(2)}} + 2\rho_v - 2\rho_h\rho_v - 2\rho_v\rho_{c^{(1)}} - 2\rho_v\rho_{c^{(2)}} \\ &\quad -2\rho_h\rho_{c^{(1)}} - 2\rho_h\rho_{c^{(2)}} + \rho_v^2 + \rho_h^2 + \rho_{c^{(2)}}^2 + \rho_{c^{(1)}}^2 \end{aligned}$$

Hence, the prediction for Y_i, $i = 1, \ldots, N$, $i \neq j$, yields

$$\begin{aligned} \hat{Y}_i &= \hat{\mu} + \sum_{j \in N_h} \hat{a}_h(Y_j - \hat{\mu}) + \sum_{j \in N_v} \hat{a}_v(Y_j - \hat{\mu}) + \sum_{j \in N_{c^{(1)}}} \hat{a}_{c^{(1)}}(Y_j - \hat{\mu}) \\ &\quad + \sum_{j \in N_{c^{(2)}}} \hat{a}_{c^{(2)}}(Y_j - \hat{\mu}). \end{aligned} \tag{6.76}$$

where $\hat{a}$'s are explained below.

If Y_i has no neighbors, it is predicted by the estimated mean. If it has neighbors, it is predicted using the weighted function of the $\hat{\rho}$'s for the difference $(Y_j - \hat{\mu})$. Assuming $\rho_h = \rho_v = \rho_{hv}$, the prediction equation becomes

$$\begin{aligned} \hat{Y}_i &= \hat{\mu} + \sum_{j \in N_{hv}} \hat{a}_{hv}(Y_j - \hat{\mu}) + \sum_{j \in N_{c^{(1)}}} \hat{a}_{c^{(1)}}(Y_j - \hat{\mu}) \\ &\quad + \sum_{j \in N_{c^{(2)}}} \hat{a}_{c^{(2)}}(Y_j - \hat{\mu}) \end{aligned} \tag{6.77}$$

where the a's are

$$\begin{aligned} a_{hv} &= \frac{1 + 2(\hat{\rho}_{hv} - \hat{\rho}_{c^{(1)}} - \hat{\rho}_{c^{(2)}})}{4(2 - 4\hat{\rho}_{hv} + \hat{\rho}_{c^{(1)}} + \hat{\rho}_{c^{(2)}})} \\ a_{c^{(1)}} &= \frac{1 - 6\hat{\rho}_{hv} + 7\hat{\rho}_{c^{(1)}} - \hat{\rho}_{c^{(2)}} + 8\hat{\rho}_{hv}(\hat{\rho}_{c^{(2)}} - \hat{\rho}_{c^{(1)}}) + 2(\hat{\rho}_{c^{(1)}}^2 - \hat{\rho}_{c^{(2)}}^2)}{4(2 - 4\hat{\rho}_{hv} + \hat{\rho}_{c^{(1)}} + \hat{\rho}_{c^{(2)}})} \\ a_{c^{(1)}} &= \frac{1 - 6\hat{\rho}_{hv} + 7\hat{\rho}_{c^{(2)}} - \hat{\rho}_{c^{(1)}} + 8\hat{\rho}_{hv}(\hat{\rho}_{c^{(2)}} - \hat{\rho}_{c^{(1)}}) - 2(\hat{\rho}_{c^{(1)}}^2 - \hat{\rho}_{c^{(2)}}^2)}{4(2 - 4\hat{\rho}_{hv} + \hat{\rho}_{c^{(1)}} + \hat{\rho}_{c^{(2)}})} \end{aligned}$$

Assuming $\rho_{c^{(1)}} = \rho_{c^{(2)}} = \rho_c$, the prediction equation is

$$\hat{Y}_i = \hat{\mu} + \sum_{j \in N_h} \hat{a}_h(Y_j - \hat{\mu}) + \sum_{j \in N_v} \hat{a}_v(Y_j - \hat{\mu}) + \sum_{j \in N_c} \hat{a}_c(Y_j - \hat{\mu}) \tag{6.78}$$

Thus

$$\begin{aligned}
\hat{a}_h &= \frac{1+6\hat{\rho}_h-4\hat{\rho}_v-6\hat{\rho}_c+6\hat{\rho}_h\hat{\rho}_v+4(2\hat{\rho}_h-3\hat{\rho}_v)\hat{\rho}_c-10\hat{\rho}_h^2+4\hat{\rho}_v^2+8\hat{\rho}_c^2}{4D} \\
\hat{a}_v &= \frac{1+6\hat{\rho}_v-4\hat{\rho}_h-6\hat{\rho}_c+6\hat{\rho}_h\hat{\rho}_v+4(2\hat{\rho}_v-3\hat{\rho}_h)\hat{\rho}_c-10\hat{\rho}_v^2+4\hat{\rho}_h^2+8\hat{\rho}_c^2}{4D} \\
\hat{a}_c &= \frac{1-3(\hat{\rho}_h+\hat{\rho}_v)+4\hat{\rho}_c-4\hat{\rho}_h\hat{\rho}_v+6(\hat{\rho}_h+\hat{\rho}_v)\hat{\rho}_c+2(\hat{\rho}_h^2+\hat{\rho}_v^2-6\hat{\rho}_c^2)}{4D}
\end{aligned}$$

and

$$D = 2-\hat{\rho}_c-2\hat{\rho}_h-2\hat{\rho}_v+2\hat{\rho}_h\hat{\rho}_v+4\hat{\rho}_v\hat{\rho}_c+4\hat{\rho}_h\hat{\rho}_c-\hat{\rho}_h^2-\hat{\rho}_v^2-4\hat{\rho}_c^2$$

Assuming $\rho_h = \rho_v = \rho_{hv}$ and $\rho_{c^{(1)}} = \rho_{c^{(2)}} = \rho_{hvc}$, the prediction equation is

$$\hat{Y}_i = \hat{\mu} + \sum_{j\in N_{hv}} \hat{a}_{hv}(Y_j-\hat{\mu}) + \sum_{j\in N_c} a_c(Y_j-\hat{\mu}). \tag{6.79}$$

Thus

$$a_{hv} = \frac{1+2\hat{\rho}_{hv}-4\hat{\rho}_c}{8(1-2\hat{\rho}_{hv}+\hat{\rho}_c)} \tag{6.80}$$

$$a_c = \frac{1-6\hat{\rho}_{hv}+6\hat{\rho}_c}{8(1-2\hat{\rho}_{hv}+\hat{\rho}_c)} \tag{6.81}$$

Assuming $\rho_h = \rho_v = \rho_{hv} = \rho_{c^{(1)}} = \rho_{c^{(2)}} = \rho_c$, the prediction equation yields

$$\hat{Y}_i = \hat{\mu} + \sum_{j\in N_{hv}} a'_{hv}(Y_j-\hat{\mu}) + \sum_{j\in N_c} a'_c(Y_j-\hat{\mu}). \tag{6.82}$$

Thus

$$a'_{hv} = \frac{1-2\hat{\rho}_{hv}}{8(1-\hat{\rho}_{hv})} \tag{6.83}$$

$$a'_c = \frac{1}{8(1-\hat{\rho}_{hv})} \tag{6.84}$$

Assuming $\rho_{c^{(1)}} = \rho_{c^{(2)}} = 0$, the prediction equation is

$$\hat{Y}_i = \hat{\mu} + \sum_{j\in N_h} a_h(Y_j-\hat{\mu}) + \sum_{j\in N_v} a_v(Y_j-\hat{\mu}). \tag{6.85}$$

Thus

$$a_h = \frac{1}{4} + \frac{\hat{\rho}_h - \hat{\rho}_v}{2} \tag{6.86}$$

$$a_v = \frac{1}{4} + \frac{\hat{\rho}_v - \hat{\rho}_h}{2} \tag{6.87}$$

Assuming $\rho_h = \rho_v = \rho_{hv}$ and $\rho_{c^{(1)}} = \rho_{c^{(2)}} = 0$, the prediction equation yields

$$\hat{Y}_i = \hat{\mu} + \sum_{j \in N_h} a_h (Y_j - \hat{\mu}) + \sum_{j \in N_v} a_v (Y_j - \hat{\mu}). \tag{6.88}$$

Thus $a_h = 1/4 = a_v$ and the prediction equation yields

$$\hat{Y}_i = \hat{\mu} + 1/4[(Y_{lh} - \hat{\mu}) + (Y_{rh} - \hat{\mu}) + (Y_{uv} - \hat{\mu}) + (Y_{uv} - \hat{\mu})].$$

4. predict with estimated conditional mean and generated Bernoulli B's.

   If $B_i = 1$, then the prediction of Y_i is $\hat{Y}_i = \hat{\mathcal{E}}[Y_i \mid B_i = 1]$; otherwise $\hat{Y}_i = 0$.

6.8.1 Prediction equations for SZIP

To simplify parametrization, we will investigate a 1-fold (leave-one-out) cross-validation (1CV) for estimating RMSEs of the data based on 7 candidate SZIP models with the following parameter assumptions:

(1a) assume that $\lambda_h = \lambda_v = \lambda_{hv}$ and $\lambda_{c^{(1)}} = \lambda_{c^{(2)}} = 0$,

(1b) assume that $\lambda_{c^{(1)}} = \lambda_{c^{(2)}} = 0$,

(2a) assume that $\lambda_h = \lambda_v = \lambda_{c^{(1)}} = \lambda_{c^{(2)}} = \lambda_{hvc}$,

(2b) assume that $\lambda_h = \lambda_v = \lambda_{hv}$ and $\lambda_{c^{(1)}} = \lambda_{c^{(2)}} = \lambda_c$,

(2c) assume that $\lambda_{c^{(1)}} = \lambda_{c^{(2)}} = \lambda_c$,

(2d) assume that $\lambda_h = \lambda_v = \lambda_{hv}$,

(2e) assume that $\lambda_h, \lambda_v, \hat{\lambda}_{c^{(1)}}$, are $\hat{\lambda}_{c^{(2)}}$ are all different.

Under the assumptions above, the directional mean schemes yield the same correlation schemes mentioned in Chapter 4 (see sec. 4.1).

The prediction rules for SZIP are

1. predict with estimated mean

$$\hat{Y}_i = \hat{p}(\hat{\lambda} + 1) \tag{6.89}$$

where $\hat{\lambda} = \hat{\lambda}_s + 2(\hat{\lambda}_h + \hat{\lambda}_v + \hat{\lambda}_{c^{(1)}} + \hat{\lambda}_{c^{(2)}})$. Under the different scheme assumptions, the number of parameters λ_i's is reduced.

2. predict with neighboring observations by weighting them with the estimated correlations.

From Eqn. (6.20), the estimators of ρ_i's are given by

$$\hat{\rho}_i = \frac{\hat{p}\hat{\lambda}_i}{\hat{\lambda} + (1-\hat{p})(\hat{\lambda}+1)^2}, \quad i = h,\ v,\ c^{(1)},\ \text{and}\ c^{(2)}$$

and

$$\hat{\lambda} = \hat{\lambda}_s + 2(\hat{\lambda}_h + \hat{\lambda}_v + \hat{\lambda}_{c^{(1)}} + \hat{\lambda}_{c^{(2)}})$$

3. predict with neighboring observations by weighting them with estimated correlations and rescaled so that the weights sum up to 1

The prediction of Y_i, $i = 1, \ldots, N$, $i \neq j$, can be expressed as

$$\begin{aligned} \hat{Y}_i \;=\; & \sum_{j \in N_h} \hat{a}_h Y_j + \sum_{j \in N_v} \hat{a}_v Y_j + \sum_{j \in N_{c^{(1)}}} \hat{a}_{c^{(1)}} Y_j \\ & + \sum_{j \in N_{c^{(2)}}} \hat{a}_{c^{(2)}} Y_j, \end{aligned} \tag{6.90}$$

where N_h, N_v, $N_{c^{(1)}}$, and $N_{c^{(1)}}$ respectively denote the neighborhood of plot i in the horizontal, vertical, $C^{(1)}$-corner and $C^{(2)}$-corner directional neighbors and $\hat{a}_i$ is a function of $\hat{\rho}_i$ in the previous prediction rule.

4. predict with estimated conditional mean and generated Bernoulli B's.

   If $B_i = 1$, then the prediction of Y_i is $\hat{Y}_i = \hat{\lambda} + 1$; otherwise $\hat{Y}_i = 0$.

6.8.2 Prediction equations for SZINB

The prediction rules for SZINB are

1. predict with estimated mean

$$\hat{Y}_i = \hat{p}(\hat{\lambda} + 1) \tag{6.91}$$

   where $\hat{\lambda} = \hat{\lambda}_s + 2(\hat{\lambda}_h + \hat{\lambda}_v + \hat{\lambda}_{c^{(1)}} + \hat{\lambda}_{c^{(2)}})$ and $\hat{\lambda}_i = \hat{\alpha}_i \hat{\zeta}$ for $i = h,\ v,\ c^{(1)},$ and $c^{(2)}$. Under the different scheme assumptions, the number of λ_i's is reduced.

2. predict with neighboring observations by weighting them with estimated correlations.

   From Eqn. (6.32), the estimators of ρ_i's are given by

$$\hat{\rho}_i = \frac{\hat{p}\hat{\lambda}_i(1+\hat{\zeta})}{\hat{\lambda}(1+\hat{\zeta}) + (1-\hat{p})(\hat{\lambda}+1)^2}, \quad i = h,\ v,\ c^{(1)},\ \text{and } c^{(2)}$$

   and

$$\hat{\lambda} = \hat{\lambda}_s + 2(\hat{\lambda}_h + \hat{\lambda}_v + \hat{\lambda}_{c^{(1)}} + \hat{\lambda}_{c^{(2)}})$$

3. predict with neighboring observations by weighting them with estimated correlations and rescaled so that the weights sum up to 1

   The prediction of Y_i, $i = 1, \ldots, N$, $i \neq j$, can be expressed as

$$\begin{aligned} \hat{Y}_i &= \sum_{j \in N_h} \hat{a}_h Y_j + \sum_{j \in N_v} \hat{a}_v Y_j + \sum_{j \in N_{c^{(1)}}} \hat{a}_{c^{(1)}} Y_j \\ &\quad + \sum_{j \in N_{c^{(2)}}} \hat{a}_{c^{(2)}} Y_j, \end{aligned} \tag{6.92}$$

where N_h, N_v, $N_{c^{(1)}}$, and $N_{c^{(1)}}$ respectively denote the neighborhood of plot i in the horizontal, vertical, $C^{(1)}$-corner and $C^{(2)}$-corner directional neighbors and $\hat{a}_i$ is a function of $\hat{\rho}_i$ in the previous prediction rule.

4. predict with estimated conditional mean and generated Bernoulli B's.

   If $B_i = 1$, then the prediction of Y_i is $\hat{Y}_i = \hat{\lambda} + 1 = \hat{\alpha}_i \hat{\zeta} + 1$, otherwise $\hat{Y}_i = 0$.

6.8.3 Prediction equations for SZIE

The prediction rules for SZIE are

1. predict with estimated mean

$$\hat{Y}_i = \hat{p}\hat{\beta}. \tag{6.93}$$

   where $\hat{\beta} = \hat{\beta}_s + 2(\hat{\beta}_h + \hat{\beta}_v + \hat{\beta}_{c^{(1)}} + \hat{\beta}_{c^{(2)}})$. Under the different scheme assumptions, the number of β_i's is reduced.

2. predict with neighboring observations by weighting them with estimated correlations.

   From Eqns. (6.47) and (6.48), the estimators of ρ_i's are given by

$$\hat{\rho}_i = \frac{\hat{p}^2 \hat{\beta}_i^2}{g(\hat{\theta})}, \quad i = h,\ v,\ c^{(1)},\ \text{and}\ c^{(2)}$$

   and

$$\begin{aligned} g(\hat{\theta}) &= \hat{p}[\hat{\beta}_s^2 + 2(\hat{\beta}_h^2 + \hat{\beta}_v^2 + \hat{\beta}_{c^{(1)}}^2 + \hat{\beta}_{c^{(2)}}^2)] + \hat{p}(1-\hat{p})\hat{\beta}^2 \\ \hat{\beta} &= \hat{\beta}_s + 2(\hat{\beta}_h + \hat{\beta}_v + \hat{\beta}_{c^{(1)}} + \hat{\beta}_{c^{(2)}}) \end{aligned}$$

3. predict with neighboring observations by weighting them with estimated correlations and rescaled so that the weights sum up to 1

The prediction of Y_i, $i = 1, \ldots, N$, $i \neq j$, can be expressed as

$$\begin{aligned} \hat{Y}_i &= \sum_{j \in N_h} \hat{a}_h Y_j + \sum_{j \in N_v} \hat{a}_v Y_j + \sum_{j \in N_{c(1)}} \hat{a}_{c(1)} Y_j \\ &\quad + \sum_{j \in N_{c(2)}} \hat{a}_{c(2)} Y_j, \end{aligned} \tag{6.94}$$

where N_h, N_v, $N_{c(1)}$, and $N_{c(1)}$ respectively denote the neighborhood of plot i in the horizontal, vertical, $\mathbf{C}^{(1)}$-corner and $\mathbf{C}^{(2)}$-corner directional neighbors and $\hat{a}_i$ is a function of $\hat{\rho}_i$ in the previous prediction rule.

4. predict with estimated conditional mean and generated Bernoulli B's.

   If $B_i = 1$, then the prediction of Y_i is $\hat{Y}_i = \hat{\beta}$, otherwise $\hat{Y}_i = 0$.

6.8.4 Prediction equations for SZIG

The prediction rules for SZIG are

1. predict with estimated mean

$$\hat{Y}_i = \hat{p}\hat{\alpha}. \tag{6.95}$$

   where $\hat{\alpha} := \hat{\alpha}_s + 2(\hat{\alpha}_h + \hat{\alpha}_v + \hat{\alpha}_{c(1)} + \hat{\alpha}_{c(2)})$. Under the different scheme assumptions, the number of parameters α_i's is reduced.

2. predict with neighboring observations by weighting them with estimated correlations.

   From Eqns. (6.60) and (6.60), the estimators of ρ_i's are given by

$$\hat{\rho}_{Y_i, Y_j} = \frac{\hat{p}\hat{\alpha}_h\hat{\gamma}}{\hat{\alpha}\hat{\gamma} + (1 - \hat{p})\hat{\alpha}^2} \tag{6.96}$$

   Similarly, the correlation between vertical neighbors, $\mathbf{C}^{(1)}$-corner neighbors, and $\mathbf{C}^{(2)}$-corner neighbors is

$$\hat{\rho}_i = \frac{\hat{p}\hat{\alpha}_i\hat{\gamma}}{\hat{\alpha}\hat{\gamma} + (1 - \hat{p})\hat{\alpha}^2} \quad i = v, c^{(1)}, \text{ and } c^{(2)}. \tag{6.97}$$

3. predict with neighboring observations by weighting them with estimated correlations and rescaled so that the weights sum up to 1

   The prediction of Y_i, $i = 1, \ldots, N$, $i \neq j$, can be expressed as

$$\hat{Y}_i = \sum_{j \in N_h} \hat{a}_h Y_j + \sum_{j \in N_v} \hat{a}_v Y_j + \sum_{j \in N_{c(1)}} \hat{a}_{c(1)} Y_j + \sum_{j \in N_{c(2)}} \hat{a}_{c(2)} Y_j, \tag{6.98}$$

   where N_h, N_v, $N_{c(1)}$, and $N_{c(1)}$ respectively denote the neighborhood of plot i in the horizontal, vertical, $C^{(1)}$-corner and $C^{(2)}$-corner directional neighbors and $\hat{a}_i$ is a function of $\hat{\rho}_i$ in the previous prediction rule.

4. predict with estimated conditional mean and generated Bernoulli B's.
   If $B_i = 1$, then the prediction of Y_i is $\hat{Y}_i = \hat{\alpha}$, otherwise $\hat{Y}_i = 0$.

6.9 Results

6.9.1 Simulated data

In the following, we conduct a simulation study for spatial zero-inflated models based on artificial locations and actual locations in the Siuslaw Forest.

6.9.1.1 spatial zero-inflated Poisson (SZIP) models

The results for simulated data from the SZIP model for comparison of methods (a)-(c), described in p.113, are summarized in Table 6.1. For small p (≤ 0.7), methods (b) and (c) are about the same and much better than method (a). When p increases up to 0.9, the results for method (b) get better, but are still worse than those for the other two methods. For small value $p = 0.3$ and large values of λ's, all three moment estimations for parameters λ_s, λ_h and λ_v are not reliable when applied to small lattices such as 10x10. In general method (a) is the worst.

For small lattice and small p, the histograms for 10×10 with $p = 0.3$, as shown in Figure 6.2, don't resemble the normal distribution. Even though we increase to the larger lattice 20×20, from Figure 6.3 we note that the histograms of estimated λ_s, λ_h, λ_v don't tend to normality, either. For large samples of non-zero sites (i.e. large lattices and large p values), as shown in Figure 6.5 the MOMEs by all of methods tends to normality.

6.9.1.2 spatial zero-inflated Negative Binomial (SZINB) models

We conducted a simulation study for SZINB models with two parameter setting: (1). ζ=0.10, λ_s=0.10, λ_h=1.30, λ_v=0.80, (2). ζ=0.01, λ_s=2.00, λ_h=10.00, λ_v=5.00. In the cross products for S_h, S_v, $S_{c(1)}$, and $S_{c(2)}$, only the overall mean was used in the cross product of the sample covariances. The results for simulated data from SZINB model are summarized Table 6.2.

For parameter setting (1), under 10×10 the average of 100 estimated ζ, $m(\hat{\zeta})$ ranged from 0.16 to 0.12, slightly off away from the actual value $\zeta = 0.10$, and $s(\hat{\zeta})$'s are as high as the value of 0.14 even when $p = 0.99$. When the sample size increases to 20×20 lattice, $m(\hat{\zeta})$'s are about the same as the actual after roundoff to 2 decimals and $s(\hat{\zeta})$'s are half as those under 10×10.

In the parameter setting (2), we chose $\zeta = 0.01$ to generate a SZINB model with the same values of λ's in the second set of generated SZIP realizations. As mentioned earlier, if $\zeta \to 0$ and $\alpha \to \infty$ such that $\alpha\zeta = \lambda$, a constant, then the negative binomial distribution converges to the Poisson distribution with parameter λ. We can see that the estimates of λ's are close to those in Table 6.1.

6.9.1.3 spatial zero-inflated Exponential (SZIE) models

We conducted a simulation study for SZIE models with two parameter setting: (1). β_s=5.00, β_h=2.00, β_v=1.00, and so from Eqn. 6.47 and 6.48 for the values of

p, (0.30 0.50 0.70 0.90 0.99), ρ_s is (0.06, 0.13, 0.25, 0.48, 0.68), ρ_h is (0.01, 0.02, 0.04, 0.08, 0.11), and ρ_v is (0.00, 0.01, 0.01, 0.02, 0.03).

(2). β_s=16.77, β_h=88.84, β_v=88.84, and so from Eqn. 6.47 and 6.48 for the values of p, (0.30 0.50 0.70 0.90 0.99), ρ_s is (0.00, 0.00, 0.00, 0.01, 0.01), and $\rho_h = \rho_v$ is (0.02, 0.04, 0.08, 0.16, 0.24).

In the cross products for S_h, S_v, $S_{c^{(1)}}$, and $S_{c^{(2)}}$, only the overall mean was used in the cross product of the sample covariances. The results for simulated data from SZIE model are summarized Table 6.3.

For the parameter setting, the MOME estimates are getting better as lattice size increases from 10×10, to Siulaw N.F. lattice with holes (312 lattice points), to 20×20. The $s(\beta_{(.)})$ are decreasing, too. But for the second parameter setting, i.e. large β's, the MOME's don't work well even with lattice size 20×20 and $p = 0.99$. From the top panel of Table 6.3, we note that $s(\hat{\beta}_s)$ decrease much faster than the other. It is because the ρ_s in the setting (1) is much larger than the others; it yields 0.68 at $p = 0.99$. It is opposite for the estimates from the second parameter setting; because the ρ_s's are almost zero no matter what values of p we use.

6.9.1.4 spatial zero-inflated Gamma (SZIG) models

We conducted a simulation study for SZIG models with two parameter setting: (1). $\gamma = 10.48$, α_s=8.15, α_h=0.93, α_v=1.89, (2). $\gamma = 192.66$, α_s=208.26, α_h=40.97, α_v=40.97, In the cross products for S_h, S_v, $S_{c^{(1)}}$, and $S_{c^{(2)}}$, only the overall mean was used in the cross product of the sample covariances. The results for simulated data from SZIG model are summarized Table 6.4.

For the parameter setting, the MOME estimates are getting better as lattice size increases from 10×10, to Siulaw N.F. lattice with holes (312 lattice points), to 20×20. The $s(\alpha_{(.)})$ are decreasing, too. But for the second parameter setting,

i.e. large α's, the MOME's don't work well even when lattice size 20×20 and p increases up to 0.9. But it becomes better for p=0.99.

6.9.2 Real data: plot based data

6.9.2.1 spatial zero-inflated Poisson (SZIP) models

We applied the SZIP model to the real data and the results are shown in Table 6.5. The estimates of p for M, Lt, and Tba are over 95%. Hence, the non-zero mean, overall mean, and directional means are quite close and so the estimates of λ's are not obviously different. From the simulation results, we note that for large p (≥ 0.7) the MOMEs are close to the actual parameter values. For M_s, Lt_s, Tba_s, the values of the estimated p are respectively 0.14, 0.56, and 0.42. The λ's estimation by the overall mean in the product (i.e. Method (a)) is the worst scenario. From the simulation results for small p and small lattice, we noted that the MOMEs were not reliable. The reality is that we have only 313 plots, which is not adequate to get reliable MOMEs for M_s and Lt_s. From Table 6.5, we can see that the estimated spatial correlations of the SZIP model with different directional mean schemes (1a)–(2e)ρ's are almost zero for variables M, M_s, Lt, and Lt_s. Under mean scheme (1b), Tba has the estimated spatial correlations, $\hat{\rho}_h = 0$, and $\hat{\rho}_v = 0.27$ by Method (b). When the neighborhood increases up to the second order neighborhood, the correlation seems to disappear due to overestimating in directional Poisson means.

Since the likelihood function is intractable, we couldn't compute the AIC, BIC, and AICC values. We will adopt the RMSEs from leave-one-out cross-validation (1CV) to select the model. Under the neighborhood system we chose, not all of the plots have an equal number of neighbors. To be fair in comparing the performance of models, we only used the 33 plots which have all 8 neighbors under the second neighborhood structure in estimating the 1CV RMSE. We left one out

of 312 plots for each of 33 plots at a time and using the remaining 311 observations to estimate the parameters of models. The y-value for that plot is predicted using the prediction rules 1-4 based on the estimated model. This is repeated until we obtain the predictions for all 33 plots. The 1CV RMSE is calculated by the squared root of the average of squared prediction errors (actual values - predicted values), for all 33 plots. There are no obvious differences between the data with and without the isolated plot (see Table 5.2), so we only discuss the performance of models in the 1CV RMSE estimation for the 312 plots (without the isolated plot). The results for the overall RMSE using the prediction rules from SZIP models are shown in Figure 6.6.

Prediction rules (1) and (2) are about the same since the estimated correlations are almost zeros for all variables except for Tba and Tba_s. The averages of 33 1CV estimated means for M for mean schemes (1a) –(2e) range from 2715.12 (scheme (1b)) to 4058.39 (scheme (2e)). Since the estimated mean plays a large role in the prediction rules, we can see that the prediction rules 1-3, under scheme (1b), is the one with the smallest estimated RMSE. The prediction rule 4, mark by +, has the smallest among all prediction rules and it is much smaller than the others.

Under scheme (1a), $\lambda_h = \lambda_v = \lambda_{hv}$ and $\lambda_{c(1)} = \lambda_{c(2)} = 0$, the weights of predictions using the prediction rule (4) become equal, 1/4 for those four neighboring observations. The estimated overall RMSEs are quite close to the estimated RMSE from the prediction 4 under scheme (1b), $\lambda_{c(1)} = \lambda_{c(2)} = 0$. $\hat{\rho}_h$ is very close to $\hat{\rho}_v$ and so the weights in Eqn. (6.86) and (6.87) are close to 1/4.

For M_s, we note that jumps at mean schemes (1b) and (2c) for predictions rules (1) and (2) due to large values of the estimated mean, the average of the 33 estimated means is 48537.51 for (1b) and 51995.20 for (2c) (the others are under 5079.96).

For Lt, the averages of the 33 estimated means for schemes (2a) -(2e) are almost twice as those of schemes (1a) and (1b). The line for prediction rule (4) is flat except for scheme (2a) due to zero estimated correlations and so the scaled weights become equal among those mean schemes. The big jump of (2c) in the RMSE graph of Lt_s is due to the large estimated mean, 643804.9, which is 1.4 times as that of schemes (1b), (2a), (2b), (2d) and (2e), and 3 times as that of scheme (1a). The jumps in the RMSE graph of Tba_s are for a similar reason.

Finally, we look at the RMSE graph of Tba. All the RMSEs for schemes (1a) and (1b) using the 4 prediction rules are pretty close. That's because of the averages of 33 estimated means are respectively 43.06 and 42.49 and the spatial correlations among the neighbors are higher than the other schemes. The averages of 33 1CV estimated means for the other schemes range from 366.68 to 398.96 which are about ten times that of the schemes (1a) and (1b).

6.9.2.2 spatial zero-inflated Negative Binomial (SZINB) models

The SZINB model has an additional parameter, so it is more flexible than the SZIP model. From Figure 6.7, among the four prediction rules we note that the prediction rule, "+", has the smallest cross-validated (CV) estimated RMSE and the largest one belongs to the prediction rule, "x". We also note that prediction rules, "$\triangle$" and "+", which are a function of the estimated ρ_{ij} work better than the others; which is not surprising because there exits stronger spatial dependence among subplot based data than the plot based data.

To compare with SZIP model, as shown in Figure 6.6, except for Tba_s the CV estimated RMSEs are much smaller than those from SZIP models.

6.9.2.3 Comparison of SZIE model and SZIG model

The SZIG model has an additional parameter γ, so it is more flexible. The 1CV estimated RMSE is used to evaluate the performances of SZI models with dif-

ferent mean directional distributions. The results of leave-one-out cross-validation (1CV) estimated RMSE for the 33 plots with all 8 neighbors for both of the spatial zero-inflated models are depicted in Figures 6.8 and 6.9. From Figure 6.8, we see that there is a jump in the graph of RMSE of Tba fitted by the SZIE model. It is caused by overestimating μ under the scheme (2a) The average of estimated mean for the scheme (2a) is about 53 which is much larger than the average of the schemes (1a), (1b), (2b), (2c), (2d) and (2e) (respectively, 44.1, 44.0, 42, 46.8, 46.7, and 46.6). The estimated correlations $\hat{\rho}_h$ and $\hat{\rho}_v$ are zero if rounded to 2 decimals, so the prediction rules (1), (2), and (4) become approximately the same rule, i.e. predict the values by the estimated mean. Since the estimated p for Tba is 0.99, the prediction rule (4) $\hat{Y}_i = \hat{\theta} \simeq \hat{\theta}\hat{p}$. While the prediction rule (3), +, weighted the neighboring observations by approximate 1/4 and exact 1/4 on horizontal and vertical neighbors for scheme (1a) and (1b), respectively. That's why it has the larger estimated RMSE (about 25, against the sample deviation $\sqrt{592.06} \simeq 24.33$ for 312 plot Tba values). The RMSE of Tba for prediction rule (3), + is much smaller when the neighborhood includes the second order neighbors. The smallest RMSE for Tba, 19.2, is produced by the SZIE model with scheme (2b). Similar trends appear in the other prediction rules.

For M, Lt, and Tba, prediction rule (3) based on spatial ZIG model, marked by + in Figure 4.4, has the largest 1CV estimated RMSE. The graph of RMSE of Tba fitted by SZIG is similar to the one fitted by SZIE except for the big jump for scheme (2a). An additional scale parameter γ reduces the overestimation of $\mu = \alpha p$ in SZIG model under scheme (2a). The average of estimated means is 43.62 which is close to the sample mean of Tba values. The smallest RMSE of Tba belongs to SZIG under scheme (2b).

Both spatial zero-inflated models for Tba data agree on the same scheme (2b) with the 1CV estimated RMSE of 19.5 which is close to the smallest RMSE for spatial model (3.4) described earlier.

6.9.3 Real data: subplot based data

The total number of subplots is $N = 1470$. For model performance, we used leave-one-out cross-validated (1CV) estimated RMSE for a spatial zero-inflated Poisson (SZIP) model, a spatial zero-inflated Negative Binomial (SZINB) model, a spatial zero-inflated Exponential (SZIE) model, and a spatial zero-inflated Gamma (SZIG) model.

6.9.3.1 spatial zero-inflated Poisson (SZIP) models

From Figure 6.10, we note that the prediction 3, marked by +, (see Sec. 6.8.1), has the smallest CV estimated RMSE for all variables and the prediction 4, marked by x has the largest CV estimated RMSE for all variables. This is not surprising. Because the prediction 4 is a conditional mean given by generating Bernoulli, B_i, to predict the value at the deleted subplot, the prediction is over-estimated when the conditional mean $\lambda + 1$ is overestimated. Prediction rule 4 is similar to prediction rule 1 using the concept of estimated conditional mean $\hat{\lambda} + 1$ and estimated mean $\hat{p}(\hat{\lambda} + 1)$ when p is large. With the conditional mean, prediction rule 4 marked by x is slightly worse than prediction 1 marked by o for M, Lt, and Tba, but prediction rule x is a lot worse than prediction rule o since the estimated p for M_s, Lt_s, and Tba_s, are small. As we expected, prediction rule 3 is the best prediction rule among four prediction rules and prediction rule 2 is the second best. This is because the prediction is a function of neighboring observations by weighting them with the estimated correlations and rescaled so that the weights sum up to 1. The "best" prediction belong to the prediction rule 4 with directional mean scheme (2a), which assumes that $\lambda_h = \lambda_v = \lambda_{c(1)} = \lambda_{c(2)} = \lambda_{hvc}$.

6.9.3.2 spatial zero-inflated Negative Binomial (SZINB) models

The advantage of spatial zero-inflated Negative Binomial models over SZIP models is that a SZINB model has an additional parameter ζ in the mean and variance structure. Set $\lambda_i = \alpha\zeta_i$. From Figure 6.11, we can see that prediction rule 4 has the smallest CV estimated RMSE for mortality. The estimated RMSEs are about the same using the prediction rule 4 under directional mean schemes (1a), (1b), and (2a), but there is big jump at directional mean schemes (2b), (2c), (2d) and (2e). Compared to the estimated RMSE of SZIP in Figure 6.10, SZINB models have much smaller estimated RMSE than SZIP models; the best SZINB model is about 18 and the best SZIP model is about 400. Both of them happened when prediction rule 4 with scheme (2a) is used. Prediction rule 4 using SZINB model with directional mean scheme (2a) is the best model for all variables. Except for Tba_s, SZINB models for all variables have the smaller estimated RMSEs than SZIP models.

6.9.3.3 spatial zero-inflated Exponential (SZIE) models

Prediction rule 4 using SZIE models under all directional mean schemes , marked by +, has the smallest estimated RMSE for all variables, as shown in Figure 6.12, and the one under scheme (2a) $\beta_h = \beta_v = \beta_{c^{(1)}} = \beta_{c^{(2)}} = \beta_{hvc}$, is the "best" one for all variables. For M and Lt, compared to Figures 6.10 and 6.11, the estimated RMSEs using SZIE models are twice as much as those using SZINB models, but much smaller than those using SZIP models, denoted as SZINB $<$ SZIE $<\ll$ SZIP. While the estimated RMSEs using SZIE models are slightly smaller than SZINB models and much smaller than those using SZIP models, denoted as SZIE $<$ SZINB $\ll$ SZIP. For M_s and Lt_s, the estimated RMSEs using SZIE models are much smaller than SZINB, and so denoted as SZIE $\ll$ SZINB $\ll$ SZIP. Once again, there is not much improvement for Tba_s.

6.9.3.4 spatial zero-inflated Gamma (SZIG) models

The results of the estimated cross-validated RMSEs from prediction rules using SZIG models under different schemes are depicted in Figure 6.13. For all variables, prediction rule 4 using SZIG models under scheme (2a) has the smallest estimated cross-validated RMSEs. For mortality, the smallest estimated RMSE belongs to prediction rule 4 and it is hard to tell the difference among the RMSEs using prediction rule 4 under schemes (1a), (1b), and (2a). Compared to Figures 6.10, 6.11,and 6.12, the estimated RMSEs using SZIG models are close to those using SZINB models, smaller than those using SZIE, and much smaller than those using SZIP, denoted as SZIG $\approx$ SZINB $<$ SZIE $\ll$ SZIP. The estimated RMSE of Lt predictions, as shown in Figures 6.13, prediction rule 4 using a SZIG model under scheme (2a) $\alpha_h = \alpha_v = \alpha_{c(1)} = \alpha_{c(2)} = \alpha_{hvc}$, has the smallest estimated RMSE and its RMSE is close to SZINB's , smaller than SZIE's, and much smaller than SZIP's, denoted as SZIG $\approx$ SZINB $<$ SZIE $\ll$ SZIP. For Tba, SZIE $<$ SZIG $<$ SZINB $\ll$ SZIP; for M_s, SZIE $\ll$ SZIG $<$ SZINB $\ll$ SZIP; for Lt_s, SZIE $<$ SZINB $\approx$ SZIG $\ll$ SZIP. Once again, there is no obvious improvement for the RMSEs of Tba_s predictions using SZIG models.

6.9.4 Discussion

For plot based data, the ranking of the estimated RMSE of predictions using SZI models are: for M, Lt, and Tba, SZIG $\approx$ SZINB $<$ SZIE $\ll$ SZIP; for M_s, SZIE $<$ SZIG $<$ SZINB $\ll$ SZIP; for Lt_s, SZINB $<$ SZIG $<$ SZIE $\ll$ SZIP; for Tba_s, SZIG $<$ SZIE $<$ SZINB $\approx$ SZIP. The suggested prediction rule for M, Lt, and Tba are prediction rules 1 (marked by o) or 4 (marked by x) using SZIG models under scheme (2a) because of simplicity on parameter setting and the small estimated RMSE. For M_s, prediction 4 using SZIE model with scheme (2a) is suggested although the overall estimated RMSEs are about the same for all 7

schemes. SZIE model under scheme (2a) includes the same number of neighbors as (2b)–(2e), but it only needs to estimate one directional mean, β_{hvc}. For Lt_s we suggest prediction rule 4 SZINB model with scheme (2a) and for Tba_s prediction rule 4 SZIG model with scheme (2a).

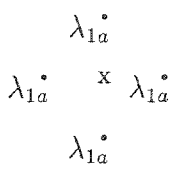
For subplot based data, the estimated RMSEs of M, M_s, Lt, Lt_s, Tba, and Tba_s predictions using the average of the remaining $(N-1) = 1469$ observations to predict the values at the leave-one-out subplot, are 18.95, 12.66, 67.52, 88.27, 6.09, and 0.14.

Among all spatial zero-inflated models, for M prediction we suggest rule 3 from a SZINB model under scheme $\lambda_h = \lambda_v = \lambda_{c(1)} = \lambda_{c(2)} = \lambda_{hvc}$. Its estimated cross-validated RMSE is 19.75 which is about the same as that of predicting using the average of the remaining observations for the deleted subplot. For Lt, prediction rule 3, from the SZINB model under $\lambda_h = \lambda_v = \lambda_{c(1)} = \lambda_{c(2)} = \lambda_{hvc}$, is suggested. Compared to the RMSE of the average prediction, 67.52, the estimate 80.56 of the RMSE for Lt prediction from SZINB is higher, but a SZINB model has advantage of getting a smaller mean squared prediction errors (m.s.p.e.) of predictions. For Tba, the "best" prediction rule is rule 3 from the SZIE model under directional scheme $\beta_h = \beta_v = \beta_{c(1)} = \beta_{c(2)} = \beta_{hvc}$. And the estimated RMSE is 7.78 which is a little bit larger than that of the average, but SZIE has an advantage of obtaining a smaller m.s.p.e. for individual subplot prediction than the average; especially for zero-outcome subplots.

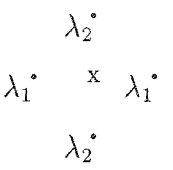
The estimated overall RMSEs of M_s, and Lt_s for SZI models are much larger than those using the average of the remaining observations to predict the deleted subplots. But for some hard-to-predict locations, SZI models have a smaller m.s.p.e.; particularly for the data with excessive zeros. The percentages of zero for M_s and for Lt_s are 0.96% and 70%, respectively. Although the overall estimated RMSEs are much larger than the average's, SZI models work better than

the average on m.s.p.e of individual predictions. No matter what kind of spatial zero-inflated model we used, the estimated overall RMSEs of Tba_s predictions from prediction rule 3 marked by + are about the smallest; under scheme (2a), 0.15 for SZIP, 0.15 for SZINB, 0.17 for SZIE, and 0.18 for SZIG.

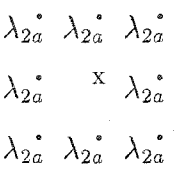
Figure 6.1: The directional Poisson mean schemes for the neighborhood of plot i [1]



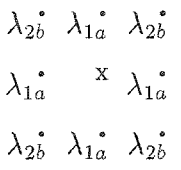
(1a). symmetric first order scheme $(S_{\lambda_1\lambda_1 00})$



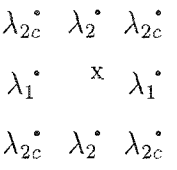
(1b). asymmetric first order scheme $(S_{\lambda_1\lambda_2 00})$



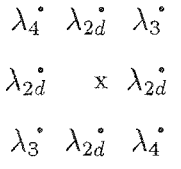
(2a). symmetric second order scheme $(S_{\lambda_1\lambda_1\lambda_1\lambda_1})$



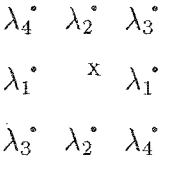
(2b). asymmetric second order scheme $(S_{\lambda_1\lambda_1\lambda_3\lambda_3})$



(2c). asymmetric second order scheme $(S_{\lambda_1\lambda_2\lambda_3\lambda_3})$



(2d). asymmetric second order scheme $(S_{\lambda_1\lambda_1\lambda_3\lambda_4})$



(2e). asymmetric second order scheme $(S_{\lambda_1\lambda_2\lambda_3\lambda_4})$

[1] x indicates the location of plot i.

Table 6.1: The sample means and standard deviations (s.d.) of 100 estimated parameters from the 100 realizations of a SZIP model.

		(a)			(b)			(c)		
True	p	λ_s	λ_h	λ_v	λ_s	λ_h	λ_v	λ_s	λ_h	λ_v
mean	$m(\hat{p})$	$m(\hat{\lambda}_s)$	$m(\hat{\lambda}_h)$	$m(\hat{\lambda}_v)$	$m(\hat{\lambda}_s)$	$m(\hat{\lambda}_h)$	$m(\hat{\lambda}_v)$	$m(\hat{\lambda}_s)$	$m(\hat{\lambda}_h)$	$m(\hat{\lambda}_v)$
s.d.	$s(\hat{p})$	$s(\hat{\lambda}_s)$	$s(\hat{\lambda}_h)$	$s(\hat{\lambda}_v)$	$s(\hat{\lambda}_s)$	$s(\hat{\lambda}_h)$	$s(\hat{\lambda}_v)$	$s(\hat{\lambda}_s)$	$s(\hat{\lambda}_h)$	$s(\hat{\lambda}_v)$
10x10	0.30	5.00	2.00	1.00	5.00	2.00	1.00	5.00	2.00	1.00
	0.30	0.00	919.62	912.38	2.82	18.1	18.50	2.83	17.95	18.40
	0.05	0.00	500.63	504.93	4.68	24.3	28.86	4.69	24.24	28.82
10x10	0.50	5.00	2.00	1.00	5.00	2.00	1.00	5.00	2.00	1.00
	0.50	0.00	159.38	159.01	2.93	6.53	8.29	2.94	6.48	8.24
	0.05	0.00	85.17	80.58	4.67	9.37	11.23	4.67	9.33	11.19
10x10	0.70	5.00	2.00	1.00	5.00	2.00	1.00	5.00	2.00	1.00
	0.71	0.00	25.58	27.59	4.61	2.76	3.83	4.63	2.75	3.81
	0.04	0.00	13.92	14.21	4.90	4.71	5.87	4.90	4.69	5.85
10x10	0.90	5.00	2.00	1.00	5.00	2.00	1.00	5.00	2.00	1.00
	0.91	3.02	2.65	3.05	5.39	1.55	1.67	5.42	1.54	1.66
	0.03	3.89	2.73	2.71	4.06	2.04	2.01	4.06	2.03	2.00
10x10	0.99	5.00	2.00	1.00	5.00	2.00	1.00	5.00	2.00	1.00
	0.99	5.49	1.86	0.94	5.55	1.84	0.93	5.56	1.83	0.92
	0.01	3.36	1.41	0.92	3.38	1.41	0.91	3.37	1.40	0.91
10x10	0.30	2.00	10.00	5.00	2.00	10.00	5.00	2.00	10.00	5.00
	0.30	0.00	6583.36	6597.93	10.04	85.11	92.18	10.07	84.67	91.72
	0.05	0.00	3196.04	3234.55	14.94	133.03	158.57	14.94	132.56	158.27
10x10	0.50	2.00	10.00	5.00	2.00	10.00	5.00	2.00	10.00	5.00
	0.50	0.00	1182.80	1193.86	13.06	35.49	36.74	13.10	35.19	36.53
	0.05	0.00	522.24	521.90	15.44	60.78	64.94	15.45	60.59	64.77
10x10	0.70	2.00	10.00	5.00	2.00	10.00	5.00	2.00	10.00	5.00
	0.70	0.00	198.23	216.26	10.39	15.46	25.52	10.48	15.37	25.38
	0.04	0.00	95.70	96.35	14.10	28.62	37.90	14.17	28.52	37.87
10x10	0.90	2.00	10.00	5.00	2.00	10.00	5.00	2.00	10.00	5.00
	0.90	5.45	18.72	16.36	15.95	7.34	5.69	16.01	7.29	5.65
	0.03	10.27	21.16	14.85	13.68	12.00	8.80	13.68	11.97	8.77
10x10	0.99	2.00	10.00	5.00	2.00	10.00	5.00	2.00	10.00	5.00
	0.99	10.17	7.54	4.57	10.68	7.35	4.43	10.74	7.33	4.41
	0.01	9.34	4.60	4.11	9.52	4.59	4.08	9.52	4.58	4.07

Note: The location points we chose are (1) artificial locations: 10x10 lattice, (2) the actual location points (N=312) in the Siuslaw Forest, or (3) artificial locations: 20x20 lattice. The first row in each lattice size and each p represent the actual values of parameters in the chosen spatial zero-inflated Poisson (SZIP) model. Methods (a), (b), and (c) respectively denote that the sample covariances are modified by using mean of non-zeros, overall mean, and individual means for each of the directional neighbors.

(Continued)

		(a)			(b)			(c)		
True	p	λ_s	λ_h	λ_v	λ_s	λ_h	λ_v	λ_s	λ_h	λ_v
mean	$m(\hat{p})$	$m(\hat{\lambda}_s)$	$m(\hat{\lambda}_h)$	$m(\hat{\lambda}_v)$	$m(\hat{\lambda}_s)$	$m(\hat{\lambda}_h)$	$m(\hat{\lambda}_v)$	$m(\hat{\lambda}_s)$	$m(\hat{\lambda}_h)$	$m(\hat{\lambda}_v)$
s.d.	$s(\hat{p})$	$s(\hat{\lambda}_s)$	$s(\hat{\lambda}_h)$	$s(\hat{\lambda}_v)$	$s(\hat{\lambda}_s)$	$s(\hat{\lambda}_h)$	$s(\hat{\lambda}_v)$	$s(\hat{\lambda}_s)$	$s(\hat{\lambda}_h)$	$s(\hat{\lambda}_v)$
312	0.30	5.00	2.00	1.00	5.00	2.00	1.00	5.00	2.00	1.00
	0.30	0.00	810.50	820.53	2.44	9.39	13.23	2.50	9.24	13.07
	0.02	0.00	193.81	192.48	4.37	13.78	16.01	4.39	13.71	15.92
312	0.50	5.00	2.00	1.00	5.00	2.00	1.00	5.00	2.00	1.00
	0.50	0.00	147.07	149.33	2.85	4.43	6.53	2.88	4.36	6.46
	0.03	0.00	38.91	39.41	4.45	5.86	8.32	4.46	5.80	8.28
312	0.70	5.00	2.00	1.00	5.00	2.00	1.00	5.00	2.00	1.00
	0.70	0.00	27.37	27.63	4.46	2.43	2.52	4.50	2.39	2.49
	0.02	0.00	8.45	8.46	4.72	3.67	3.50	4.73	3.65	3.49
312	0.90	5.00	2.00	1.00	5.00	2.00	1.00	5.00	2.00	1.00
	0.90	3.10	2.30	2.49	6.89	0.98	1.10	6.95	0.97	1.09
	0.02	3.68	2.04	2.05	3.53	1.44	1.44	3.51	1.42	1.44
312	0.99	5.00	2.00	1.00	5.00	2.00	1.00	5.00	2.00	1.00
	0.99	7.39	0.98	0.80	7.45	0.96	0.79	7.49	0.96	0.78
	0.01	2.18	0.80	0.66	2.18	0.80	0.65	2.17	0.80	0.65
312	0.30	2.00	10.00	5.00	2.00	10.00	5.00	2.00	10.00	5.00
	0.30	0.00	5880.56	5895.69	9.84	58.99	68.07	10.13	58.28	67.07
	0.02	0.00	1306.22	1335.98	14.43	95.82	101.96	14.50	95.11	101.09
312	0.50	2.00	10.00	5.00	2.00	10.00	5.00	2.00	10.00	5.00
	0.50	0.00	1084.32	1107.34	9.01	32.78	35.05	9.26	32.20	34.58
	0.02	0.00	228.55	237.92	13.99	48.48	49.75	14.10	48.11	49.46
312	0.70	2.00	10.00	5.00	2.00	10.00	5.00	2.00	10.00	5.00
	0.70	0.00	196.54	202.05	8.48	13.83	20.63	8.75	13.56	20.39
	0.02	0.00	53.80	58.45	12.54	19.18	24.71	12.69	19.07	24.61
312	0.90	2.00	10.00	5.00	2.00	10.00	5.00	2.00	10.00	5.00
	0.90	2.32	15.28	17.82	14.67	5.06	6.65	14.83	4.96	6.57
	0.02	5.74	11.60	10.94	12.92	7.06	8.39	12.97	6.98	8.35
312	0.99	2.00	10.00	5.00	2.00	10.00	5.00	2.00	10.00	5.00
	0.99	15.31	4.86	3.63	15.70	4.74	3.54	15.84	4.71	3.50
	0.00	8.30	3.61	2.74	8.39	3.61	2.72	8.38	3.60	2.72

Note: The location points we chose are (1) artificial locations: 10x10 lattice, (2) the actual location points (N=312) in the Siuslaw Forest, or (3) artificial locations: 20x20 lattice. The first row in each lattice size and each p represent the actual values of parameters in the chosen spatial zero-inflated Poisson (SZIP) model. Methods (a), (b), and (c) respectively denote that the sample covariances are modified by using mean of non-zeros, overall mean, and individual means for each of the directional neighbors.

(Continued)

		(a)			(b)			(c)		
True	p	λ_s	λ_h	λ_v	λ_s	λ_h	λ_v	λ_s	λ_h	λ_v
mean	$m(\hat{p})$	$m(\hat{\lambda}_s)$	$m(\hat{\lambda}_h)$	$m(\hat{\lambda}_v)$	$m(\hat{\lambda}_s)$	$m(\hat{\lambda}_h)$	$m(\hat{\lambda}_v)$	$m(\hat{\lambda}_s)$	$m(\hat{\lambda}_h)$	$m(\hat{\lambda}_v)$
s.d.	$s(\hat{p})$	$s(\hat{\lambda}_s)$	$s(\hat{\lambda}_h)$	$s(\hat{\lambda}_v)$	$s(\hat{\lambda}_s)$	$s(\hat{\lambda}_h)$	$s(\hat{\lambda}_v)$	$s(\hat{\lambda}_s)$	$s(\hat{\lambda}_h)$	$s(\hat{\lambda}_v)$
20x20	0.30	5.00	2.00	1.00	5.00	2.00	1.00	5.00	2.00	1.00
	0.30	0.00	785.41	789.62	3.31	5.72	9.00	3.32	5.71	8.98
	0.02	0.00	189.40	185.68	4.71	9.58	11.64	4.72	9.58	11.62
20x20	0.50	5.00	2.00	1.00	5.00	2.00	1.00	5.00	2.00	1.00
	0.50	0.00	147.58	146.80	0.14	0.06	4.09	4.09	3.80	3.29
	0.03	0.00	33.69	33.88	2.17	1.58	4.81	4.81	5.27	5.10
20x20	0.70	5.00	2.00	1.00	5.00	2.00	1.00	5.00	2.00	1.00
	0.70	0.00	27.66	27.85	4.87	1.96	1.95	4.88	1.95	1.95
	0.02	0.00	7.13	6.89	4.24	2.47	2.66	4.24	2.47	2.65
20x20	0.90	5.00	2.00	1.00	5.00	2.00	1.00	5.00	2.00	1.00
	0.90	2.72	2.13	2.65	7.18	0.77	1.13	7.18	0.77	1.13
	0.02	2.99	1.57	1.48	2.74	0.93	1.05	2.74	0.93	1.05
20x20	0.99	5.00	2.00	1.00	5.00	2.00	1.00	5.00	2.00	1.00
	0.99	5.47	1.73	1.03	5.53	1.71	1.01	5.53	1.71	1.01
	0.00	2.03	0.68	0.71	2.03	0.68	0.71	2.03	0.68	0.71
20x20	0.30	2.00	10.00	5.00	2.00	10.00	5.00	2.00	10.00	5.00
	0.30	0.00	6084.27	6088.59	9.79	49.55	61.11	9.79	49.48	61.04
	0.02	0.00	1257.21	1248.15	14.60	74.29	100.82	14.61	74.25	100.79
20x20	0.50	2.00	10.00	5.00	2.00	10.00	5.00	2.00	10.00	5.00
	0.50	0.00	1127.32	1116.88	7.80	28.36	20.06	7.81	28.33	20.03
	0.02	0.00	203.78	197.90	13.37	37.86	33.70	13.37	37.84	33.68
20x20	0.70	2.00	10.00	5.00	2.00	10.00	5.00	2.00	10.00	5.00
	0.70	0.00	208.73	210.11	12.63	9.61	11.54	12.64	9.59	11.52
	0.02	0.00	48.02	46.89	14.70	14.67	16.23	14.70	14.65	16.22
20x20	0.90	2.00	10.00	5.00	2.00	10.00	5.00	2.00	10.00	5.00
	0.90	1.18	13.47	16.8	16.74	3.35	5.25	16.75	3.35	5.24
	0.01	3.43	8.97	8.6	11.86	5.03	5.82	11.86	5.02	5.82
20x20	0.99	2.00	10.00	5.00	2.00	10.00	5.00	2.00	10.00	5.00
	0.99	7.08	7.97	4.77	7.56	7.83	4.64	7.58	7.82	4.64
	0.01	5.63	1.99	2.34	5.81	2.05	2.32	5.81	2.05	2.32

Note: The location points we chose are (1) artificial locations: 10x10 lattice, (2) the actual location points (N=312) in the Siuslaw Forest, or (3) artificial locations: 20x20 lattice. The first row in each lattice size and each p represent the actual values of parameters in the chosen spatial zero-inflated Poisson (SZIP) model. Methods (a), (b), and (c) respectively denote that the sample covariances are modified by using mean of non-zeros, overall mean, and individual means for each of the directional neighbors.

Table 6.2: The sample means and standard deviations (s.d.) of 100 estimated parameters from the 100 realizations of a SZINB model.

			(b) overall		
True	p	ζ	λ_s	λ_h	λ_v
mean	$m(\hat{p})$	$m(\hat{\zeta})$	$m(\hat{\lambda}_s)$	$m(\hat{\lambda}_h)$	$m(\hat{\lambda}_v)$
s.d.	$s(\hat{p})$	$s(\hat{\zeta})$	$s(\hat{\lambda}_s)$	$s(\hat{\lambda}_h)$	$s(\hat{\lambda}_v)$
10x10	0.30	0.10	0.10	1.30	0.80
	0.30	0.16	1.60	3.40	2.67
	0.05	0.20	1.99	8.20	4.40
10x10	0.50	0.10	0.10	1.30	0.80
	0.50	0.16	1.67	1.19	1.77
	0.05	0.19	1.98	2.06	2.57
10x10	0.70	0.10	0.10	1.30	0.80
	0.69	0.14	1.63	0.63	1.27
	0.05	0.17	1.74	1.01	1.51
10x10	0.90	0.10	0.10	1.30	0.80
	0.90	0.12	2.04	0.47	0.77
	0.03	0.14	1.52	0.54	0.67
10x10	0.99	0.10	0.10	1.30	0.80
	0.99	0.12	1.03	1.04	0.73
	0.01	0.14	1.11	0.51	0.51
10x10	0.30	0.01	2.00	10.00	5.00
	0.30	0.25	9.51	83.08	75.80
	0.05	0.28	14.43	136.44	137.72
10x10	0.50	0.01	2.00	10.00	5.00
	0.50	0.19	8.15	43.25	25.52
	0.05	0.21	13.88	53.61	44.36
10x10	0.70	0.01	2.00	10.00	5.00
	0.70	0.13	8.97	20.39	18.83
	0.04	0.14	14.07	29.32	24.20
10x10	0.90	0.01	2.00	10.00	5.00
	0.90	0.08	12.37	8.60	6.59
	0.03	0.12	12.95	11.78	10.16
10x10	0.99	0.01	2.00	10.00	5.00
	0.99	0.07	10.36	8.17	3.78
	0.01	0.11	8.72	4.42	3.77

Note: The location points we chose are (1) artificial locations: 10x10 lattice, (2) the actual location points (N=312) in the Siuslaw Forest, or (3) artificial locations: 20x20 lattice. The first row in each lattice size and each p represent the actual values of parameters in the chosen spatial zero-inflated Negative Binomial (SZINB) model with $\lambda_i = \alpha_i \zeta$.

(Continued)

			(b) overall		
True	p	ζ	λ_s	λ_h	λ_v
mean	$m(\hat{p})$	$m(\hat{\zeta})$	$m(\hat{\lambda}_s)$	$m(\hat{\lambda}_h)$	$m(\hat{\lambda}_v)$
s.d.	$s(\hat{p})$	$s(\hat{\zeta})$	$s(\hat{\lambda}_s)$	$s(\hat{\lambda}_h)$	$s(\hat{\lambda}_v)$
312	0.30	0.10	0.10	1.30	0.80
	0.30	0.12	1.52	1.59	2.15
	0.03	0.11	1.94	2.45	3.24
312	0.50	0.10	0.10	1.30	0.80
	0.49	0.12	1.85	0.91	1.20
	0.03	0.12	1.86	1.33	1.66
312	0.70	0.10	0.10	1.30	0.80
	0.69	0.11	1.97	0.51	0.82
	0.03	0.09	1.54	0.77	0.89
312	0.90	0.10	0.10	1.30	0.80
	0.90	0.11	2.71	0.22	0.57
	0.02	0.09	1.04	0.31	0.49
312	0.99	0.10	0.10	1.30	0.80
	0.99	0.10	1.99	0.60	0.55
	0.01	0.08	1.12	0.38	0.35
312	0.30	0.01	2.00	10.00	5.00
	0.30	0.10	9.62	61.58	71.06
	0.02	0.12	14.56	94.20	103.84
312	0.50	0.01	2.00	10.00	5.00
	0.50	0.07	7.95	25.54	35.45
	0.02	0.09	13.05	39.00	48.85
312	0.70	0.01	2.00	10.00	5.00
	0.70	0.05	10.55	11.61	14.83
	0.03	0.07	13.49	19.03	20.18
312	0.90	0.01	2.00	10.00	5.00
	0.90	0.03	17.68	3.34	5.24
	0.02	0.05	12.06	5.07	6.78
312	0.99	0.01	2.00	10.00	5.00
	0.99	0.04	16.18	5.08	2.92
	0.01	0.06	7.44	3.28	2.46

Note: The location points we chose are (1) artificial locations: 10x10 lattice, (2) the actual location points (N=312) in the Siuslaw Forest, or (3) artificial locations: 20x20 lattice. The first row in each lattice size and each p represent the actual values of parameters in the chosen spatial zero-inflated Negative Binomial (SZINB) model with $\lambda_i = \alpha_i \zeta$.

(Continued)

			(b) overall		
True	p	ζ	λ_s	λ_h	λ_v
mean	$m(\hat{p})$	$m(\hat{\zeta})$	$m(\hat{\lambda}_s)$	$m(\hat{\lambda}_h)$	$m(\hat{\lambda}_v)$
s.d.	$s(\hat{p})$	$s(\hat{\zeta})$	$s(\hat{\lambda}_s)$	$s(\hat{\lambda}_h)$	$s(\hat{\lambda}_v)$
20x20	0.30	0.10	0.10	1.30	0.80
	0.30	0.10	1.52	1.56	1.75
	0.02	0.11	1.89	2.51	2.46
20x20	0.50	0.10	0.10	1.30	0.80
	0.50	0.09	1.79	0.71	0.93
	0.02	0.09	1.73	1.08	1.20
20x20	0.70	0.10	0.10	1.30	0.80
	0.70	0.10	2.12	0.34	0.83
	0.02	0.09	1.49	0.47	0.79
20x20	0.90	0.10	0.10	1.30	0.80
	0.90	0.10	2.30	0.24	0.76
	0.02	0.08	1.03	0.27	0.44
20x20	0.99	0.10	0.10	1.30	0.80
	0.99	0.09	0.70	1.07	0.74
	0.01	0.07	0.58	0.26	0.23
20x20	0.30	0.01	2.00	10.00	5.00
	0.30	0.07	10.66	51.59	32.23
	0.03	0.10	14.58	72.34	68.09
20x20	0.50	0.01	2.00	10.00	5.00
	0.50	0.06	11.37	26.49	16.99
	0.03	0.08	14.66	36.49	31.26
20x20	0.70	0.01	2.00	10.00	5.00
	0.70	0.05	12.78	9.86	9.17
	0.02	0.07	13.58	16.62	13.37
20x20	0.90	0.01	2.00	10.00	5.00
	0.90	0.04	15.88	3.67	5.05
	0.02	0.06	11.16	5.09	4.77
20x20	0.99	0.01	2.00	10.00	5.00
	0.99	0.03	6.77	8.01	4.91
	0.01	0.05	5.49	2.31	2.31

Note: The location points we chose are (1) artificial locations: 10x10 lattice, (2) the actual location points (N=312) in the Siuslaw Forest, or (3) artificial locations: 20x20 lattice. The first row in each lattice size and each p represent the actual values of parameters in the chosen spatial zero-inflated Negative Binomial (SZINB) model with $\lambda_i = \alpha_i \zeta$.

Table 6.3: The sample means and standard deviations (s.d.) of 100 estimated parameters from the 100 realizations of a SZIE model.

		(b) overall		
True	p	β_s	β_h	β_v
mean	$m(\hat{p})$	$m(\hat{\beta}_s)$	$m(\hat{\beta}_h)$	$m(\hat{\beta}_v)$
s.d.	$s(\hat{p})$	$s(\hat{\beta}_s)$	$s(\hat{\beta}_h)$	$s(\hat{\beta}_v)$
10x10	0.30	5.00	2.00	1.00
	0.30	5.08	2.00	2.11
	0.05	4.97	2.84	3.00
10x10	0.50	5.00	2.00	1.00
	0.50	5.05	1.73	1.68
	0.04	4.29	2.02	2.18
10x10	0.70	5.00	2.00	1.00
	0.70	5.95	1.24	1.45
	0.04	3.88	1.44	1.66
10x10	0.90	5.00	2.00	1.00
	0.90	6.49	1.21	1.06
	0.03	3.30	1.19	1.20
10x10	0.99	5.00	2.00	1.00
	0.99	5.88	1.54	1.03
	0.01	2.41	0.95	1.04
10x10	0.30	16.77	88.84	88.84
	0.30	172.22	65.36	75.87
	0.05	169.87	95.13	102.88
10x10	0.50	16.77	88.84	88.84
	0.50	154.84	52.37	74.27
	0.04	150.58	72.20	77.74
10x10	0.70	16.77	88.84	88.84
	0.70	175.94	38.94	69.45
	0.04	136.60	52.16	62.45
10x10	0.90	16.77	88.84	88.84
	0.90	148.76	44.47	70.87
	0.03	109.78	40.80	41.18
10x10	0.99	16.77	88.84	88.84
	0.99	84.88	70.08	79.39
	0.01	71.49	33.13	30.07

Note: The location points we chose are (1) artificial locations: 10x10 lattice, (2) the actual location points (N=312) in the Siuslaw Forest, or (3) artificial locations: 20x20 lattice. The first row in each lattice size and each p represent the actual values of parameters in the chosen spatial zero-inflated Exponential (SZIE) model.

(Continued)

		(b) overall		
True	p	β_s	β_h	β_v
mean	$m(\hat{p})$	$m(\hat{\beta}_s)$	$m(\hat{\beta}_h)$	$m(\hat{\beta}_v)$
s.d.	$s(\hat{p})$	$s(\hat{\beta}_s)$	$s(\hat{\beta}_h)$	$s(\hat{\beta}_v)$
312	0.30	5.00	2.00	1.00
	0.30	4.90	2.31	1.81
	0.03	4.91	2.66	2.61
312	0.50	5.00	2.00	1.00
	0.50	5.39	1.49	1.55
	0.03	4.10	1.72	1.81
312	0.70	5.00	2.00	1.00
	0.70	6.29	0.99	1.38
	0.03	3.19	1.21	1.32
312	0.90	5.00	2.00	1.00
	0.90	7.40	0.83	0.96
	0.02	2.68	0.96	0.97
312	0.99	5.00	2.00	1.00
	0.99	7.01	1.18	0.82
	0.01	2.28	0.90	0.84
312	0.30	16.77	88.84	88.84
	0.30	150.87	77.62	73.58
	0.03	167.53	87.51	96.65
312	0.50	16.77	88.84	88.84
	0.50	151.36	48.00	71.19
	0.03	137.28	53.93	63.60
312	0.70	16.77	88.84	88.84
	0.70	173.25	29.84	70.73
	0.03	102.45	36.93	44.10
312	0.90	16.77	88.84	88.84
	0.90	192.57	25.15	64.70
	0.02	92.22	30.25	32.48
312	0.99	16.77	88.84	88.84
	0.99	131.31	51.52	69.05
	0.01	82.34	30.47	22.47

Note: The location points we chose are (1) artificial locations: 10x10 lattice, (2) the actual location points (N=312) in the Siuslaw Forest, or (3) artificial locations: 20x20 lattice. The first row in each lattice size and each p represent the actual values of parameters in the chosen spatial zero-inflated Exponential (SZIE) model.

(Continued)

		(b) overall		
True	p	β_s	β_h	β_v
mean	$m(\hat{p})$	$m(\hat{\beta}_s)$	$m(\hat{\beta}_h)$	$m(\hat{\beta}_v)$
s.d.	$s(\hat{p})$	$s(\hat{\beta}_s)$	$s(\hat{\beta}_h)$	$s(\hat{\beta}_v)$
20x20	0.30	5.00	2.00	1.00
	0.30	5.01	1.41	1.98
	0.02	4.32	1.98	2.19
20x20	0.50	5.00	2.00	1.00
	0.50	5.74	1.02	1.68
	0.03	3.43	1.31	1.55
20x20	0.70	5.00	2.00	1.00
	0.70	6.78	0.88	1.24
	0.02	2.83	1.05	1.19
20x20	0.90	5.00	2.00	1.00
	0.90	7.25	0.79	1.08
	0.01	2.23	0.86	0.90
20x20	0.99	5.00	2.00	1.00
	0.99	5.93	1.56	0.97
	0.00	2.15	0.70	0.73
20x20	0.30	16.77	88.84	88.84
	0.30	148.65	54.67	79.20
	0.02	136.06	66.92	80.87
20x20	0.50	16.77	88.84	88.84
	0.50	155.01	36.27	76.55
	0.03	126.62	46.50	60.27
20x20	0.70	16.77	88.84	88.84
	0.70	150.47	27.43	84.22
	0.02	94.69	34.65	36.82
20x20	0.90	16.77	88.84	88.84
	0.90	147.53	27.27	84.39
	0.01	72.57	26.91	21.30
20x20	0.99	16.77	88.84	88.84
	0.99	43.59	77.45	87.36
	0.00	37.11	15.91	12.74

Note: The location points we chose are (1) artificial locations: 10x10 lattice, (2) the actual location points (N=312) in the Siuslaw Forest, or (3) artificial locations: 20x20 lattice. The first row in each lattice size and each p represent the actual values of parameters in the chosen spatial zero-inflated Exponential (SZIE) model.

Table 6.4: The sample means and standard deviations (s.d.) of 100 estimated parameters from the 100 realizations of a SZIG model.

		$\hat{\gamma}$				$\hat{\gamma}_{alt}$			
True	p	γ	α_s	α_h	α_v	γ_{alt}	α_s	α_h	α_v
mean	$m(\hat{p})$	$m(\hat{\gamma})$	$m(\hat{\alpha}_s)$	$m(\hat{\alpha}_h)$	$m(\hat{\alpha}_v)$	$m(\hat{\gamma_{alt}})$	$m(\hat{\alpha}_s)$	$m(\hat{\alpha}_h)$	$m(\hat{\alpha}_v)$
s.d.	$s(\hat{p})$	$s(\hat{\gamma})$	$s(\hat{\alpha}_s)$	$s(\hat{\alpha}_h)$	$s(\hat{\alpha}_v)$	$s(\hat{\gamma_{alt}})$	$s(\hat{\alpha}_s)$	$s(\hat{\alpha}_h)$	$s(\hat{\alpha}_v)$
10x10	0.30	10.48	8.15	0.93	1.89	10.48	8.15	0.93	1.89
	0.30	10.09	5.85	4.71	4.66	10.24	5.87	4.64	4.60
	0.04	4.16	6.48	8.68	6.97	4.26	6.48	8.55	6.87
10x10	0.50	10.48	8.15	0.93	1.89	10.48	8.15	0.93	1.89
	0.50	10.70	7.29	1.75	2.68	10.75	7.31	1.74	2.67
	0.04	3.42	5.77	2.73	3.75	3.45	5.76	2.72	3.74
10x10	0.70	10.48	8.15	0.93	1.89	10.48	8.15	0.93	1.89
	0.69	10.63	7.53	1.04	2.30	10.64	7.53	1.04	2.30
	0.04	3.04	5.20	1.63	2.66	3.05	5.20	1.63	2.66
10x10	0.90	10.48	8.15	0.93	1.89	10.48	8.15	0.93	1.89
	0.90	10.79	8.74	0.73	1.80	10.78	8.74	0.73	1.80
	0.03	2.63	3.81	1.06	1.52	2.63	3.81	1.06	1.52
10x10	0.99	10.48	8.15	0.93	1.89	10.48	8.15	0.93	1.89
	0.99	10.76	7.80	1.11	1.96	10.75	7.80	1.11	1.96
	0.01	2.41	3.47	1.39	1.25	2.41	3.47	1.39	1.25
10x10	0.30	192.66	208.26	40.97	40.97	192.66	208.26	40.97	40.97
	0.30	173.94	180.02	92.00	122.24	175.69	180.30	91.35	121.21
	0.05	71.94	172.13	181.54	221.26	73.58	172.15	180.32	219.71
10x10	0.50	192.66	208.26	40.97	40.97	192.66	208.26	40.97	40.97
	0.49	182.71	165.69	72.72	84.82	182.77	165.73	72.82	84.83
	0.05	50.99	159.30	119.85	113.38	51.42	159.29	120.19	113.38
10x10	0.70	192.66	208.26	40.97	40.97	192.66	208.26	40.97	40.97
	0.70	183.83	211.83	34.13	54.68	183.50	211.62	34.22	54.79
	0.04	43.99	137.57	53.41	77.18	44.13	137.66	53.56	77.32
10x10	0.90	192.66	208.26	40.97	40.97	192.66	208.26	40.97	40.97
	0.90	187.05	231.09	32.70	38.98	186.89	230.97	32.73	39.02
	0.03	38.68	105.88	33.23	43.79	38.72	105.92	33.26	43.83
10x10	0.99	192.66	208.26	40.97	40.97	192.66	208.26	40.97	40.97
	0.99	190.87	208.32	42.13	41.13	190.85	208.30	42.14	41.14
	0.01	40.15	98.99	36.63	40.73	40.16	99.00	36.63	40.73

Note: The location points we chose are (1) artificial locations: 10x10 lattice, (2) the actual location points (N=312) in the Siuslaw Forest, or (3) artificial locations: 20x20 lattice. The first row in each lattice size and each p represent the actual values of parameters in the chosen spatial zero-inflated Gamma (SZIG) model.

(Continued)

		$\hat{\gamma}$				$\hat{\gamma}_{alt}$			
True	p	γ	α_s	α_h	α_v	γ_{alt}	α_s	α_h	α_v
mean	$m(\hat{p})$	$m(\hat{\gamma})$	$m(\hat{\alpha}_s)$	$m(\hat{\alpha}_h)$	$m(\hat{\alpha}_v)$	$m(\hat{\gamma}_{alt})$	$m(\hat{\alpha}_s)$	$m(\hat{\alpha}_h)$	$m(\hat{\alpha}_v)$
s.d.	$s(\hat{p})$	$s(\hat{\gamma})$	$s(\hat{\alpha}_s)$	$s(\hat{\alpha}_h)$	$s(\hat{\alpha}_v)$	$s(\hat{\gamma}_{alt})$	$s(\hat{\alpha}_s)$	$s(\hat{\alpha}_h)$	$s(\hat{\alpha}_v)$
312	0.30	10.48	8.15	0.93	1.89	10.48	8.15	0.93	1.89
	0.30	10.07	7.27	2.46	3.47	10.11	7.28	2.45	3.45
	0.03	2.27	6.23	4.35	5.49	2.28	6.22	4.33	5.47
312	0.50	10.48	8.15	0.93	1.89	10.48	8.15	0.93	1.89
	0.50	10.14	7.72	1.16	2.08	10.15	7.73	1.16	2.08
	0.03	1.84	4.86	2.11	2.22	1.85	4.86	2.11	2.21
312	0.70	10.48	8.15	0.93	1.89	10.48	8.15	0.93	1.89
	0.70	10.04	8.83	0.77	1.70	10.04	8.83	0.77	1.70
	0.02	1.45	4.07	1.31	1.79	1.45	4.07	1.31	1.79
312	0.90	10.48	8.15	0.93	1.89	10.48	8.15	0.93	1.89
	0.90	10.39	10.94	0.43	0.98	10.39	10.94	0.43	0.98
	0.02	1.38	2.57	0.60	1.02	1.38	2.57	0.60	1.02
312	0.99	10.48	8.15	0.93	1.89	10.48	8.15	0.93	1.89
	0.99	10.29	9.60	0.76	1.31	10.29	9.60	0.76	1.31
	0.01	1.19	2.41	0.81	0.95	1.19	2.41	0.81	0.95
312	0.30	192.66	208.26	40.97	40.97	192.66	208.26	40.97	40.97
	0.30	187.96	162.41	92.07	78.00	188.56	162.59	91.81	77.76
	0.03	35.26	165.44	132.83	124.79	35.53	165.40	132.44	124.41
312	0.50	192.66	208.26	40.97	40.97	192.66	208.26	40.97	40.97
	0.50	190.91	191.91	45.08	55.55	190.93	191.91	45.08	55.54
	0.03	29.06	139.11	69.59	69.48	29.14	139.11	69.59	69.46
312	0.70	192.66	208.26	40.97	40.97	192.66	208.26	40.97	40.97
	0.70	191.35	267.19	18.28	32.67	191.26	267.14	18.29	32.69
	0.02	25.53	98.28	29.08	43.35	25.56	98.31	29.10	43.37
312	0.90	192.66	208.26	40.97	40.97	192.66	208.26	40.97	40.97
	0.90	191.07	290.16	9.79	29.35	191.02	290.14	9.79	29.36
	0.02	19.48	68.12	16.13	27.78	19.48	68.13	16.14	27.78
312	0.99	192.66	208.26	40.97	40.97	192.66	208.26	40.97	40.97
	0.99	191.97	277.68	18.71	26.4	191.97	277.68	18.71	26.40
	0.01	19.73	68.25	20.43	25.5	19.73	68.25	20.43	25.50

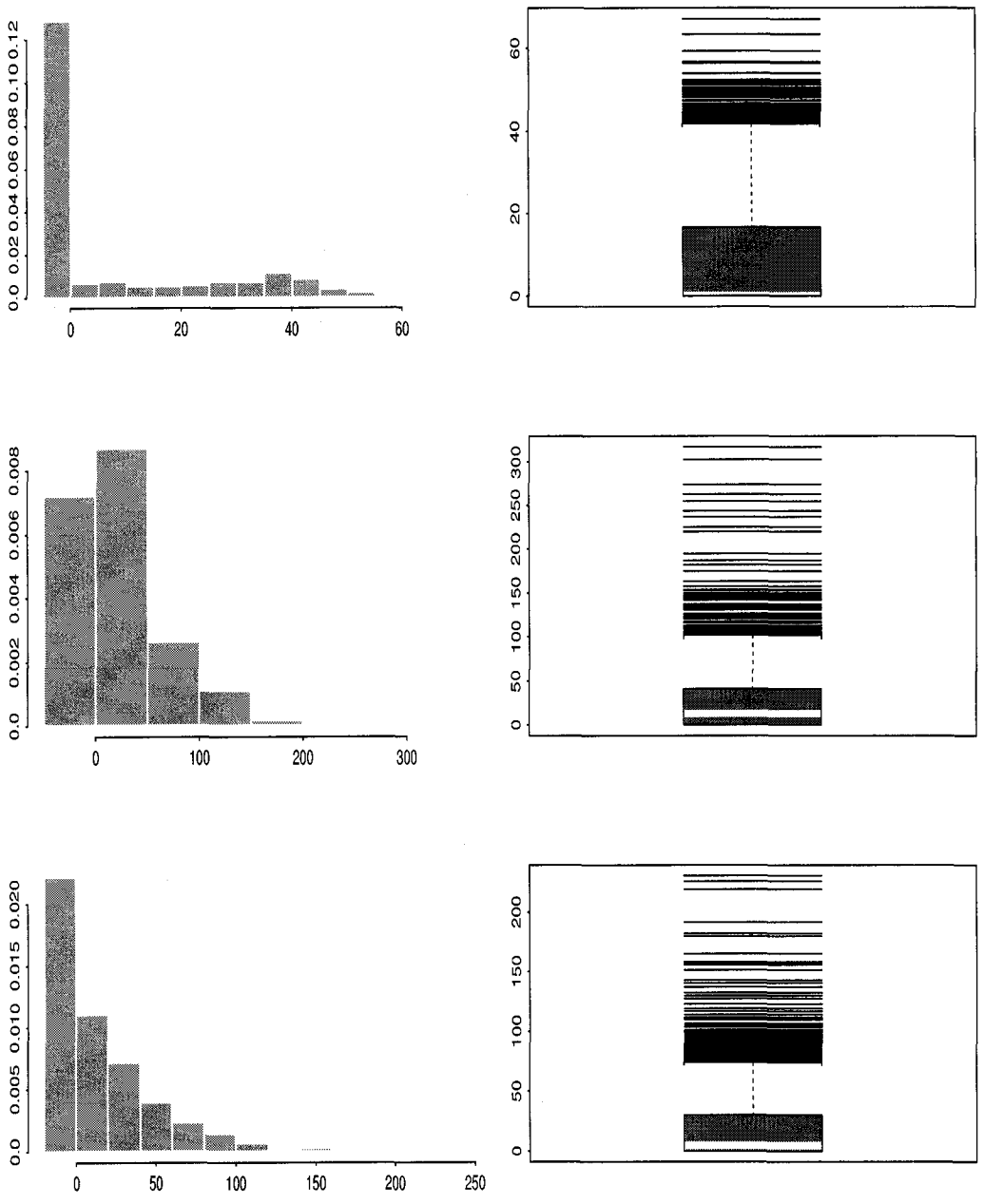
Note: The location points we chose are (1) artificial locations: 10x10 lattice, (2) the actual location points (N=312) in the Siuslaw Forest, or (3) artificial locations: 20x20 lattice. The first row in each lattice size and each p represent the actual values of parameters in the chosen spatial zero-inflated Gamma (SZIG) model.

(Continued)

		$\hat{\gamma}$				$\hat{\gamma}_{alt}$			
True	p	γ	α_s	α_h	α_v	γ_{alt}	α_s	α_h	α_v
mean	$m(\hat{p})$	$m(\hat{\gamma})$	$m(\hat{\alpha}_s)$	$m(\hat{\alpha}_h)$	$m(\hat{\alpha}_v)$	$m(\hat{\gamma_{alt}})$	$m(\hat{\alpha}_s)$	$m(\hat{\alpha}_h)$	$m(\hat{\alpha}_v)$
s.d.	$s(\hat{p})$	$s(\hat{\gamma})$	$s(\hat{\alpha}_s)$	$s(\hat{\alpha}_h)$	$s(\hat{\alpha}_v)$	$s(\hat{\gamma_{alt}})$	$s(\hat{\alpha}_s)$	$s(\hat{\alpha}_h)$	$s(\hat{\alpha}_v)$
20x20	0.30	10.48	8.15	0.93	1.89	10.48	8.15	0.93	1.89
	0.30	9.84	7.19	2.12	2.33	9.88	7.20	2.11	2.32
	0.02	1.82	5.70	3.51	3.05	1.83	5.69	3.50	3.04
20x20	0.50	10.48	8.15	0.93	1.89	10.48	8.15	0.93	1.89
	0.50	10.26	8.06	0.78	2.04	10.27	8.06	0.78	2.04
	0.02	1.60	4.51	1.40	2.08	1.61	4.51	1.40	2.08
20x20	0.70	10.48	8.15	0.93	1.89	10.48	8.15	0.93	1.89
	0.70	10.33	8.39	0.62	2.09	10.33	8.39	0.62	2.09
	0.02	1.27	3.51	0.79	1.55	1.27	3.51	0.79	1.55
20x20	0.90	10.48	8.15	0.93	1.89	10.48	8.15	0.93	1.89
	0.90	10.32	9.36	0.36	1.84	10.32	9.36	0.36	1.84
	0.01	1.23	2.32	0.50	0.98	1.23	2.32	0.50	0.98
20x20	0.99	10.48	8.15	0.93	1.89	10.48	8.15	0.93	1.89
	0.99	10.39	8.79	0.75	1.74	10.39	8.79	0.75	1.74
	0.00	1.18	2.04	0.63	0.79	1.18	2.04	0.63	0.79
20x20	0.30	192.66	208.26	40.97	40.97	192.66	208.26	40.97	40.97
	0.30	191.58	166.35	61.68	86.86	192.05	166.56	61.52	86.65
	0.02	37.20	154.17	93.52	120.92	37.41	154.11	93.28	120.60
20x20	0.50	192.66	208.26	40.97	40.97	192.66	208.26	40.97	40.97
	0.50	194.34	228.07	23.24	54.22	194.36	228.08	23.23	54.21
	0.03	27.02	132.84	43.39	58.12	27.07	132.84	43.39	58.11
20x20	0.70	192.66	208.26	40.97	40.97	192.66	208.26	40.97	40.97
	0.70	194.51	250.93	16.29	45.03	194.43	250.88	16.29	45.04
	0.02	22.73	80.83	22.33	36.87	22.75	80.86	22.347	36.89
20x20	0.90	192.66	208.26	40.97	40.97	192.66	208.26	40.97	40.97
	0.90	193.18	268.58	11.84	40.34	193.14	268.56	11.84	40.35
	0.02	20.67	61.18	18.17	26.89	20.67	61.19	18.17	26.89
20x20	0.99	192.66	208.26	40.97	40.97	192.66	208.26	40.97	40.97
	0.99	191.40	231.57	31.27	39.45	191.39	231.56	31.27	39.45
	0.00	19.89	53.69	23.05	21.18	19.89	53.70	23.05	21.18

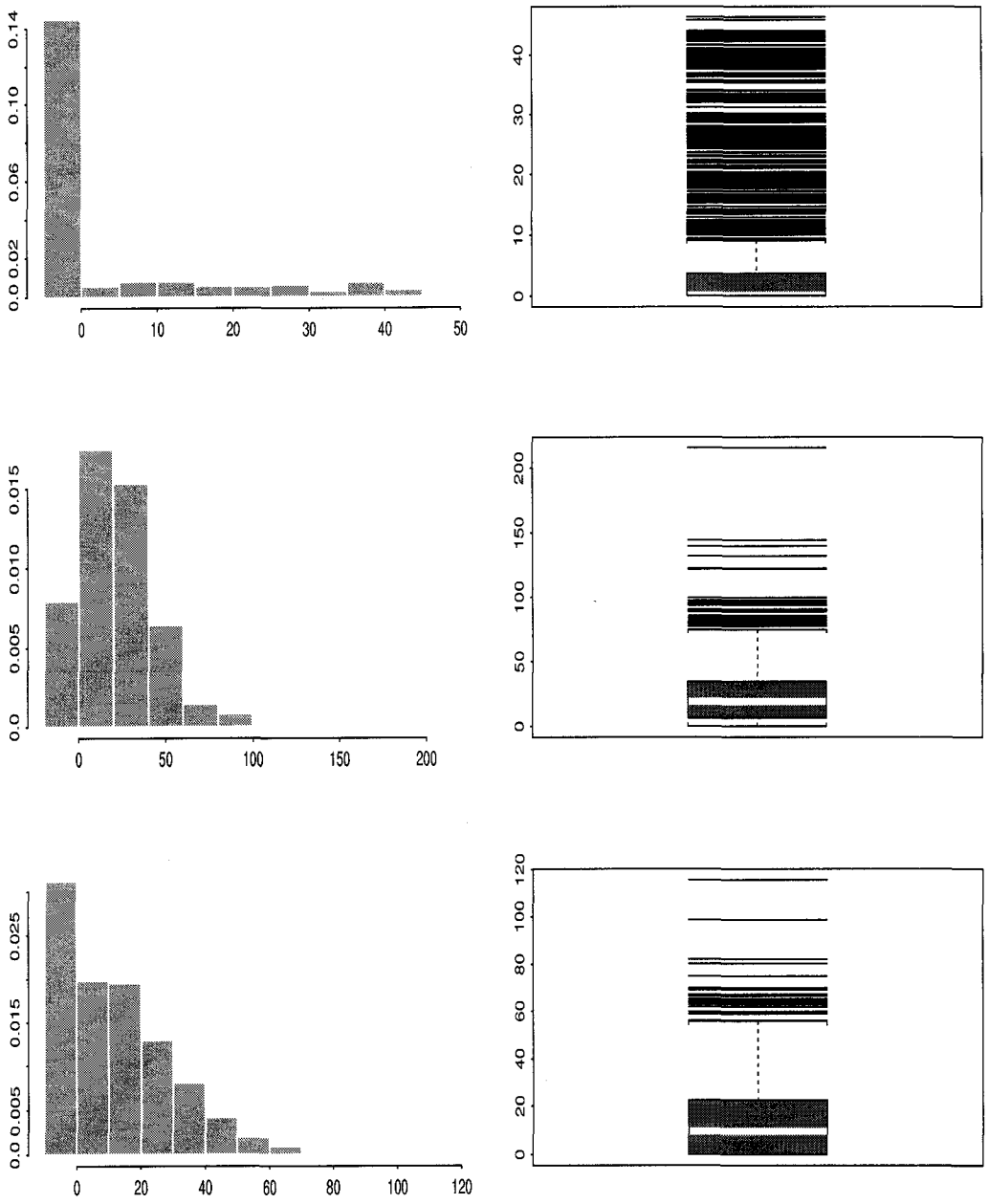
Note: The location points we chose are (1) artificial locations: 10x10 lattice, (2) the actual location points (N=312) in the Siuslaw Forest, or (3) artificial locations: 20x20 lattice. The first row in each lattice size and each p represent the actual values of parameters in the chosen spatial zero-inflated Gamma (SZIG) model.

Figure 6.2: Histograms and boxplots for 1,000 estimated λ_s, λ_h, and λ_v for 10x10 lattice realization with $p = 0.3$, $\lambda_s = 5$, $\lambda_h = 2$, and $\lambda_v = 1$



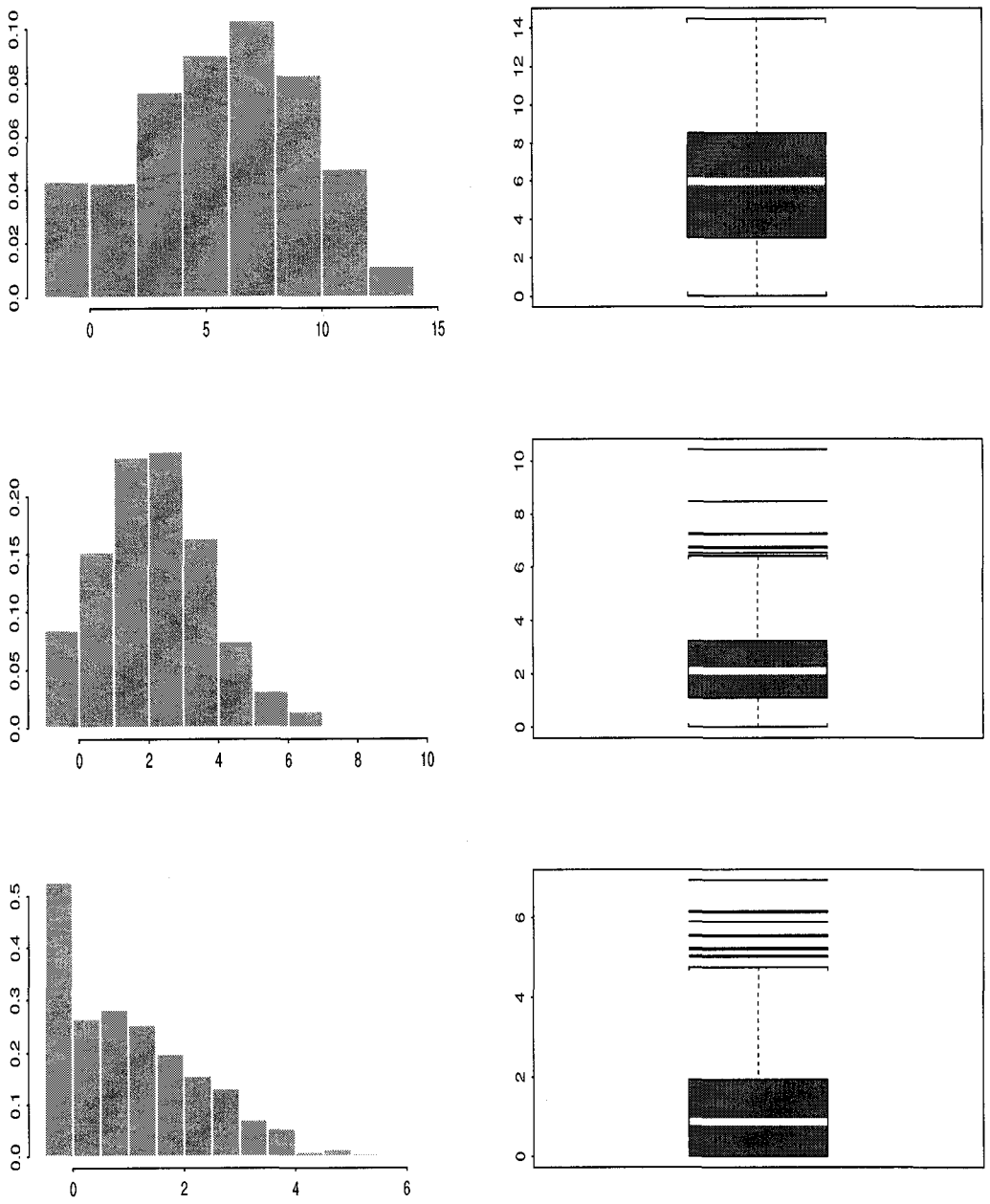
(a) the first row diplays the histogram and boxplot for 1,000 estimated λ_s, (b) the second row diplays the histogram and boxplot for 1,000 estimated λ_h, and (c) the third row diplays the histogram and boxplot for 1,000 estimated λ_v.

Figure 6.3: Histograms and boxplots for 1,000 estimated λ_s, λ_h, and λ_v for 20x20 lattice realization with $p = 0.3$, $\lambda_s = 5$, $\lambda_h = 2$, and $\lambda_v = 1$



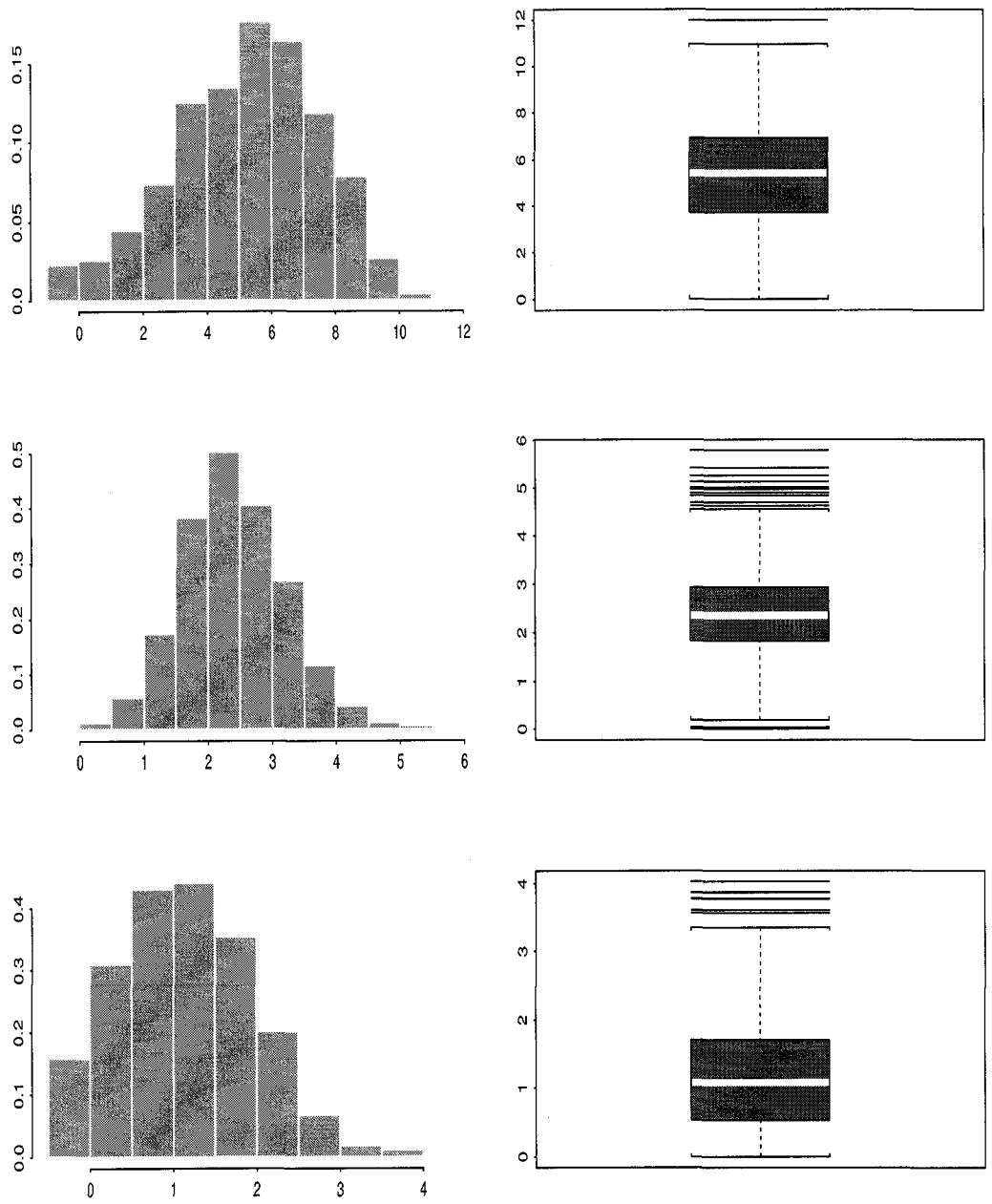
(a) the first row diplays the histogram and boxplot for 1,000 estimated λ_s, (b) the second row diplays the histogram and boxplot for 1,000 estimated λ_h, and (c) the third row diplays the histogram and boxplot for 1,000 estimated λ_v.

Figure 6.4: Histograms and boxplots for 1,000 estimated λ_s, λ_h, and λ_v for 10x10 lattice realization with $p = 0.9$, $\lambda_s = 5$, $\lambda_h = 2$, and $\lambda_v = 1$



(a) the first row diplays the histogram and boxplot for 1,000 estimated λ_s, (b) the second row diplays the histogram and boxplot for 1,000 estimated λ_h, and (c) the third row diplays the histogram and boxplot for 1,000 estimated λ_v.

Figure 6.5: Histograms and boxplots for 1,000 estimated λ_s, λ_h, and λ_v for 20x20 lattice realization with $p = 0.9$, $\lambda_s = 5$, $\lambda_h = 2$, and $\lambda_v = 1$



(a) the first row diplays the histogram and boxplot for 1,000 estimated λ_s, (b) the second row diplays the histogram and boxplot for 1,000 estimated λ_h, and (c) the third row diplays the histogram and boxplot for 1,000 estimated λ_v.

Table 6.5: The three moment estimations for the SZIP model applied to real data

		(a)			(b)			(c)		
	$\hat{p}$	$\hat{\lambda}_s$	$\hat{\lambda}_{hv}$	$\hat{\rho}_{hv}$	$\hat{\lambda}_s$	$\hat{\lambda}_{hv}$	$\hat{\rho}_{hv}$	$\hat{\lambda}_s$	$\hat{\lambda}_{hv}$	$\hat{\rho}_{hv}$
	313 plots									
M	0.96	0.00	757.76	0.00	0.00	747.56	0.00	0.00	746.62	0.00
M_s	0.14	0.00	301502.64	0.00	0.00	2817.51	0.00	0.00	2816.17	0.00
Lt	0.99	0.00	7876.95	0.00	0.00	7889.82	0.00	0.00	7875.80	0.00
Lt_s	0.56	0.00	190528.70	0.00	0.00	108290.12	0.00	0.00	108211.21	0.00
Tba	0.99	0.00	5.61	0.08	0.00	5.05	0.07	0.00	5.01	0.07
Tba_s	0.42	0.00	0.96	0.02	0.00	0.40	0.03	0.00	0.40	0.03
	312 plots									
M	0.96	0.00	759.06	0.00	0.00	748.05	0.00	0.00	746.80	0.00
M_s	0.14	0.00	299579.18	0.00	0.00	2800.28	0.00	0.00	2798.21	0.00
Lt	0.99	0.00	7877.10	0.00	0.00	7892.68	0.00	0.00	7876.45	0.00
Lt_s	0.56	0.00	188715.94	0.00	0.00	108781.19	0.00	0.00	108753.19	0.00
Tba	0.99	0.00	5.57	0.08	0.00	5.04	0.07	0.00	5.02	0.07
Tba_s	0.42	0.00	0.94	0.02	0.00	0.40	0.03	0.00	0.40	0.03

(1a) assume that $\lambda_h = \lambda_v = \lambda_{hv}$ and $\lambda_{c(1)} = \lambda_{c(2)} = 0$

	$\hat{\lambda}_s$	$\hat{\lambda}_h$	$\hat{\lambda}_v$	$\hat{\rho}_h$	$\hat{\rho}_v$	$\hat{\lambda}_s$	$\hat{\lambda}_h$	$\hat{\lambda}_v$	$\hat{\rho}_h$	$\hat{\rho}_v$
	313 plots									
	(a)					(b)				
M	0.00	672.72	845.79	0.00	0.00	0.00	664.27	833.77	0.00	0.00
M_s	0.00	298166.07	304955.99	0.00	0.00	0.00	3245.39	2374.66	0.00	0.00
Lt	0.00	7032.29	8751.18	0.00	0.00	0.00	7008.51	8801.98	0.00	0.00
Lt_s	0.00	173104.34	208562.92	0.00	0.00	0.00	93811.56	123275.43	0.00	0.00
Tba	0.00	0.00	20.74	0.00	0.30	0.00	0.00	19.48	0.00	0.28
Tba_s	0.00	0.87	1.04	0.02	0.03	0.00	0.33	0.47	0.03	0.04
	313 plots (c)					312 plots (c)				
M	0.00	663.91	831.94	0.00	0.00	0.00	664.07	832.13	0.00	0.00
M_s	0.00	3240.13	2352.29	0.00	0.00	0.00	3219.46	2337.28	0.00	0.00
Lt	0.00	7008.51	8743.00	0.00	0.00	0.00	7009.09	8743.73	0.00	0.00
Lt_s	0.00	93798.71	123069.04	0.00	0.00	0.00	94268.50	123685.44	0.00	0.00
Tba	0.00	0.00	18.81	0.00	0.27	0.00	0.00	18.81	0.00	0.27
Tba_s	0.00	0.33	0.47	0.03	0.04	0.00	0.33	0.47	0.03	0.04
	312 plots									
	(a)					(b)				
M	0.00	673.87	847.23	0.00	0.00	0.00	664.63	834.39	0.00	0.00
M_s	0.00	296263.90	303010.50	0.00	0.00	0.00	3223.46	2362.28	0.00	0.00
Lt	0.00	7030.30	8753.53	0.00	0.00	0.00	7009.14	8807.14	0.00	0.00
Lt_s	0.00	171255.83	206787.15	0.00	0.00	0.00	94268.50	123801.82	0.00	0.00
Tba	0.00	0.00	20.67	0.00	0.30	0.00	0.00	19.44	0.00	0.28
Tba_s	0.00	0.86	1.03	0.02	0.03	0.00	0.33	0.47	0.03	0.04

(1b) assume that $\lambda_{c(1)} = \lambda_{c(2)} = 0$

(Continued)

		(a)			(b)			(c)	
	$\hat{p}$	$\hat{\lambda}_s$	$\hat{\lambda}_{hvc}$	$\hat{\rho}_{hvc}$	$\hat{\lambda}_s$	$\hat{\lambda}_{hvc}$	$\hat{\rho}_{hvc}$	$\hat{\lambda}_s$	$\hat{\lambda}_{hvc}$	$\hat{\rho}_{hvc}$
	313 plots									
M	0.96	0.00	507.12	0.00	0.00	494.93	0.00	0.00	493.01	0.00
M_s	0.14	0.00	302951.83	0.00	0.00	4352.83	0.00	0.00	4351.67	0.00
Lt	0.99	0.00	5654.40	0.00	0.00	5624.15	0.00	0.00	5623.62	0.00
Lt_s	0.56	0.00	184045.38	0.00	0.00	97452.65	0.00	0.00	97173.30	0.00
Tba	0.99	0.00	46.49	0.02	0.00	45.93	0.02	0.00	45.89	0.02
Tba_s	0.42	0.00	0.91	0.01	0.00	0.33	0.01	0.00	0.33	0.01
	312 plots									
M	0.96	0.00	508.49	0.00	0.00	495.49	0.00	0.00	493.12	0.00
M_s	0.14	0.00	301019.13	0.00	0.00	4325.76	0.00	0.00	4323.91	0.00
Lt	0.99	0.00	5651.93	0.00	0.00	5624.28	0.00	0.00	5624.09	0.00
Lt_s	0.56	0.00	182123.98	0.00	0.00	97832.69	0.00	0.00	97660.00	0.00
Tba	0.99	0.00	46.45	0.02	0.00	45.93	0.02	0.00	45.90	0.02
Tba_s	0.42	0.00	0.90	0.01	0.00	0.33	0.01	0.00	0.33	0.01

(2a) assume that $\lambda_h = \lambda_v = \lambda_{c^{(1)}} = \lambda_{c^{(2)}} = \lambda_{hvc}$

	$\hat{\lambda}_s$	$\hat{\lambda}_{hv}$	$\hat{\lambda}_c$	$\hat{\rho}_{hv}$	$\hat{\rho}_c$	$\hat{\lambda}_s$	$\hat{\lambda}_{hv}$	$\hat{\lambda}_c$	$\hat{\rho}_{hv}$	$\hat{\rho}_c$
	313 plots									
	(a)					(b)				
M	0.00	757.76	235.81	0.00	0.00	0.00	747.56	221.48	0.00	0.00
M_s	0.00	301502.64	304520.50	0.00	0.00	0.00	2817.51	6014.73	0.00	0.00
Lt	0.00	7876.95	3248.60	0.00	0.00	0.00	7889.82	3171.67	0.00	0.00
Lt_s	0.00	190528.70	177027.53	0.00	0.00	0.00	108290.12	85721.67	0.00	0.00
Tba	0.00	5.61	90.74	0.00	0.04	0.00	5.05	90.18	0.00	0.04
Tba_s	0.00	0.96	0.86	0.01	0.01	0.00	0.40	0.25	0.02	0.01
	313 plots (c)					312 plots (c)				
M	0.00	746.62	218.09	0.00	0.00	0.00	746.80	218.14	0.00	0.00
M_s	0.00	2816.17	6013.75	0.00	0.00	0.00	2798.21	5975.38	0.00	0.00
Lt	0.00	7875.80	3140.67	0.00	0.00	0.00	7876.45	3140.93	0.00	0.00
Lt_s	0.00	108211.21	85087.14	0.00	0.00	0.00	108753.19	85513.30	0.00	0.00
Tba	0.00	5.01	90.14	0.00	0.04	0.00	5.02	90.15	0.00	0.04
Tba_s	0.00	0.40	0.25	0.02	0.01	0.00	0.40	0.25	0.02	0.01
	312 plots									
	(a)					(b)				
M	0.00	759.06	237.26	0.00	0.00	0.00	748.05	222.11	0.00	0.00
M_s	0.00	299579.18	302577.79	0.00	0.00	0.00	2800.28	5977.01	0.00	0.00
Lt	0.00	7877.10	3243.31	0.00	0.00	0.00	7892.68	3168.87	0.00	0.00
Lt_s	0.00	188715.94	174988.55	0.00	0.00	0.00	108781.19	85981.53	0.00	0.00
Tba	0.00	5.57	90.70	0.00	0.04	0.00	5.04	90.18	0.00	0.04
Tba_s	0.00	0.94	0.85	0.01	0.01	0.00	0.40	0.26	0.02	0.01

(2b) assume that $\lambda_h = \lambda_v = \lambda_{hv}$ and $\lambda_{c^{(1)}} = \lambda_{c^{(2)}} = \lambda_c$

(Continued)

	$\hat{p}$	$\hat{\lambda}_s$	$\hat{\lambda}_h$	$\hat{\lambda}_v$	$\hat{\lambda}_c$	$\hat{\rho}_h$	$\hat{\rho}_v$	$\hat{\rho}_c$
	313 plots							
		(a)						
M	0.96	0.00	672.72	845.79	235.81	0.00	0.00	0.00
M_s	0.14	0.00	298166.07	304955.99	304520.50	0.00	0.00	0.00
Lt	0.99	0.00	7032.29	8751.18	3248.60	0.00	0.00	0.00
Lt_s	0.56	0.00	173104.34	208562.92	177027.53	0.00	0.00	0.00
Tba	0.99	0.00	0.00	20.74	90.74	0.00	0.01	0.04
Tba_s	0.42	0.00	0.87	1.04	0.86	0.01	0.01	0.01
		(b)						
M	0.96	0.00	664.27	833.77	221.48	0.00	0.00	0.00
M_s	0.14	0.00	3245.39	2374.66	6014.73	0.00	0.00	0.00
Lt	0.99	0.00	7008.51	8801.98	3171.67	0.00	0.00	0.00
Lt_s	0.56	0.00	93811.56	123275.43	85721.67	0.00	0.00	0.00
Tba	0.99	0.00	0.00	19.48	90.18	0.00	0.01	0.04
Tba_s	0.42	0.00	0.33	0.47	0.25	0.01	0.02	0.01
		(c)						
M	0.96	0.00	663.91	831.94	218.09	0.00	0.00	0.00
M_s	0.14	0.00	3240.13	2352.29	6013.75	0.00	0.00	0.00
Lt	0.99	0.00	7008.51	8743.00	3140.67	0.00	0.00	0.00
Lt_s	0.56	0.00	93798.71	123069.04	85087.14	0.00	0.00	0.00
Tba	0.99	0.00	0.00	18.81	90.14	0.00	0.01	0.04
Tba_s	0.42	0.00	0.33	0.47	0.25	0.01	0.02	0.01
	312 plots							
		(a)						
M	0.96	0.00	673.87	847.23	237.26	0.00	0.00	0.00
M_s	0.14	0.00	296263.90	303010.50	302577.79	0.00	0.00	0.00
Lt	0.99	0.00	7030.30	8753.53	3243.31	0.00	0.00	0.00
Lt_s	0.56	0.00	171255.83	206787.15	174988.55	0.00	0.00	0.00
Tba	0.99	0.00	0.00	20.67	90.70	0.00	0.01	0.04
Tba_s	0.42	0.00	0.86	1.03	0.85	0.01	0.01	0.01
		(b)						
M	0.96	0.00	664.63	834.39	222.11	0.00	0.00	0.00
M_s	0.14	0.00	3223.46	2362.28	5977.01	0.00	0.00	0.00
Lt	0.99	0.00	7009.14	8807.14	3168.87	0.00	0.00	0.00
Lt_s	0.56	0.00	94268.50	123801.82	85981.53	0.00	0.00	0.00
Tba	0.99	0.00	0.00	19.44	90.18	0.00	0.01	0.04
Tba_s	0.42	0.00	0.33	0.47	0.26	0.01	0.02	0.01
		(c)						
M	0.96	0.00	664.07	832.13	218.14	0.00	0.00	0.00
M_s	0.14	0.00	3219.46	2337.28	5975.38	0.00	0.00	0.00
Lt	0.99	0.00	7009.09	8743.73	3140.93	0.00	0.00	0.00
Lt_s	0.56	0.00	94268.50	123685.44	85513.30	0.00	0.00	0.00
Tba	0.99	0.00	0.00	18.81	90.15	0.00	0.01	0.04
Tba_s	0.42	0.00	0.33	0.47	0.25	0.01	0.02	0.01

(2c) assume that $\lambda_{c(1)} = \lambda_{c(2)} = \lambda_c$

(Continued)

	$\hat{p}$	$\hat{\lambda}_s$	$\hat{\lambda}_{hv}$	$\hat{\lambda}_{c(1)}$	$\hat{\lambda}_{c(2)}$	$\hat{\rho}_{hv}$	$\hat{\rho}_{c(1)}$	$\hat{\rho}_{c(2)}$
	313 plots							
		(a)						
M	0.96	0.00	757.76	595.36	0.00	0.00	0.00	0.00
M_s	0.14	0.00	301502.64	297559.03	311408.30	0.00	0.00	0.00
Lt	0.99	0.00	7876.95	5434.60	1085.72	0.00	0.00	0.00
Lt_s	0.56	0.00	190528.70	247948.69	106856.87	0.00	0.00	0.00
Tba	0.99	0.00	5.61	110.29	71.38	0.00	0.05	0.03
Tba_s	0.42	0.00	0.96	1.10	0.62	0.01	0.01	0.01
		(b)						
M	0.96	0.00	747.56	574.98	0.00	0.00	0.00	0.00
M_s	0.14	0.00	2817.51	1901.72	10084.23	0.00	0.00	0.00
Lt	0.99	0.00	7889.82	5379.62	987.09	0.00	0.00	0.00
Lt_s	0.56	0.00	108290.12	161436.38	10808.16	0.00	0.00	0.00
Tba	0.99	0.00	5.05	109.50	71.07	0.00	0.05	0.03
Tba_s	0.42	0.00	0.40	0.52	0.00	0.02	0.02	0.00
		(c)						
M	0.96	0.00	746.62	565.25	0.00	0.00	0.00	0.00
M_s	0.14	0.00	2816.17	1899.10	10071.43	0.00	0.00	0.00
Lt	0.99	0.00	7875.80	5368.78	925.94	0.00	0.00	0.00
Lt_s	0.56	0.00	108211.21	161161.85	9671.17	0.00	0.00	0.00
Tba	0.99	0.00	5.01	109.34	71.07	0.00	0.05	0.03
Tba_s	0.42	0.00	0.40	0.52	0.00	0.02	0.02	0.00
	312 plots							
		(a)						
M	0.96	0.00	759.06	597.32	0.00	0.00	0.00	0.00
M_s	0.14	0.00	299579.18	295660.73	309421.65	0.00	0.00	0.00
Lt	0.99	0.00	7877.10	5430.73	1079.04	0.00	0.00	0.00
Lt_s	0.56	0.00	188715.94	246348.76	104383.47	0.00	0.00	0.00
Tba	0.99	0.00	5.57	110.25	71.36	0.00	0.05	0.03
Tba_s	0.42	0.00	0.94	1.09	0.61	0.01	0.01	0.01
		(b)						
M	0.96	0.00	748.05	576.07	0.00	0.00	0.00	0.00
M_s	0.14	0.00	2800.28	1888.75	10022.01	0.00	0.00	0.00
Lt	0.99	0.00	7892.68	5378.28	982.83	0.00	0.00	0.00
Lt_s	0.56	0.00	108781.19	162137.94	10631.01	0.00	0.00	0.00
Tba	0.99	0.00	5.04	109.49	71.08	0.00	0.05	0.03
Tba_s	0.42	0.00	0.40	0.53	0.00	0.02	0.02	0.00
		(c)						
M	0.96	0.00	746.80	565.39	0.00	0.00	0.00	0.00
M_s	0.14	0.00	2798.21	1886.98	10007.18	0.00	0.00	0.00
Lt	0.99	0.00	7876.45	5369.23	926.01	0.00	0.00	0.00
Lt_s	0.56	0.00	108753.19	161969.03	9719.61	0.00	0.00	0.00
Tba	0.99	0.00	5.02	109.35	71.07	0.00	0.05	0.03
Tba_s	0.42	0.00	0.40	0.53	0.00	0.02	0.02	0.00

(2d) assume that $\lambda_h = \lambda_v = \lambda_{hv}$

(Continued)

	$\hat{\lambda}_s$	$\hat{\lambda}_h$	$\hat{\lambda}_v$	$\hat{\lambda}_{c(1)}$	$\hat{\lambda}_{c(2)}$	$\hat{\rho}_h$	$\hat{\rho}_v$	$\hat{\rho}_{c(1)}$	$\hat{\rho}_{c(2)}$
	313 plots								
	(a)								
M	0.00	672.72	845.79	595.36	0.00	0.00	0.00	0.00	0.00
M_s	0.00	298166.07	304955.99	297559.03	311408.30	0.00	0.00	0.00	0.00
Lt	0.00	7032.29	8751.18	5434.60	1085.72	0.00	0.00	0.00	0.00
Lt_s	0.00	173104.34	208562.92	247948.69	106856.87	0.00	0.00	0.00	0.00
Tba	0.00	0.00	20.74	110.29	71.38	0.00	0.01	0.05	0.03
Tba_s	0.00	0.87	1.04	1.10	0.62	0.01	0.01	0.01	0.01
	(b)								
M	0.00	664.27	833.77	574.98	0.00	0.00	0.00	0.00	0.00
M_s	0.00	3245.39	2374.66	1901.72	10084.23	0.00	0.00	0.00	0.00
Lt	0.00	7008.51	8801.98	5379.62	987.09	0.00	0.00	0.00	0.00
Lt_s	0.00	93811.56	123275.43	161436.38	10808.16	0.00	0.00	0.00	0.00
Tba	0.00	0.00	19.48	109.50	71.07	0.00	0.01	0.05	0.03
Tba_s	0.00	0.33	0.47	0.52	0.00	0.02	0.02	0.02	0.00
	(c)								
M	0.00	663.91	831.94	565.25	0.00	0.00	0.00	0.00	0.00
M_s	0.00	3240.13	2352.29	1899.10	10071.43	0.00	0.00	0.00	0.00
Lt	0.00	7008.51	8743.00	5368.78	925.94	0.00	0.00	0.00	0.00
Lt_s	0.00	93798.71	123069.04	161161.85	9671.17	0.00	0.00	0.00	0.00
Tba	0.00	0.00	18.81	109.34	71.07	0.00	0.01	0.05	0.03
Tba_s	0.00	0.33	0.47	0.52	0.00	0.02	0.02	0.02	0.00
	312 plots								
	(a)								
M	0.00	673.87	847.23	597.32	0.00	0.00	0.00	0.00	0.00
M_s	0.00	296263.90	303010.50	295660.73	309421.65	0.00	0.00	0.00	0.00
Lt	0.00	7030.30	8753.53	5430.73	1079.04	0.00	0.00	0.00	0.00
Lt_s	0.00	171255.83	206787.15	246348.76	104383.47	0.00	0.00	0.00	0.00
Tba	0.00	0.00	20.67	110.25	71.36	0.00	0.01	0.05	0.03
Tba_s	0.00	0.86	1.03	1.09	0.61	0.01	0.01	0.01	0.01
	(b)								
M	0.00	664.63	834.39	576.07	0.00	0.00	0.00	0.00	0.00
M_s	0.00	3223.46	2362.28	1888.75	10022.01	0.00	0.00	0.00	0.00
Lt	0.00	7009.14	8807.14	5378.28	982.83	0.00	0.00	0.00	0.00
Lt_s	0.00	94268.50	123801.82	162137.94	10631.01	0.00	0.00	0.00	0.00
Tba	0.00	0.00	19.44	109.49	71.08	0.00	0.01	0.05	0.03
Tba_s	0.00	0.33	0.47	0.53	0.00	0.02	0.02	0.02	0.00
	(c)								
M	0.00	664.07	832.13	565.39	0.00	0.00	0.00	0.00	0.00
M_s	0.00	3219.46	2337.28	1886.98	10007.18	0.00	0.00	0.00	0.00
Lt	0.00	7009.09	8743.73	5369.23	926.01	0.00	0.00	0.00	0.00
Lt_s	0.00	94268.50	123685.44	161969.03	9719.61	0.00	0.00	0.00	0.00
Tba	0.00	0.00	18.81	109.35	71.07	0.00	0.01	0.05	0.03
Tba_s	0.00	0.33	0.47	0.53	0.00	0.02	0.02	0.02	0.00

(2e) assume that $\lambda_h, \lambda_v, \hat{\lambda}_{c(1)}$, are $\hat{\lambda}_{c(2)}$ are all different.

Table 6.6: The three moment estimations for the SZINB model applied to real data

			(a)			(b)			(c)		
	$\hat{p}$	$\hat{\theta}$	$\hat{\lambda}_s$	$\hat{\lambda}_{hv}$	$\hat{\rho}_{hv}$	$\hat{\lambda}_s$	$\hat{\lambda}_{hv}$	$\hat{\rho}_{hv}$	$\hat{\lambda}_s$	$\hat{\lambda}_{hv}$	$\hat{\rho}_{hv}$
	313 plots										
M	0.96	70.35	21.55	10.62	0.16	22.12	10.48	0.15	22.17	10.46	0.15
M_s	0.14	53.66	0.00	5515.94	0.00	0.00	51.55	0.01	0.00	51.52	0.01
Lt	0.99	191.48	207.70	40.92	0.11	207.44	40.99	0.11	207.73	40.92	0.11
Lt_s	0.56	488.33	0.00	389.36	0.06	0.00	221.30	0.08	0.00	221.14	0.08
Tba	0.99	12.36	41.38	0.42	0.01	41.55	0.38	0.01	41.56	0.38	0.01
Tba_s	0.42	0.00	0.00	0.96	0.03	0.00	0.40	0.06	0.00	0.40	0.06
	312 plots										
M	0.96	70.29	21.60	10.65	0.16	22.22	10.49	0.15	22.29	10.48	0.15
M_s	0.14	53.66	0.00	5480.63	0.00	0.00	51.23	0.01	0.00	51.19	0.01
Lt	0.99	192.19	208.03	40.77	0.11	207.71	40.86	0.11	208.04	40.77	0.11
Lt_s	0.56	491.22	0.00	383.40	0.06	0.00	221.00	0.08	0.00	220.95	0.08
Tba	0.99	12.40	41.37	0.42	0.01	41.53	0.38	0.01	41.54	0.37	0.01
Tba_s	0.42	0.00	0.00	0.94	0.03	0.00	0.40	0.05	0.00	0.40	0.05

(1a) assume that $\lambda_h = \lambda_v = \lambda_{hv}$ and $\lambda_{c(1)} = \lambda_{c(2)} = 0$

	$\hat{\lambda}_s$	$\hat{\lambda}_h$	$\hat{\lambda}_v$	$\hat{\rho}_h$	$\hat{\rho}_v$	$\hat{\lambda}_s$	$\hat{\lambda}_h$	$\hat{\lambda}_v$	$\hat{\rho}_h$	$\hat{\rho}_v$
	313 plots									
	(a)					(b)				
M	21.47	9.43	11.85	0.14	0.17	22.04	9.31	11.69	0.14	0.17
M_s	0.00	5454.90	5579.12	0.00	0.00	0.00	59.37	43.44	0.01	0.01
Lt	207.40	36.54	45.47	0.09	0.12	207.12	36.41	45.73	0.09	0.12
Lt_s	0.00	353.76	426.22	0.05	0.06	0.00	191.71	251.93	0.07	0.09
Tba	39.95	0.00	1.55	0.00	0.03	40.14	0.00	1.46	0.00	0.03
Tba_s	0.00	0.87	1.04	0.03	0.04	0.00	0.33	0.47	0.05	0.06
	313 plots (c)					312 plots (c)				
M	22.10	9.31	11.66	0.14	0.17	21.52	9.45	11.88	0.14	0.17
M_s	0.00	59.28	43.03	0.01	0.01	0.00	5419.98	5543.41	0.00	0.00
Lt	207.73	36.41	45.42	0.09	0.12	207.72	36.39	45.31	0.09	0.12
Lt_s	0.00	191.69	251.50	0.07	0.09	0.00	347.93	420.11	0.05	0.06
Tba	40.24	0.00	1.41	0.00	0.03	39.95	0.00	1.54	0.00	0.03
Tba_s	0.00	0.33	0.47	0.05	0.06	0.00	0.86	1.03	0.03	0.04
	312 plots									
	(a)					(b)				
M	21.52	9.45	11.88	0.14	0.17	22.14	9.32	11.70	0.14	0.17
M_s	0.00	5419.98	5543.41	0.00	0.00	0.00	58.97	43.22	0.01	0.01
Lt	207.72	36.39	45.31	0.09	0.12	207.39	36.28	45.59	0.09	0.12
Lt_s	0.00	347.93	420.11	0.05	0.06	0.00	191.52	251.52	0.07	0.09
Tba	39.95	0.00	1.54	0.00	0.03	40.13	0.00	1.45	0.00	0.03
Tba_s	0.00	0.86	1.03	0.03	0.04	0.00	0.33	0.47	0.05	0.06

(1b) assume that $\lambda_{c(1)} = \lambda_{c(2)} = 0$

(Continued)

		(a)			(b)			(c)		
	$\hat{\theta}$	$\hat{\lambda}_s$	$\hat{\lambda}_{hvc}$	$\hat{\rho}_{hvc}$	$\hat{\lambda}_s$	$\hat{\lambda}_{hvc}$	$\hat{\rho}_{hvc}$	$\hat{\lambda}_s$	$\hat{\lambda}_{hvc}$	$\hat{\rho}_{hvc}$
	313 plots									
M	70.35	7.17	7.11	0.10	8.54	6.94	0.10	8.75	6.91	0.10
M_s	53.66	0.00	5542.46	0.00	0.00	79.63	0.00	0.00	79.61	0.00
Lt	191.48	136.38	29.38	0.08	137.64	29.22	0.08	137.66	29.22	0.08
Lt_s	488.33	0.00	376.11	0.02	0.00	199.15	0.03	0.00	198.58	0.03
Tba	12.36	15.22	3.48	0.08	15.55	3.44	0.08	15.57	3.44	0.08
Tba_s	0.00	0.00	0.91	0.01	0.00	0.33	0.02	0.00	0.33	0.02
	312 plots									
M	70.29	7.13	7.13	0.10	8.59	6.95	0.10	8.85	6.92	0.10
M_s	53.66	0.00	5506.97	0.00	0.00	79.14	0.00	0.00	79.10	0.00
Lt	192.19	137.08	29.26	0.08	138.22	29.11	0.08	138.23	29.11	0.08
Lt_s	491.22	0.00	370.01	0.02	0.00	198.76	0.03	0.00	198.41	0.03
Tba	12.40	15.31	3.47	0.08	15.62	3.43	0.08	15.64	3.42	0.08
Tba_s	0.00	0.00	0.90	0.01	0.00	0.33	0.02	0.00	0.33	0.02

(2a) assume that $\lambda_h = \lambda_v = \lambda_{c(1)} = \lambda_{c(2)} = \lambda_{hvc}$

	$\hat{\lambda}_s$	$\hat{\lambda}_{hv}$	$\hat{\lambda}_c$	$\hat{\rho}_{hv}$	$\hat{\rho}_c$	$\hat{\lambda}_s$	$\hat{\lambda}_{hv}$	$\hat{\lambda}_c$	$\hat{\rho}_{hv}$	$\hat{\rho}_c$
	313 plots									
	(a)					(b)				
M	8.33	10.62	3.31	0.16	0.05	9.70	10.48	3.10	0.15	0.05
M_s	0.00	5515.94	5571.15	0.00	0.00	0.00	51.55	110.04	0.00	0.00
Lt	140.19	40.92	16.88	0.11	0.04	141.52	40.99	16.48	0.11	0.04
Lt_s	0.00	389.36	361.77	0.02	0.02	0.00	221.30	175.18	0.03	0.03
Tba	14.21	0.42	6.79	0.01	0.15	14.54	0.38	6.75	0.01	0.15
Tba_s	0.00	0.96	0.86	0.01	0.01	0.00	0.40	0.25	0.03	0.02
	313 plots (c)					312 plots (c)				
M	9.95	10.46	3.06	0.15	0.04	10.05	10.48	3.06	0.15	0.04
M_s	0.00	51.52	110.02	0.00	0.00	0.00	51.19	109.32	0.00	0.00
Lt	142.46	40.92	16.32	0.11	0.04	143.01	40.77	16.26	0.11	0.04
Lt_s	0.00	221.14	173.88	0.03	0.03	0.00	220.95	173.73	0.03	0.03
Tba	14.56	0.38	6.75	0.01	0.15	14.63	0.37	6.73	0.01	0.15
Tba_s	0.00	0.40	0.25	0.03	0.02	0.00	0.40	0.25	0.03	0.02
	312 plots									
	(a)					(b)				
M	8.29	10.65	3.33	0.16	0.05	9.76	10.49	3.12	0.15	0.05
M_s	0.00	5480.63	5535.49	0.00	0.00	0.00	51.23	109.35	0.00	0.00
Lt	140.88	40.77	16.79	0.11	0.04	142.10	40.86	16.40	0.11	0.04
Lt_s	0.00	383.40	355.51	0.02	0.02	0.00	221.00	174.68	0.03	0.03
Tba	14.30	0.42	6.77	0.01	0.15	14.62	0.38	6.73	0.01	0.15
Tba_s	0.00	0.94	0.85	0.01	0.01	0.00	0.40	0.26	0.03	0.02

(2b) assume that $\lambda_h = \lambda_v = \lambda_{hv}$ and $\lambda_{c(1)} = \lambda_{c(2)} = \lambda_c$

(Continued)

	$\hat{\theta}$	$\hat{\lambda}_s$	$\hat{\lambda}_h$	$\hat{\lambda}_v$	$\hat{\lambda}_c$	$\hat{\rho}_h$	$\hat{\rho}_v$	$\hat{\rho}_c$
	313 plots							
	(a)							
M	70.35	8.24	9.43	11.85	3.31	0.14	0.17	0.05
M_s	53.66	0.00	5454.90	5579.12	5571.15	0.00	0.00	0.00
Lt	191.48	139.89	36.54	45.47	16.88	0.09	0.12	0.04
Lt_s	488.33	0.00	353.76	426.22	361.77	0.02	0.02	0.02
Tba	12.36	12.78	0.00	1.55	6.79	0.00	0.03	0.15
Tba_s	0.00	0.00	0.87	1.04	0.86	0.01	0.01	0.01
	(b)							
M	70.35	9.62	9.31	11.69	3.10	0.14	0.17	0.05
M_s	53.66	0.00	59.37	43.44	110.04	0.00	0.00	0.00
Lt	191.48	141.20	36.41	45.73	16.48	0.09	0.12	0.04
Lt_s	488.33	0.00	191.71	251.93	175.18	0.03	0.04	0.03
Tba	12.36	13.14	0.00	1.46	6.75	0.00	0.03	0.15
Tba_s	0.00	0.00	0.33	0.47	0.25	0.02	0.03	0.02
	(c)							
M	70.35	9.87	9.31	11.66	3.06	0.14	0.17	0.04
M_s	53.66	0.00	59.28	43.03	110.02	0.00	0.00	0.00
Lt	191.48	142.46	36.41	45.42	16.32	0.09	0.12	0.04
Lt_s	488.33	0.00	191.69	251.50	173.88	0.03	0.04	0.03
Tba	12.36	13.25	0.00	1.41	6.75	0.00	0.03	0.15
Tba_s	0.00	0.00	0.33	0.47	0.25	0.02	0.03	0.02
	312 plots							
	(a)							
M	70.29	8.20	9.45	11.88	3.33	0.14	0.17	0.05
M_s	53.66	0.00	5419.98	5543.41	5535.49	0.00	0.00	0.00
Lt	192.19	140.57	36.39	45.31	16.79	0.09	0.12	0.04
Lt_s	491.22	0.00	347.93	420.11	355.51	0.02	0.02	0.02
Tba	12.40	12.88	0.00	1.54	6.77	0.00	0.03	0.15
Tba_s	0.00	0.00	0.86	1.03	0.85	0.01	0.01	0.01
	(b)							
M	70.29	9.67	9.32	11.70	3.12	0.14	0.17	0.05
M_s	53.66	0.00	58.97	43.22	109.35	0.00	0.00	0.00
Lt	192.19	141.78	36.28	45.59	16.40	0.09	0.12	0.04
Lt_s	491.22	0.00	191.52	251.52	174.68	0.03	0.04	0.03
Tba	12.40	13.22	0.00	1.45	6.73	0.00	0.03	0.15
Tba_s	0.00	0.00	0.33	0.47	0.26	0.02	0.03	0.02
	(c)							
M	70.29	9.98	9.32	11.67	3.06	0.14	0.17	0.04
M_s	53.66	0.00	58.90	42.76	109.32	0.00	0.00	0.00
Lt	192.19	143.01	36.28	45.26	16.26	0.09	0.12	0.04
Lt_s	491.22	0.00	191.52	251.28	173.73	0.03	0.04	0.03
Tba	12.40	13.32	0.00	1.40	6.73	0.00	0.03	0.15
Tba_s	0.00	0.00	0.33	0.47	0.25	0.02	0.03	0.02

(2c) assume that $\lambda_{c(1)} = \lambda_{c(2)} = \lambda_c$

(Continued)

	$\hat{\lambda}_s$	$\hat{\lambda}_{hv}$	$\hat{\lambda}_{c(1)}$	$\hat{\lambda}_{c(2)}$	$\hat{\rho}_{hv}$	$\hat{\rho}_{c(1)}$	$\hat{\rho}_{c(2)}$
	313 plots						
	(a)						
M	4.86	10.62	8.34	0.00	0.16	0.12	0.00
M_s	0.00	5515.94	5443.80	5697.17	0.00	0.00	0.00
Lt	139.95	40.92	28.23	5.64	0.11	0.07	0.01
Lt_s	0.00	389.36	506.71	218.37	0.02	0.03	0.01
Tba	14.18	0.42	8.26	5.34	0.01	0.18	0.12
Tba_s	0.00	0.96	1.10	0.62	0.01	0.01	0.01
	(b)						
M	6.00	10.48	8.06	0.00	0.15	0.12	0.00
M_s	0.00	51.55	34.79	184.49	0.00	0.00	0.00
Lt	141.28	40.99	27.95	5.13	0.11	0.07	0.01
Lt_s	0.00	221.30	329.91	22.09	0.03	0.05	0.00
Tba	14.51	0.38	8.20	5.32	0.01	0.18	0.12
Tba_s	0.00	0.40	0.52	0.00	0.03	0.03	0.00
	(c)						
M	6.33	10.46	7.92	0.00	0.15	0.12	0.00
M_s	0.00	51.52	34.74	184.26	0.00	0.00	0.00
Lt_s	142.32	40.92	27.89	4.81	0.11	0.07	0.01
Lt	0.00	221.14	329.35	19.76	0.03	0.05	0.00
Tba	14.54	0.38	8.19	5.32	0.01	0.18	0.12
Tba_s	0.00	0.40	0.52	0.00	0.03	0.03	0.00
	312 plots						
	(a)						
M	4.84	10.65	8.38	0.00	0.16	0.12	0.00
M_s	0.00	5480.63	5408.95	5660.69	0.00	0.00	0.00
Lt	140.64	40.77	28.11	5.59	0.11	0.07	0.01
Lt_s	0.00	383.40	500.49	212.07	0.02	0.03	0.01
Tba	14.27	0.42	8.23	5.32	0.01	0.18	0.12
Tba_s	0.00	0.94	1.09	0.61	0.01	0.01	0.01
	(b)						
M	6.06	10.49	8.08	0.00	0.15	0.12	0.00
M_s	0.00	51.23	34.55	183.35	0.00	0.00	0.00
Lt	141.85	40.86	27.84	5.09	0.11	0.07	0.01
Lt_s	0.00	221.00	329.40	21.60	0.03	0.05	0.00
Tba	14.59	0.38	8.17	5.30	0.01	0.18	0.12
Tba_s	0.00	0.40	0.53	0.00	0.03	0.03	0.00
	(c)						
M	6.43	10.48	7.93	0.00	0.15	0.12	0.00
M_s	0.00	51.19	34.52	183.08	0.00	0.00	0.00
Lt	142.87	40.77	27.79	4.79	0.11	0.07	0.01
Lt_s	0.00	220.95	329.06	19.75	0.03	0.05	0.00
Tba	14.62	0.37	8.16	5.30	0.01	0.18	0.12
Tba_s	0.00	0.40	0.53	0.00	0.03	0.03	0.00

(2d) assume that $\lambda_h = \lambda_v = \lambda_{hv}$

(Continued)

	$\hat{\lambda}_s$	$\hat{\lambda}_h$	$\hat{\lambda}_v$	$\hat{\lambda}_{c(1)}$	$\hat{\lambda}_{c(2)}$	$\hat{\rho}_h$	$\hat{\rho}_v$	$\hat{\rho}_{c(1)}$	$\hat{\rho}_{c(2)}$
	313 plots								
	(a)								
M	4.78	9.43	11.85	8.34	0.00	0.14	0.17	0.12	0.00
M_s	0.00	5454.90	5579.12	5443.80	5697.17	0.00	0.00	0.00	0.00
Lt	139.65	36.54	45.47	28.23	5.64	0.09	0.12	0.07	0.01
Lt_s	0.00	353.76	426.22	506.71	218.37	0.02	0.02	0.03	0.01
Tba	12.75	0.00	1.55	8.26	5.34	0.00	0.03	0.18	0.12
Tba_s	0.00	0.87	1.04	1.10	0.62	0.01	0.01	0.01	0.01
	(b)								
M	5.92	9.31	11.69	8.06	0.00	0.14	0.17	0.12	0.00
M_s	0.00	59.37	43.44	34.79	184.49	0.00	0.00	0.00	0.00
Lt	140.96	36.41	45.73	27.95	5.13	0.09	0.12	0.07	0.01
Lt_s	0.00	191.71	251.93	329.91	22.09	0.03	0.04	0.05	0.00
Tba	13.11	0.00	1.46	8.20	5.32	0.00	0.03	0.18	0.12
Tba_s	0.00	0.33	0.47	0.52	0.00	0.02	0.03	0.03	0.00
	(c)								
M	6.25	9.31	11.66	7.92	0.00	0.14	0.17	0.12	0.00
M_s	0.00	59.28	43.03	34.74	184.26	0.00	0.00	0.00	0.00
Lt	142.32	36.41	45.42	27.89	4.81	0.09	0.12	0.07	0.01
Lt_s	0.00	191.69	251.50	329.35	19.76	0.03	0.04	0.05	0.00
Tba	13.23	0.00	1.41	8.19	5.32	0.00	0.03	0.18	0.12
Tba_s	0.00	0.33	0.47	0.52	0.00	0.02	0.03	0.03	0.00
	312 plots								
	(a)								
M	4.76	9.45	11.88	8.38	0.00	0.14	0.17	0.12	0.00
M_s	0.00	5419.98	5543.41	5408.95	5660.69	0.00	0.00	0.00	0.00
Lt	140.33	36.39	45.31	28.11	5.59	0.09	0.12	0.07	0.01
Lt_s	0.00	347.93	420.11	500.49	212.07	0.02	0.02	0.03	0.01
Tba	12.85	0.00	1.54	8.23	5.32	0.00	0.03	0.18	0.12
Tba_s	0.00	0.86	1.03	1.09	0.61	0.01	0.017	0.01	0.01
	(b)								
M	5.97	9.32	11.70	8.08	0.00	0.14	0.17	0.12	0.00
M_s	0.00	58.97	43.22	34.55	183.35	0.00	0.00	0.00	0.00
Lt	141.53	36.28	45.59	27.84	5.09	0.09	0.12	0.07	0.01
Lt_s	0.00	191.52	251.52	329.40	21.60	0.03	0.04	0.05	0.00
Tba	13.19	0.00	1.45	8.17	5.30	0.00	0.03	0.18	0.12
Tba_s	0.00	0.33	0.47	0.53	0.00	0.02	0.03	0.03	0.00
	(c)								
M	6.35	9.32	11.67	7.93	0.00	0.14	0.17	0.12	0.00
M_s	0.00	58.90	42.76	34.52	183.08	0.00	0.00	0.00	0.00
Lt	142.87	36.28	45.26	27.79	4.79	0.09	0.12	0.07	0.01
Lt_s	0.00	191.52	251.28	329.06	19.75	0.03	0.04	0.05	0.00
Tba	13.31	0.00	1.40	8.16	5.30	0.00	0.03	0.18	0.12
Tba_s	0.00	0.33	0.47	0.53	0.00	0.02	0.03	0.03	0.00

(2e) assume that $\lambda_h, \lambda_v, \hat{\lambda}_{c(1)}$, are $\hat{\lambda}_{c(2)}$ are all different.

Table 6.7: The moment estimations for the SZIE model applied to real data in the Siuslaw Forest

	$\hat{p}$	$\hat{\beta}_s$	$\hat{\beta}_{hv}$	$\hat{\rho}_{hv}$
	313 plots			
M	0.96	0.00	27.34	0.20
M_s	0.14	0.00	53.08	0.00
Lt	0.99	17.10	88.82	0.23
Lt_s	0.56	0.00	329.07	0.04
Tba	0.99	35.07	2.25	0.00
Tba_s	0.42	0.00	0.63	0.02
	312 plots			
M	0.96	0.00	27.35	0.20
M_s	0.14	0.00	52.92	0.00
Lt	0.99	16.77	88.84	0.23
Lt_s	0.56	0.00	329.82	0.04
Tba	0.99	35.05	2.25	0.00
Tba_s	0.42	0.00	0.63	0.02

(1a) assume that $\beta_h = \beta_v = \beta_{hv}$ and $\beta_{c(1)} = \beta_{c(2)} = 0$

	$\hat{p}$	$\hat{\beta}_s$	$\hat{\beta}_h$	$\hat{\beta}_v$	$\hat{\rho}_h$	$\hat{\rho}_v$
	313 plots					
M	0.96	0.00	25.77	28.88	0.18	0.23
M_s	0.14	0.00	56.97	48.73	0.00	0.00
Lt	0.99	17.33	83.72	93.82	0.20	0.25
Lt_s	0.56	0.00	306.29	351.11	0.03	0.05
Tba	0.99	35.23	0.00	4.41	0.00	0.01
Tba_s	0.42	0.00	0.58	0.69	0.02	0.03
	312 plots					
M	0.96	0.00	25.78	28.89	0.18	0.23
M_s	0.14	0.00	56.78	48.60	0.00	0.00
Lt	0.99	17.00	83.72	93.85	0.20	0.25
Lt_s	0.56	0.00	307.03	351.85	0.03	0.05
Tba	0.99	35.22	0.00	4.41	0.00	0.01
Tba_s	0.42	0.00	0.58	0.69	0.02	0.03

(1b) assume that $\beta_{c(1)} = \beta_{c(2)} = 0$

(Continued)

	$\hat{p}$	$\hat{\beta}_s$	$\hat{\beta}_{hvc}$	$\hat{\rho}_{hvc}$
	313 plots			
M	0.96	0.00	22.25	0.09
M_s	0.14	0.00	65.98	0.00
Lt	0.99	0.00	74.99	0.11
Lt_s	0.56	0.00	312.17	0.01
Tba	0.99	0.00	6.78	0.11
Tba_s	0.42	0.00	0.57	0.01
	312 plots			
M	0.96	0.00	22.26	0.09
M_s	0.14	0.00	65.77	0.00
Lt	0.99	0.00	75.00	0.11
Lt_s	0.56	0.00	312.78	0.01
Tba	0.99	0.00	6.78	0.11
Tba_s	0.42	0.00	0.58	0.01

(2a) assume that $\beta_h = \beta_v = \beta_{c(1)} = \beta_{c(2)} = \beta_{hvc}$

	$\hat{p}$	$\hat{\beta}_s$	$\hat{\beta}_{hv}$	$\hat{\beta}_c$	$\hat{\rho}_{hv}$	$\hat{\rho}_c$
	313 plots					
M	0.96	0.00	27.34	14.88	0.14	0.04
M_s	0.14	0.00	53.08	77.55	0.00	0.00
Lt	0.99	0.00	88.82	56.32	0.16	0.06
Lt_s	0.56	0.00	329.07	292.78	0.01	0.01
Tba	0.99	0.00	2.25	9.50	0.01	0.22
Tba_s	0.42	0.00	0.63	0.50	0.01	0.01
	312 plots					
M	0.96	0.00	27.35	14.90	0.14	0.04
M_s	0.14	0.00	52.92	77.31	0.00	0.00
Lt	0.99	0.00	88.84	56.29	0.16	0.06
Lt_s	0.56	0.00	329.82	293.23	0.01	0.01
Tba	0.99	0.00	2.25	9.50	0.01	0.21
Tba_s	0.42	0.00	0.63	0.51	0.01	0.01

(2b) assume that $\beta_h = \beta_v = \beta_{hv}$ and $\beta_{c(1)} = \beta_{c(2)} = \beta_c$

(Continued)

	$\hat{p}$	$\hat{\beta}_s$	$\hat{\beta}_h$	$\hat{\beta}_v$	$\hat{\beta}_c$	$\hat{\rho}_h$	$\hat{\rho}_v$	$\hat{\rho}_c$
	313 plots							
M	0.96	0.00	25.77	28.88	14.88	0.13	0.16	0.04
M_s	0.14	0.00	56.97	48.73	77.55	0.00	0.00	0.00
Lt	0.99	0.00	83.72	93.82	56.32	0.14	0.18	0.06
Lt_s	0.56	0.00	306.29	351.11	292.78	0.01	0.02	0.01
Tba	0.99	0.00	0.00	4.41	9.50	0.00	0.04	0.21
Tba_s	0.42	0.00	0.58	0.69	0.50	0.01	0.01	0.01
	312 plots							
M	0.96	0.00	25.78	28.89	14.90	0.13	0.16	0.04
M_s	0.14	0.00	56.78	48.60	77.31	0.00	0.00	0.00
Lt	0.99	0.00	83.72	93.85	56.29	0.14	0.18	0.06
Lt_s	0.56	0.00	307.03	351.85	293.23	0.01	0.02	0.01
Tba	0.99	0.00	0.00	4.41	9.50	0.00	0.04	0.21
Tba_s	0.42	0.00	0.58	0.69	0.51	0.01	0.01	0.01

(2c) assume that $\beta_{c(1)} = \beta_{c(2)} = \beta_c$

	$\hat{p}$	$\hat{\beta}_s$	$\hat{\beta}_{hv}$	$\hat{\beta}_{c(1)}$	$\hat{\beta}_{c(2)}$	$\hat{\rho}_{hv}$	$\hat{\rho}_{c(1)}$	$\hat{\rho}_{c(2)}$
	313 plots							
M	0.96	0.00	27.34	23.98	0.00	0.14	0.11	0.00
M_s	0.14	0.00	53.08	43.61	100.42	0.00	0.00	0.00
Lt	0.99	0.00	88.82	73.35	31.42	0.16	0.11	0.02
Lt_s	0.56	0.00	329.07	401.79	103.96	0.02	0.02	0.00
Tba	0.99	0.00	2.25	10.46	8.43	0.01	0.26	0.17
Tba_s	0.42	0.00	0.63	0.72	0.00	0.01	0.01	0.00
	312 plots							
M	0.96	0.00	27.35	24.00	0.00	0.14	0.11	0.00
M_s	0.14	0.00	52.92	43.46	100.11	0.00	0.00	0.00
Lt	0.99	0.00	88.84	73.34	31.35	0.16	0.11	0.02
Lt_s	0.56	0.00	329.82	402.66	103.11	0.02	0.02	0.00
Tba	0.99	0.00	2.25	10.46	8.43	0.01	0.26	0.17
Tba_s	0.42	0.00	0.63	0.73	0.00	0.01	0.01	0.00

(2d) assume that $\beta_h = \beta_v = \beta_{hv}$

(Continued)

	$\hat{p}$	$\hat{\beta}_s$	$\hat{\beta}_h$	$\hat{\beta}_v$	$\hat{\beta}_{c(1)}$	$\hat{\beta}_{c(2)}$	$\hat{\rho}_h$	$\hat{\rho}_v$	$\hat{\rho}_{c(1)}$	$\hat{\rho}_{c(2)}$
	313 plots									
M	0.96	0.00	25.77	28.88	23.98	0.00	0.12	0.16	0.11	0.00
M_s	0.14	0.00	56.97	48.73	43.61	100.42	0.00	0.00	0.00	0.00
Lt	0.99	0.00	83.72	93.82	73.35	31.42	0.14	0.18	0.11	0.02
Lt_s	0.56	0.00	306.29	351.11	401.79	103.96	0.01	0.02	0.02	0.00
Tba	0.99	0.00	0.00	4.41	10.46	8.43	0.00	0.04	0.25	0.16
Tba_s	0.42	0.00	0.58	0.69	0.72	0.00	0.01	0.01	0.01	0.00
	312 plots									
M	0.96	0.00	25.78	28.89	24.00	0.00	0.12	0.16	0.11	0.00
M_s	0.14	0.00	56.78	48.60	43.46	100.11	0.00	0.00	0.00	0.00
Lt	0.99	0.00	83.72	93.85	73.34	31.35	0.14	0.18	0.11	0.02
Lt_s	0.56	0.00	307.03	351.85	402.66	103.11	0.01	0.02	0.02	0.00
Tba	0.99	0.00	0.00	4.41	10.46	8.43	0.00	0.04	0.25	0.16
Tba_s	0.42	0.00	0.58	0.69	0.73	0.00	0.01	0.01	0.01	0.00

(2e) assume that $\beta_h, \beta_v, \hat{\beta}_{c(1)}$, are $\hat{\beta}_{c(2)}$ are all different.

Table 6.8: The moment estimations for the SZIG model applied to real data in the Siuslaw Forest

		$\hat{\gamma}$				alternative $\hat{\gamma}'$			
	$\hat{p}$	$\hat{\gamma}$	$\hat{\alpha}_s$	$\hat{\alpha}_{hv}$	$\hat{\rho}_{hv}$	$\hat{\gamma}'$	$\hat{\alpha'}_s$	$\hat{\alpha'}_{hv}$	$\hat{\rho'}_{hv}$
	313 plots								
M	0.96	70.25	22.47	10.64	0.15	70.25	22.47	10.64	0.15
M_s	0.14	54.05	0.00	52.13	0.01	54.88	0.00	51.34	0.01
Lt	0.99	191.96	207.99	41.10	0.11	191.96	207.99	41.10	0.11
Lt_s	0.56	487.96	0.00	221.92	0.08	488.68	0.00	221.60	0.08
Tba	0.99	13.05	42.51	0.39	0.01	13.05	42.51	0.39	0.01
Tba_s	0.42	0.84	0.00	0.47	0.05	0.85	0.00	0.47	0.05
	312 plots								
M	0.96	70.20	22.56	10.66	0.15	70.20	22.56	10.66	0.15
M_s	0.14	54.05	0.00	51.81	0.01	54.88	0.00	51.03	0.01
Lt	0.99	192.67	208.27	40.97	0.11	192.66	208.26	40.97	0.11
Lt_s	0.56	490.83	0.00	221.63	0.08	491.57	0.00	221.30	0.08
Tba	0.99	13.10	42.49	0.39	0.01	13.10	42.49	0.39	0.01
Tba_s	0.42	0.85	0.00	0.47	0.05	0.85	0.00	0.47	0.05

(1a) assume that $\alpha_h = \alpha_v = \alpha_{hv}$ and $\alpha_{c(1)} = \alpha_{c(2)} = 0$

	$\hat{\gamma}$					alternative $\hat{\gamma}'$				
	$\hat{\alpha}_s$	$\hat{\alpha}_h$	$\hat{\alpha}_v$	$\hat{\rho}_h$	$\hat{\rho}_v$	$\hat{\alpha'}_s$	$\hat{\alpha'}_h$	$\hat{\alpha'}_v$	$\hat{\rho'}_h$	$\hat{\rho'}_v$
	313 plots									
M	22.38	9.46	11.87	0.14	0.17	22.38	9.46	11.87	0.14	0.17
M_s	0.00	60.05	43.94	0.01	0.01	0.00	59.14	43.27	0.01	0.01
Lt	207.67	36.51	45.85	0.09	0.12	207.67	36.51	45.85	0.09	0.12
Lt_s	0.00	192.25	252.63	0.07	0.09	0.00	191.97	252.26	0.07	0.09
Tba	41.08	0.00	1.49	0.00	0.03	41.07	0.00	1.49	0.00	0.03
Tba_s	0.00	0.39	0.56	0.04	0.05	0.00	0.39	0.56	0.04	0.05
	312 plots									
M	22.48	9.47	11.89	0.14	0.17	22.48	9.47	11.89	0.14	0.17
M_s	0.00	59.64	43.71	0.01	0.01	0.00	58.74	43.05	0.01	0.01
Lt	207.95	36.38	45.71	0.09	0.12	207.94	36.38	45.71	0.09	0.12
Lt_s	0.00	192.06	252.23	0.07	0.09	0.00	191.77	251.85	0.07	0.09
Tba	41.07	0.00	1.48	0.00	0.03	41.07	0.00	1.48	0.00	0.03
Tba_s	0.00	0.39	0.56	0.04	0.05	0.00	0.39	0.55	0.04	0.05

(1b) assume that $\alpha_{c(1)} = \alpha_{c(2)} = 0$

(Continued)

	$\hat{\gamma}$			alternative $\hat{\gamma}'$		
	$\hat{\alpha}_s$	$\hat{\alpha}_{hvc}$	$\hat{\rho}_{hvc}$	$\hat{\alpha'}_s$	$\hat{\alpha'}_{hvc}$	$\hat{\rho'}_{hvc}$
	313 plots					
M	8.67	7.05	0.10	8.67	7.05	0.10
M_s	0.00	80.54	0.00	0.00	79.32	0.00
Lt	138.01	29.30	0.08	138.00	29.30	0.08
Lt_s	0.00	199.71	0.03	0.00	199.42	0.03
Tba	15.91	3.52	0.08	15.91	3.52	0.08
Tba_s	0.00	0.39	0.02	0.00	0.39	0.02
	312 plots					
M	8.72	7.06	0.10	8.72	7.06	0.10
M_s	0.00	80.03	0.00	0.00	78.83	0.00
Lt	138.60	29.19	0.08	138.59	29.19	0.08
Lt_s	0.00	199.32	0.03	0.00	199.02	0.03
Tba	15.99	3.51	0.08	15.98	3.51	0.08
Tba_s	0.00	0.39	0.02	0.00	0.39	0.02

(2a) assume that $\alpha_h = \alpha_v = \alpha_{hv} = \alpha_{c(1)} = \alpha_{c(2)} = \alpha_{hvc}$

	$\hat{\gamma}$					alternative $\hat{\gamma}'$				
	$\hat{\alpha}_s$	$\hat{\alpha}_{hv}$	$\hat{\alpha}_c$	$\hat{\rho}_{hv}$	$\hat{\rho}_c$	$\hat{\alpha'}_s$	$\hat{\alpha'}_{hv}$	$\hat{\alpha'}_c$	$\hat{\rho'}_{hv}$	$\hat{\rho'}_c$
	313 plots									
M	9.85	10.64	3.15	0.15	0.05	9.86	10.64	3.15	0.15	0.05
M_s	0.00	52.13	111.29	0.00	0.00	0.00	51.34	109.60	0.00	0.00
Lt	141.91	41.10	16.52	0.11	0.04	141.90	41.10	16.52	0.11	0.04
Lt_s	0.00	221.92	175.67	0.03	0.03	0.00	221.60	175.41	0.03	0.03
Tba	14.88	0.39	6.91	0.01	0.15	14.87	0.39	6.91	0.01	0.15
Tba_s	0.00	0.47	0.30	0.02	0.01	0.00	0.47	0.30	0.02	0.01
	312 plots									
M	9.91	10.66	3.16	0.15	0.05	9.91	10.66	3.16	0.15	0.05
M_s	0.00	51.81	110.59	0.00	0.00	0.00	51.03	108.92	0.00	0.00
Lt	142.48	40.97	16.45	0.11	0.04	142.47	40.97	16.45	0.11	0.04
Lt_s	0.00	221.63	175.18	0.03	0.03	0.00	221.30	174.91	0.03	0.03
Tba	14.96	0.39	6.88	0.01	0.15	14.95	0.39	6.89	0.01	0.15
Tba_s	0.00	0.47	0.30	0.02	0.01	0.00	0.47	0.30	0.02	0.01

(2b) assume that $\alpha_h = \alpha_v = \alpha_{hv}$ and $\alpha_{c(1)} = \alpha_{c(2)} = \alpha_c$

(Continued)

	313 plots						
	$\hat{\gamma}$						
	$\hat{\alpha}_s$	$\hat{\alpha}_h$	$\hat{\alpha}_v$	$\hat{\alpha}_c$	$\hat{\rho}_h$	$\hat{\rho}_v$	$\hat{\rho}_c$
M	9.77	9.46	11.87	3.15	0.14	0.17	0.05
M_s	0.00	60.05	43.94	111.29	0.00	0.00	0.00
Lt	141.58	36.51	45.85	16.52	0.09	0.12	0.04
Lt_s	0.00	192.25	252.63	175.67	0.03	0.04	0.03
Tba	13.44	0.00	1.49	6.91	0.00	0.03	0.15
Tba_s	0.00	0.39	0.56	0.30	0.02	0.02	0.01
	alternative $\hat{\gamma}'$						
	$\hat{\alpha'}_s$	$\hat{\alpha'}_h$	$\hat{\alpha'}_v$	$\hat{\alpha'}_c$	$\hat{\rho'}_h$	$\hat{\rho'}_v$	$\hat{\rho'}_c$
M	9.77	9.46	11.87	3.15	0.14	0.17	0.05
M_s	0.00	59.14	43.27	109.60	0.00	0.00	0.00
Lt	141.57	36.51	45.85	16.52	0.09	0.12	0.04
Lt_s	0.00	191.97	252.26	175.41	0.03	0.04	0.03
Tba	13.44	0.00	1.49	6.91	0.00	0.03	0.15
Tba_s	0.00	0.39	0.56	0.30	0.02	0.02	0.01
	312 plots						
	$\hat{\gamma}$						
	$\hat{\alpha}_s$	$\hat{\alpha}_h$	$\hat{\alpha}_v$	$\hat{\alpha}_c$	$\hat{\rho}_h$	$\hat{\rho}_v$	$\hat{\rho}_c$
M	9.82	9.47	11.89	3.16	0.14	0.17	0.05
M_s	0.00	59.64	43.71	110.59	0.00	0.00	0.00
Lt	142.16	36.38	45.71	16.45	0.09	0.12	0.04
Lt_s	0.00	192.06	252.23	175.18	0.03	0.04	0.03
Tba	13.53	0.00	1.48	6.88	0.00	0.03	0.15
Tba_s	0.00	0.39	0.56	0.30	0.02	0.02	0.01
	alternative $\hat{\gamma}'$						
	$\hat{\alpha'}_s$	$\hat{\alpha'}_h$	$\hat{\alpha'}_v$	$\hat{\alpha'}_c$	$\hat{\rho'}_h$	$\hat{\rho'}_v$	$\hat{\rho'}_c$
M	9.83	9.47	11.89	3.16	0.14	0.17	0.05
M_s	0.00	58.74	43.05	108.92	0.00	0.00	0.00
Lt	142.15	36.38	45.71	16.45	0.09	0.12	0.04
Lt_s	0.00	191.77	251.85	174.91	0.03	0.04	0.03
Tba	13.53	0.00	1.48	6.89	0.00	0.03	0.15
Tba_s	0.00	0.39	0.55	0.30	0.02	0.02	0.01

(2c) assume that $\alpha_{c(1)} = \alpha_{c(2)} = \alpha_c$

(Continued)

	313 plots						
	$\hat{\gamma}$						
	$\hat{\alpha}_s$	$\hat{\alpha}_{hv}$	$\hat{\alpha}_{c(1)}$	$\hat{\alpha}_{c(2)}$	$\hat{\rho}_{hv}$	$\hat{\rho}_{c(1)}$	$\hat{\rho}_{c(2)}$
M	6.10	10.64	8.18	0.00	0.15	0.12	0.00
M_s	0.00	52.13	35.19	186.58	0.00	0.00	0.00
Lt	141.66	41.10	28.02	5.14	0.11	0.07	0.01
Lt_s	0.00	221.92	330.84	22.15	0.03	0.05	0.00
Tba	14.85	0.39	8.39	5.44	0.01	0.18	0.12
Tba_s	0.00	0.47	0.62	0.00	0.02	0.03	0.00
	alternative $\hat{\gamma}'$						
	$\hat{\alpha}'_s$	$\hat{\alpha}'_{hv}$	$\hat{\alpha}'_{c(1)}$	$\hat{\alpha}'_{c(2)}$	$\hat{\rho}'_{hv}$	$\hat{\rho}'_{c(1)}$	$\hat{\rho}'_{c(2)}$
M	6.10	10.64	8.18	0.00	0.15	0.12	0.00
M_s	0.00	51.34	34.65	183.76	0.00	0.00	0.00
Lt	141.65	41.10	28.03	5.14	0.11	0.07	0.01
Lt_s	0.00	221.60	330.35	22.12	0.03	0.05	0.00
Tba	14.84	0.39	8.39	5.44	0.01	0.18	0.12
Tba_s	0.00	0.47	0.62	0.00	0.02	0.03	0.00
	312 plots						
	$\hat{\gamma}$						
	$\hat{\alpha}_s$	$\hat{\alpha}_{hv}$	$\hat{\alpha}_{c(1)}$	$\hat{\alpha}_{c(2)}$	$\hat{\rho}_{hv}$	$\hat{\rho}_{c(1)}$	$\hat{\rho}_{c(2)}$
M	6.15	10.66	8.21	0.00	0.15	0.12	0.00
M_s	0.00	51.81	34.95	185.43	0.00	0.00	0.00
Lt	142.24	40.97	27.91	5.10	0.11	0.07	0.01
Lt_s	0.00	221.63	330.34	21.66	0.03	0.05	0.00
Tba	14.92	0.39	8.36	5.43	0.01	0.18	0.12
Tba_s	0.00	0.47	0.62	0.00	0.02	0.03	0.00
	alternative $\hat{\gamma}'$						
	$\hat{\alpha}'_s$	$\hat{\alpha}'_{hv}$	$\hat{\alpha}'_{c(1)}$	$\hat{\alpha}'_{c(2)}$	$\hat{\rho}'_{hv}$	$\hat{\rho}'_{c(1)}$	$\hat{\rho}'_{c(2)}$
M	6.15	10.66	8.21	0.00	0.15	0.12	0.00
M_s	0.00	51.03	34.42	182.63	0.00	0.00	0.00
Lt	142.23	40.97	27.92	5.10	0.11	0.07	0.01
Lt_s	0.00	221.30	329.84	21.63	0.03	0.05	0.00
Tba	14.92	0.39	8.36	5.43	0.01	0.18	0.12
Tba_s	0.00	0.47	0.62	0.00	0.02	0.03	0.00

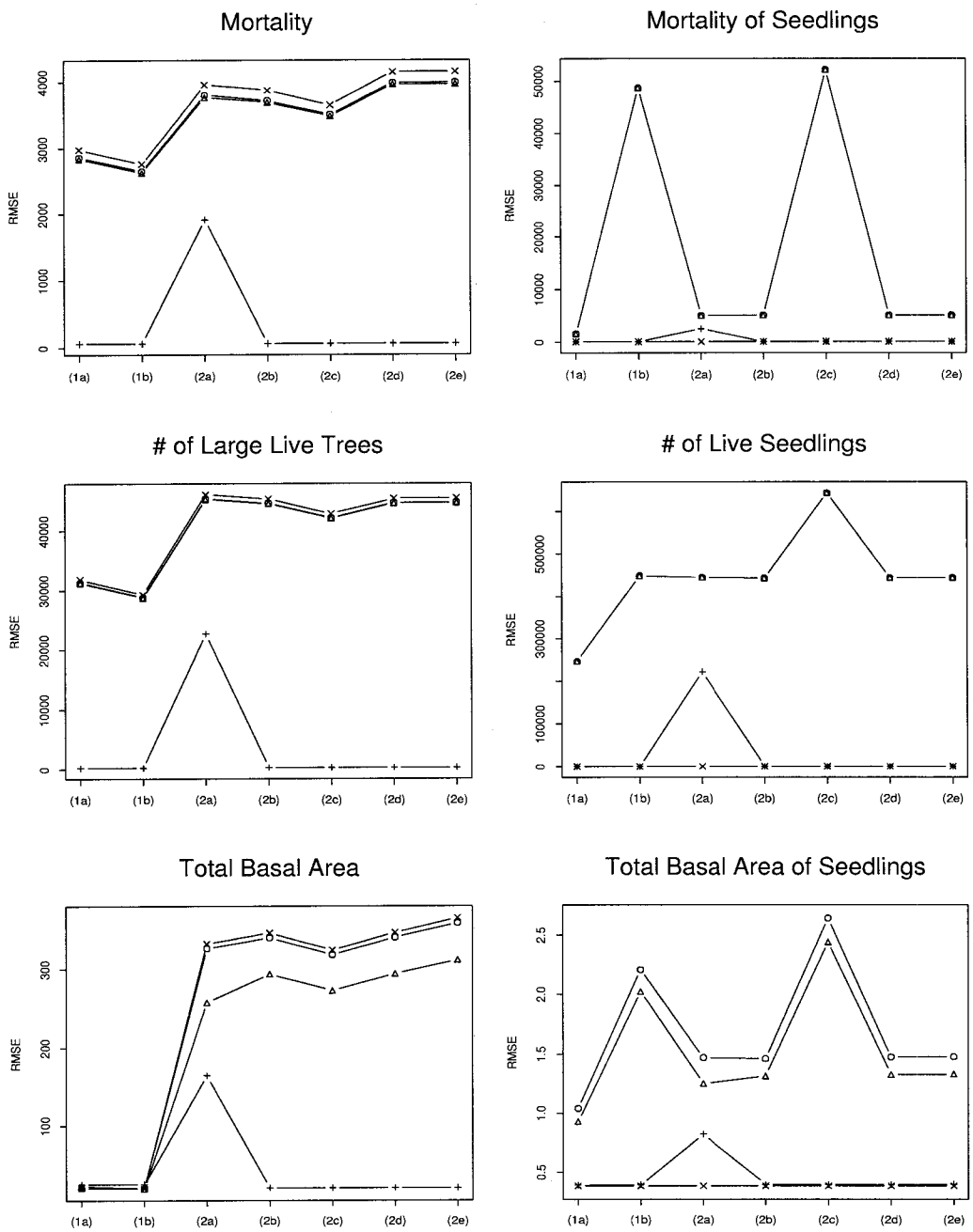
(2d) assume that $\alpha_h = \alpha_v = \alpha_{hv}$

(Continued)

	313 plots								
	$\hat{\gamma}$								
	$\hat{\alpha}_s$	$\hat{\alpha}_h$	$\hat{\alpha}_v$	$\hat{\alpha}_{c(1)}$	$\hat{\alpha}_{c(2)}$	$\hat{\rho}_h$	$\hat{\rho}_v$	$\hat{\rho}_{c(1)}$	$\hat{\rho}_{c(2)}$
M	6.01	9.46	11.87	8.18	0.00	0.14	0.17	0.12	0.00
M_s	0.00	60.05	43.94	35.19	186.58	0.00	0.00	0.00	0.00
Lt	141.34	36.51	45.85	28.02	5.14	0.09	0.12	0.07	0.01
Lt_s	0.00	192.25	252.63	330.84	22.15	0.03	0.04	0.05	0.00
Tba	13.41	0.00	1.49	8.39	5.44	0.00	0.03	0.18	0.12
Tba_s	0.00	0.39	0.56	0.62	0.00	0.02	0.02	0.03	0.00
	alternative $\hat{\gamma}'$								
	$\hat{\alpha'}_s$	$\hat{\alpha'}_h$	$\hat{\alpha'}_v$	$\hat{\alpha'}_{c(1)}$	$\hat{\alpha'}_{c(2)}$	$\hat{\rho'}_h$	$\hat{\rho'}_v$	$\hat{\rho'}_{c(1)}$	$\hat{\rho'}_{c(2)}$
M	6.01	9.46	11.87	8.18	0.00	0.14	0.17	0.12	0.00
M_s	0.00	59.14	43.27	34.65	183.76	0.00	0.00	0.00	0.00
Lt	141.33	36.51	45.85	28.03	5.14	0.09	0.12	0.07	0.01
Lt_s	0.00	191.97	252.26	330.35	22.12	0.03	0.04	0.05	0.00
Tba	13.41	0.00	1.49	8.39	5.44	0.00	0.03	0.18	0.12
Tba_s	0.00	0.39	0.56	0.62	0.00	0.02	0.02	0.03	0.00
	312 plots								
	$\hat{\gamma}$								
	$\hat{\alpha}_s$	$\hat{\alpha}_h$	$\hat{\alpha}_v$	$\hat{\alpha}_{c(1)}$	$\hat{\alpha}_{c(2)}$	$\hat{\rho}_h$	$\hat{\rho}_v$	$\hat{\rho}_{c(1)}$	$\hat{\rho}_{c(2)}$
M	6.07	9.47	11.89	8.21	0.00	0.14	0.17	0.12	0.00
M_s	0.00	59.64	43.71	34.95	185.43	0.00	0.00	0.00	0.00
Lt	141.92	36.38	45.71	27.91	5.10	0.09	0.12	0.07	0.01
Lt_s	0.00	192.06	252.23	330.34	21.66	0.03	0.04	0.05	0.00
Tba	13.50	0.00	1.48	8.36	5.43	0.00	0.03	0.18	0.12
Tba_s	0.00	0.39	0.56	0.62	0.00	0.02	0.02	0.03	0.00
	alternative $\hat{\gamma}'$								
	$\hat{\alpha'}_s$	$\hat{\alpha'}_h$	$\hat{\alpha'}_v$	$\hat{\alpha'}_{c(1)}$	$\hat{\alpha'}_{c(2)}$	$\hat{\rho'}_h$	$\hat{\rho'}_v$	$\hat{\rho'}_{c(1)}$	$\hat{\rho'}_{c(2)}$
M	6.07	9.47	11.89	8.21	0.00	0.14	0.17	0.12	0.00
M_s	0.00	58.74	43.05	34.42	182.63	0.00	0.00	0.00	0.00
Lt	141.91	36.38	45.71	27.92	5.10	0.09	0.12	0.07	0.01
Lt_s	0.00	191.77	251.85	329.84	21.63	0.03	0.04	0.05	0.00
Tba	13.49	0.00	1.48	8.36	5.43	0.00	0.03	0.18	0.12
Tba_s	0.00	0.39	0.55	0.62	0.00	0.02	0.02	0.03	0.00

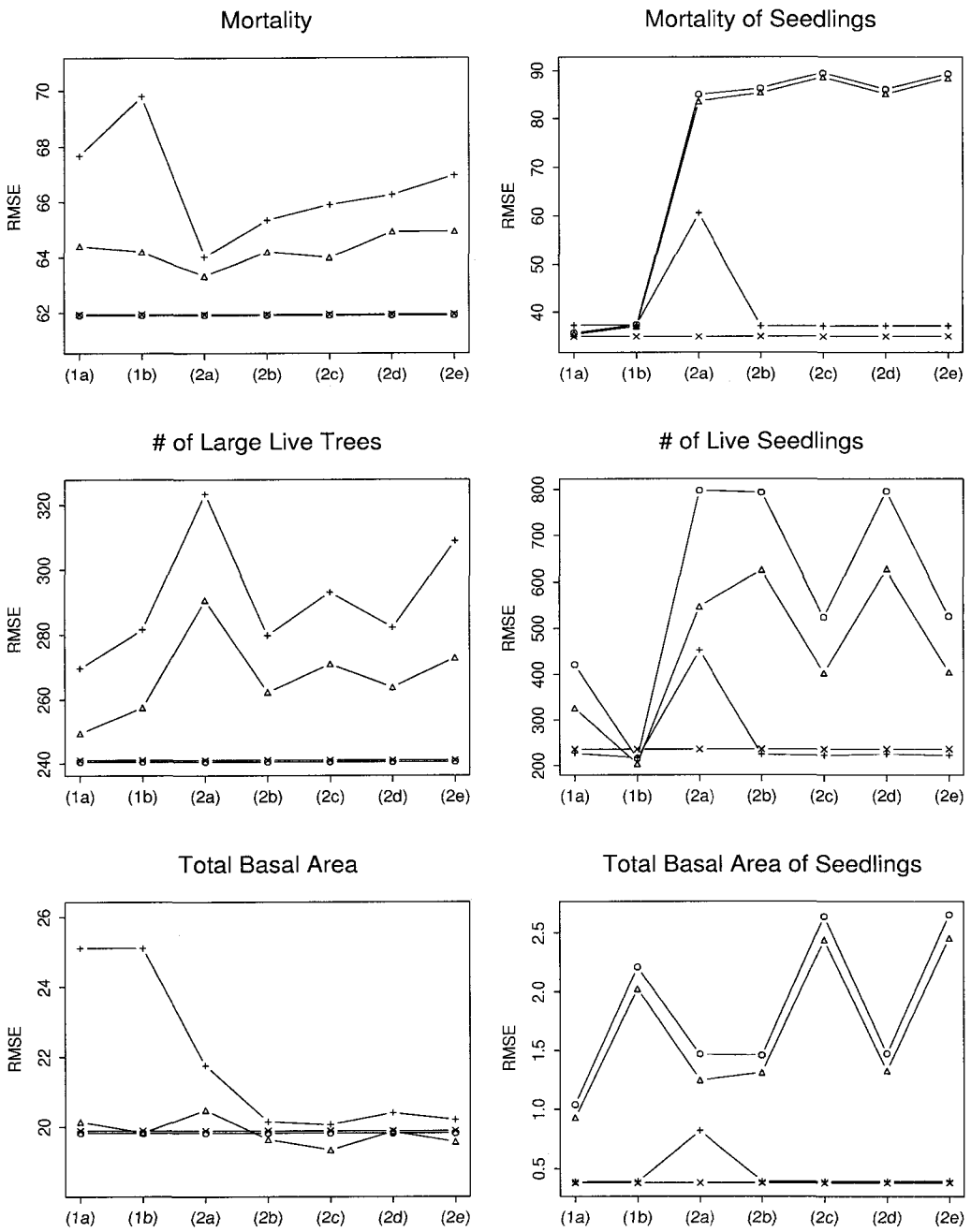
(2e) assume that α's are all different

Figure 6.6: RMSE for the plot-based data fitted by spatial ZIP models



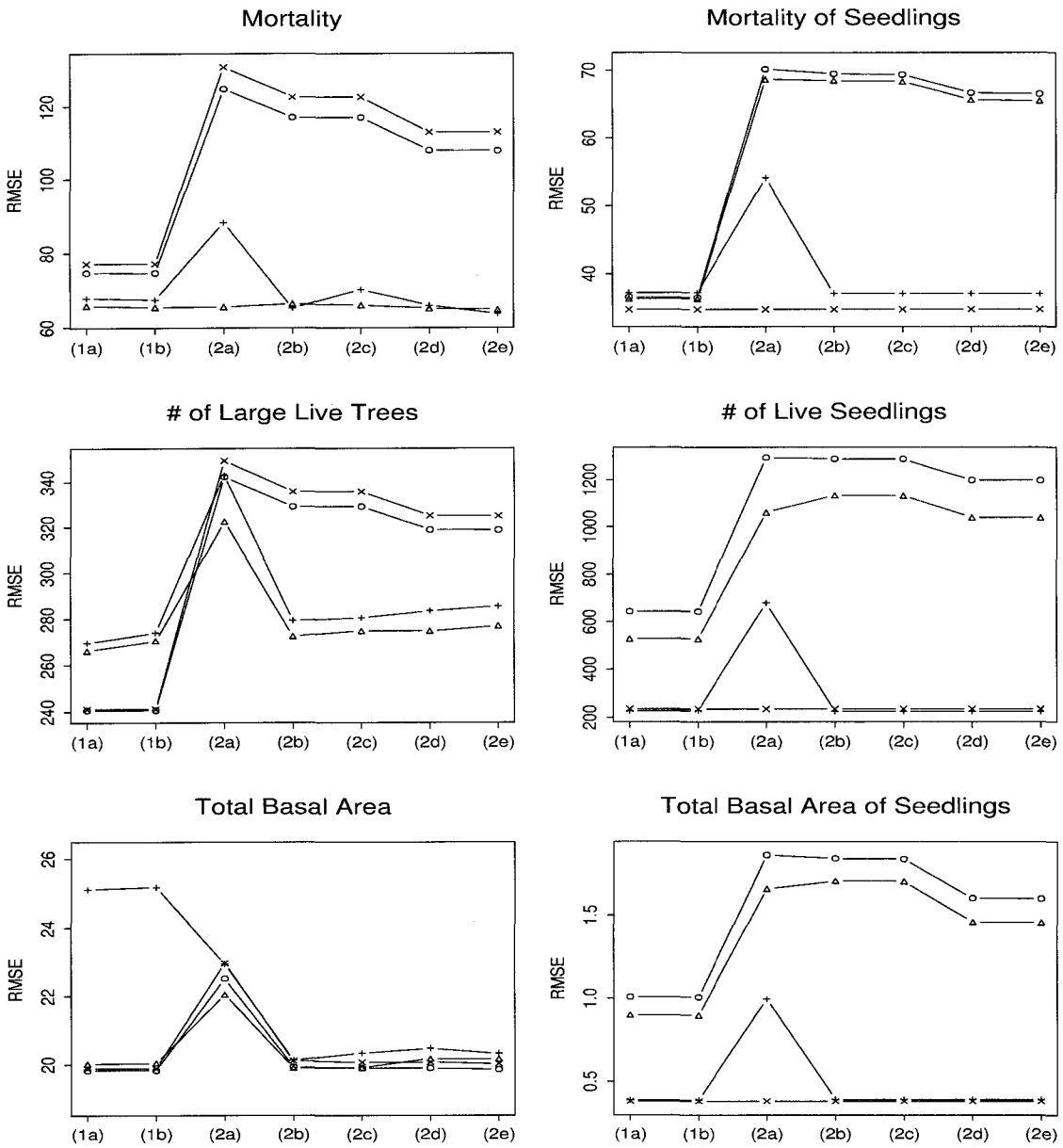
The RMSE is estimated by leave-one-out cross-validation method for 33 plots with 8 neighbors. Note that the marks, $\circ$, $\triangle$, $+$ and x respectively denote the prediction rules 1-4 (see Sec. 6.8.1). The values of x-axis are (1a) assume that $\lambda_h = \lambda_v = \lambda_{hv}$ and $\lambda_{c(1)} = \lambda_{c(2)} = 0$, (1b) assume that $\lambda_{c(1)} = \lambda_{c(2)} = 0$ (2a) assume that $\lambda_h = \lambda_v = \lambda_{c(1)} = \lambda_{c(2)} = \lambda_{hvc}$ (2b) assume that $\lambda_h = \lambda_v = \lambda_{hv}$ and $\lambda_{c(1)} = \lambda_{c(2)} = \lambda_c$ (2c) assume that $\lambda_{c(1)} = \lambda_{c(2)} = \lambda_c$ (2d) assume that $\lambda_h = \lambda_v = \lambda_{hv}$ (2e) assume that $\lambda_h, \lambda_v, \hat{\lambda}_{c(1)}$, are $\hat{\lambda}_{c(2)}$ are all different.

Figure 6.7: RMSE for the plot-based data fitted by spatial ZINB models



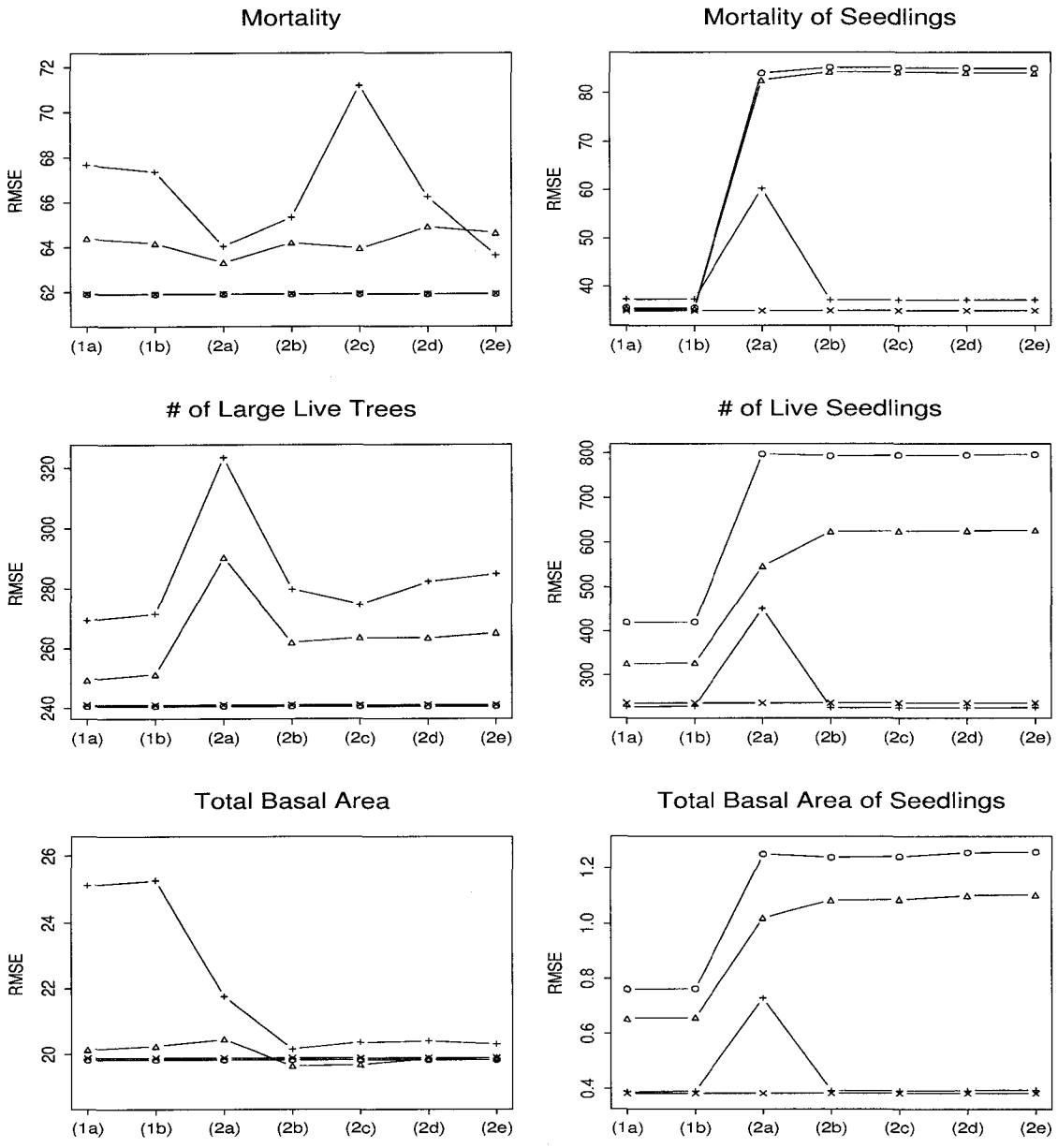
The RMSE is estimated by leave-one-out cross-validation method for 33 plots with 8 neighbors. Note that the marks, $\circ$, $\triangle$, $+$ and x respectively denote the prediction rules 1-4 (see Sec. 6.8.2). The values of x-axis are (1a) assume that $\lambda_h = \lambda_v = \lambda_{hv}$ and $\lambda_{c(1)} = \lambda_{c(2)} = 0$, (1b) assume that $\lambda_{c(1)} = \lambda_{c(2)} = 0$ (2a) assume that $\lambda_h = \lambda_v = \lambda_{c(1)} = \lambda_{c(2)} = \lambda_{hvc}$ (2b) assume that $\lambda_h = \lambda_v = \lambda_{hv}$ and $\lambda_{c(1)} = \lambda_{c(2)} = \lambda_c$ (2c) assume that $\lambda_{c(1)} = \lambda_{c(2)} = \lambda_c$ (2d) assume that $\lambda_h = \lambda_v = \lambda_{hv}$ (2e) assume that $\lambda_h, \lambda_v, \hat{\lambda}_{c(1)}$, are $\hat{\lambda}_{c(2)}$ are all different.

Figure 6.8: RMSE for the plot-based data fitted by spatial ZIE models



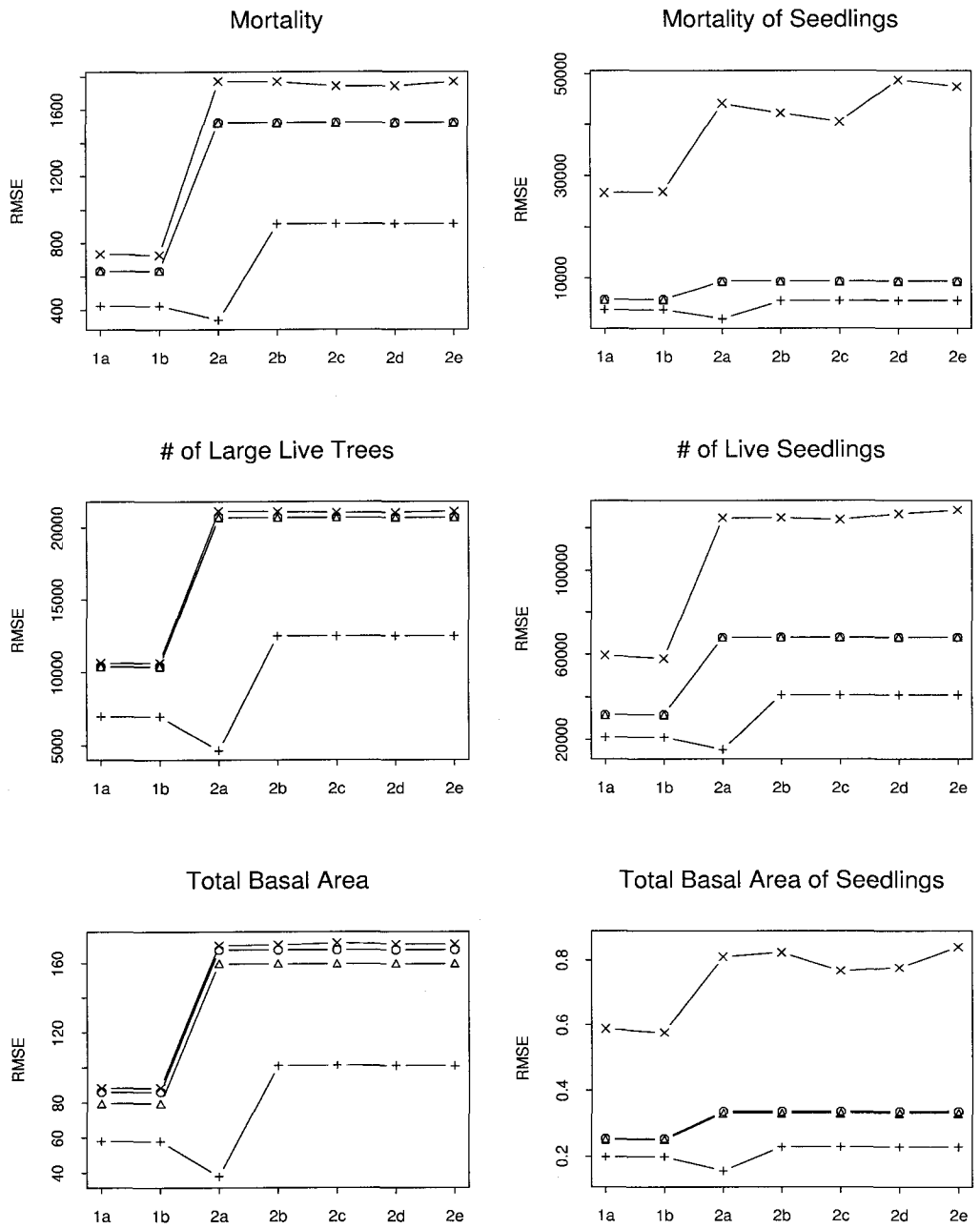
The RMSE is estimated by leave-one-out cross-validation method for 33 plots with 8 neighbors. Note that the marks, $\circ$, $\triangle$, $+$ and x respectively denote the prediction rules 1-4 (see Sec. 6.8.3). The values of x-axis are (1a) assume that $\beta_h = \beta_v = \beta_{hv}$ and $\beta_{c(1)} = \beta_{c(2)} = 0$, (1b) assume that $\beta_{c(1)} = \beta_{c(2)} = 0$, (2a) assume that $\beta_h = \beta_v = \beta_{c(1)} = \beta_{c(2)} = \beta_{hvc}$, (2b) assume that $\beta_h = \beta_v = \beta_{hv}$ and $\beta_{c(1)} = \beta_{c(2)} = \beta_c$, (2c) assume that $\beta_{c(1)} = \beta_{c(2)} = \beta_c$, (2d) assume that $\beta_h = \beta_v = \beta_{hv}$, (2e) assume that $\beta_h, \beta_v, \hat{\beta}_{c(1)}$, and $\hat{\beta}_{c(2)}$ are all different.

Figure 6.9: RMSE for the plot-based data fitted by spatial ZIG models



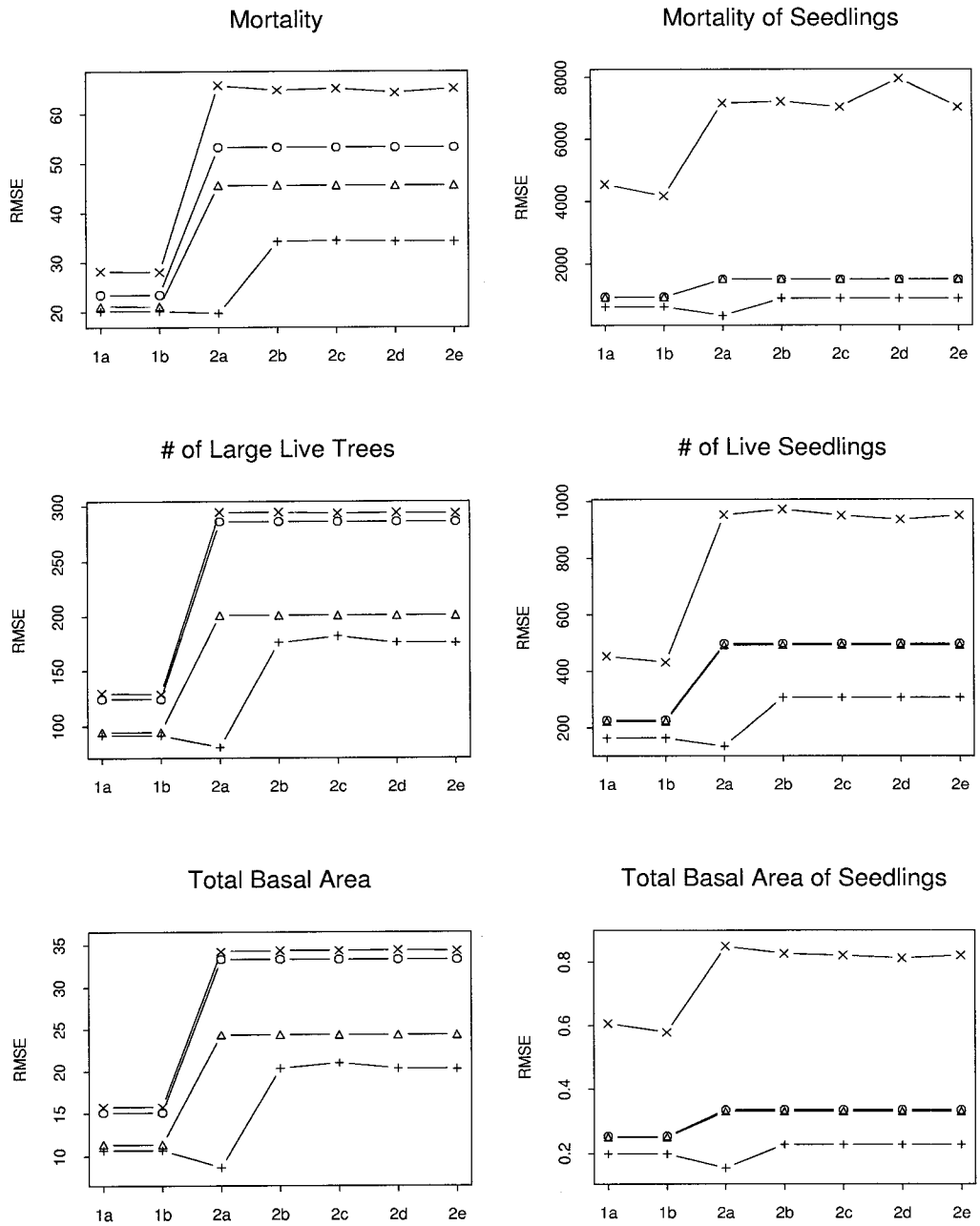
The RMSE is estimated by leave-one-out cross-validation method for 33 plots with 8 neighbors. Note that the marks, $\circ$, $\triangle$, $+$ and x respectively denote the prediction rules 1-4 (see Sed. 6.8.4). The values of x-axis are (1a) assume that $\alpha_h = \alpha_v = \alpha_{hv}$ and $\alpha_{c(1)} = \alpha_{c(2)} = 0$, (1b) assume that $\alpha_{c(1)} = \alpha_{c(2)} = 0$, (2a) assume that $\alpha_h = \alpha_v = \alpha_{c(1)} = \alpha_{c(2)} = \alpha_{hvc}$, (2b) assume that $\alpha_h = \alpha_v = \alpha_{hv}$ and $\alpha_{c(1)} = \alpha_{c(2)} = \alpha_c$, (2c) assume that $\alpha_{c(1)} = \alpha_{c(2)} = \alpha_c$, (2d) assume that $\alpha_h = \alpha_v = \alpha_{hv}$, (2e) assume that $\alpha_h, \alpha_v, \hat{\alpha}_{c(1)}$, and $\hat{\alpha}_{c(2)}$ are all different.

Figure 6.10: RMSE for the subplot-based data fitted by spatial ZIP models



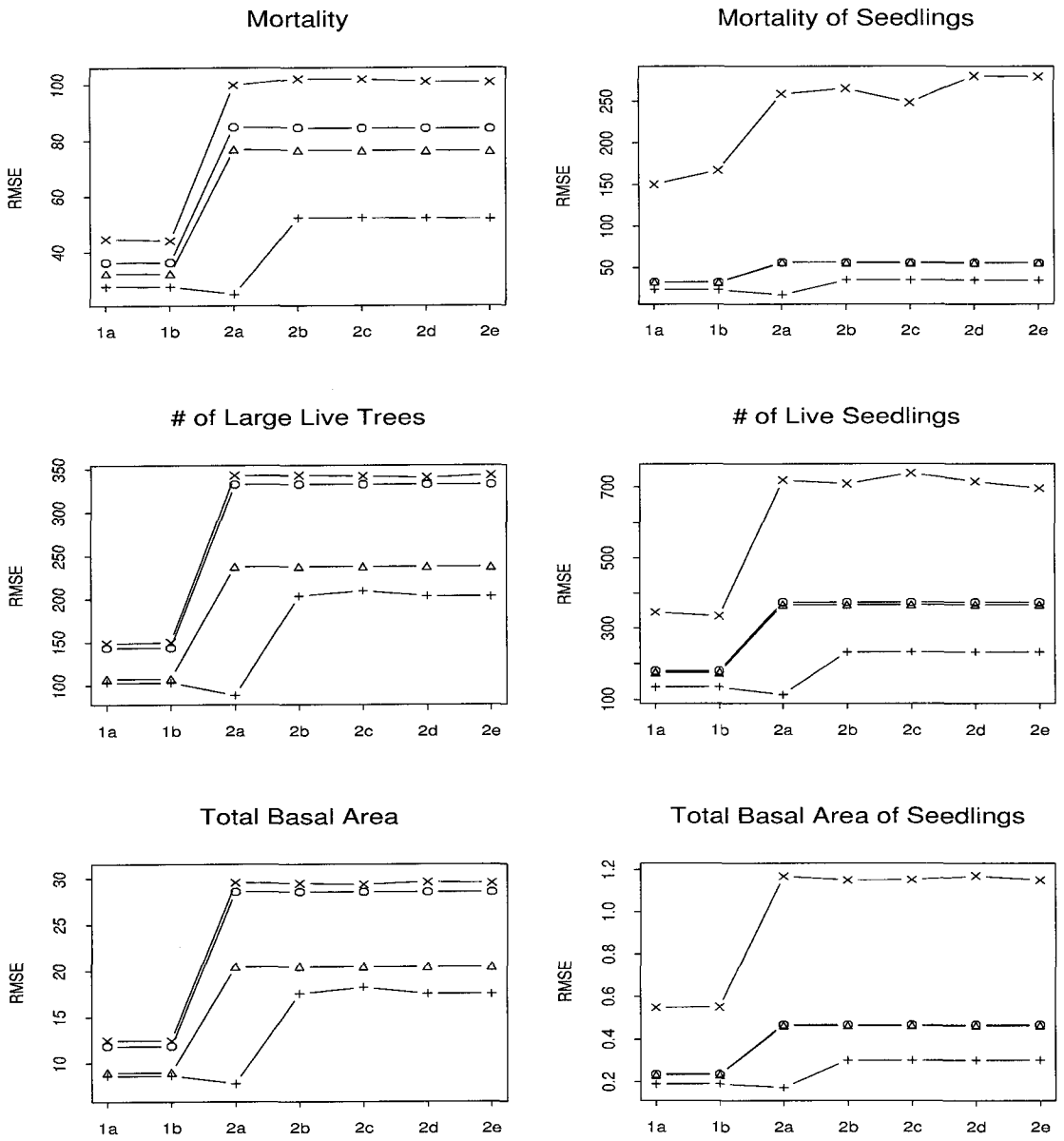
The RMSE is estimated by leave-one-out cross-validation method for 33 plots with 8 neighbors. Note that the marks, $\circ$, $\triangle$, $+$ and x respectively denote the prediction rules 1-4 (see Sec. 6.8.1). The values of x-axis are (1a) assume that $\lambda_h = \lambda_v = \lambda_{hv}$ and $\lambda_{c(1)} = \lambda_{c(2)} = 0$, (1b) assume that $\lambda_{c(1)} = \lambda_{c(2)} = 0$ (2a) assume that $\lambda_h = \lambda_v = \lambda_{c(1)} = \lambda_{c(2)} = \lambda_{hvc}$ (2b) assume that $\lambda_h = \lambda_v = \lambda_{hv}$ and $\lambda_{c(1)} = \lambda_{c(2)} = \lambda_c$ (2c) assume that $\lambda_{c(1)} = \lambda_{c(2)} = \lambda_c$ (2d) assume that $\lambda_h = \lambda_v = \lambda_{hv}$ (2e) assume that $\lambda_h, \lambda_v, \hat{\lambda}_{c(1)}$, are $\hat{\lambda}_{c(2)}$ are all different.

Figure 6.11: RMSE for the subplot-based data fitted by spatial ZINB models



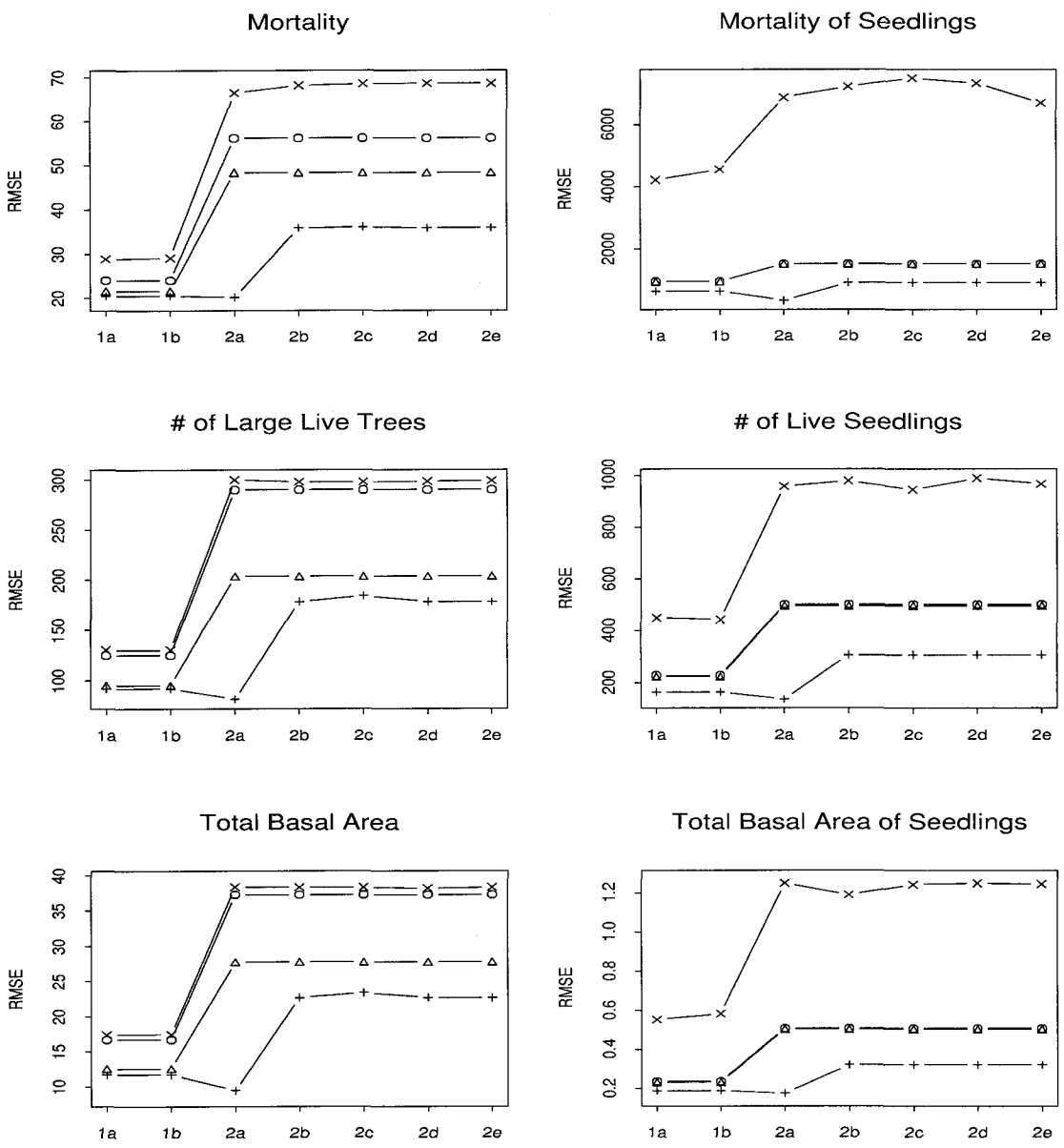
The RMSE is estimated by leave-one-out cross-validation method for 33 plots with 8 neighbors. Note that the marks, ○, △, + and x respectively denote the prediction rules 1-4 (see Sec. 6.8.2). The values of x-axis are (1a) assume that $\lambda_h = \lambda_v = \lambda_{hv}$ and $\lambda_{c(1)} = \lambda_{c(2)} = 0$, (1b) assume that $\lambda_{c(1)} = \lambda_{c(2)} = 0$ (2a) assume that $\lambda_h = \lambda_v = \lambda_{c(1)} = \lambda_{c(2)} = \lambda_{hvc}$ (2b) assume that $\lambda_h = \lambda_v = \lambda_{hv}$ and $\lambda_{c(1)} = \lambda_{c(2)} = \lambda_c$ (2c) assume that $\lambda_{c(1)} = \lambda_{c(2)} = \lambda_c$ (2d) assume that $\lambda_h = \lambda_v = \lambda_{hv}$ (2e) assume that $\lambda_h, \lambda_v, \hat{\lambda}_{c(1)}$, are $\hat{\lambda}_{c(2)}$ are all different and $\lambda_i = \alpha\zeta_i$.

Figure 6.12: RMSE for the subplot-based data fitted by spatial ZIE models



The RMSE is estimated by leave-one-out cross-validation method for 33 plots with 8 neighbors. Note that the marks, o, $\triangle$, + and x respectively denote the prediction rules 1-4 (see Sec. 6.8.3). The values of x-axis are (1a) assume that $\beta_h = \beta_v = \beta_{hv}$ and $\beta_{c(1)} = \beta_{c(2)} = 0$, (1b) assume that $\beta_{c(1)} = \beta_{c(2)} = 0$, (2a) assume that $\beta_h = \beta_v = \beta_{c(1)} = \beta_{c(2)} = \beta_{hvc}$, (2b) assume that $\beta_h = \beta_v = \beta_{hv}$ and $\beta_{c(1)} = \beta_{c(2)} = \beta_c$, (2c) assume that $\beta_{c(1)} = \beta_{c(2)} = \beta_c$, (2d) assume that $\beta_h = \beta_v = \beta_{hv}$, (2e) assume that $\beta_h, \beta_v, \hat{\beta}_{c(1)}$, and $\hat{\beta}_{c(2)}$ are all different.

Figure 6.13: RMSE for the subplot-based data fitted by spatial ZIG models



The RMSE is estimated by leave-one-out cross-validation method for 33 plots with 8 neighbors. Note that the marks, ○, △, + and x respectively denote the prediction rules 1-4 (see Sed. 6.8.4). The values of x-axis are (1a) assume that $\alpha_h = \alpha_v = \alpha_{hv}$ and $\alpha_{c(1)} = \alpha_{c(2)} = 0$, (1b) assume that $\alpha_{c(1)} = \alpha_{c(2)} = 0$, (2a) assume that $\alpha_h = \alpha_v = \alpha_{c(1)} = \alpha_{c(2)} = \alpha_{hvc}$, (2b) assume that $\alpha_h = \alpha_v = \alpha_{hv}$ and $\alpha_{c(1)} = \alpha_{c(2)} = \alpha_c$, (2c) assume that $\alpha_{c(1)} = \alpha_{c(2)} = \alpha_c$, (2d) assume that $\alpha_h = \alpha_v = \alpha_{hv}$, (2e) assume that $\alpha_h, \alpha_v, \hat{\alpha}_{c(1)}$, and $\hat{\alpha}_{c(2)}$ are all different.

7. SPATIAL MODELS WITH AUXILIARY DATA

In this chapter, we propose the following procedures to incorporate the auxiliary data: (1) general linear model, (2) spatial regression model with $\boldsymbol{\mu}$=$\mathbf{X}$ $\boldsymbol{\beta}$ where $\mathbf{X}$ is an $N \times (k+1)$ matrix and $\boldsymbol{\beta}$ is a $(k+1)$-dimensional vector of regression parameters, and (3) spatial models with similar neighbors defined by the similarity function formed by the auxiliary data.

7.1 General Linear Model (GLM)

If there exists a significant spatial dependence among the response variables, including more auxiliary data might account for the spatial dependence. First we will deal with the general linear regression model without spatial dependence. The model is written in the matrix form of

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{7.1}$$

where $\mathbf{Y} = (Y_1, \ldots, Y_N)'$, $\mathbf{X} = (\mathbf{1}, \mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(k)})'$, $\mathbf{X}^{(j)} = (X_1^{(j)}, \ldots, X_N^{(j)})'$ and $X_i^{(j)}$ denotes the jth covariate at plot i, $\boldsymbol{\beta}$ is a vector of $k+1$ regression parameters, $\boldsymbol{\varepsilon} \sim \text{MVN}(\mathbf{0}, \sigma^2\,\mathbf{I})$; where the $X_i^{(j)}$ might be Aspect (Asp), Elevation (Elev), Slope, TM bands 1-5 and 7 (Band_1-Band_5, and Band_7), Tasseled Cap band data (Cap_1-Cap_3), ratio band data (Ratio_1-Ratio_3), etc.

The ordinary least square estimate $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is found by minimizing

$$\sum_{i=1}^{N}(Y_i - \beta_0 - \sum_{j=1}^{k}\beta_j X_i^{(j)})^2$$

with respect to $\boldsymbol{\beta}$. The solution is given by

$$\hat{\boldsymbol{\beta}}_{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \tag{7.2}$$

The residuals in the model (7.1) are given by

$$\hat{\varepsilon}_i = Y_i - \hat{\beta}_0 - \sum_{j=1}^{k} \hat{\beta}_j X_i^{(j)},\ i = 1, \ldots, N.$$

The variance σ^2 is estimated by

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{N-(k+1)}(\mathbf{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) \\ &= \frac{1}{N-(k+1)} \sum_{i=1}^{N} \hat{\varepsilon}_i^2. \end{aligned} \tag{7.3}$$

The MLE of σ^2 is given by

$$\hat{\sigma}^2_{MLE} = \frac{1}{N} \sum_{i=1}^{N} \hat{\varepsilon}_i^2 \text{ under the Gaussian model.} \tag{7.4}$$

The Gaussian likelihood function for $\mathbf{Y} = (Y_1, \ldots, Y_N)'$ is given by

$$L(\boldsymbol{\beta}, \sigma^2) = \frac{1}{(2\pi)^{N/2}\sigma^N}\, exp\left[\frac{-1}{2\sigma^2}(\mathbf{Y} - \boldsymbol{X}\boldsymbol{\beta})'(\mathbf{Y} - \boldsymbol{X}\boldsymbol{\beta})\right].$$

The log-likelihood is

$$l(\boldsymbol{\beta}, \sigma^2) = -\frac{N}{2} log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(\mathbf{Y} - \boldsymbol{X}\boldsymbol{\beta})'(\mathbf{Y} - \boldsymbol{X}\boldsymbol{\beta}).$$

The MLE (maximum likelihood) estimator is found by maximizing $l(\cdot)$ with respect to $\boldsymbol{\beta}$ and σ^2.

$$\begin{aligned} \frac{\partial l(\boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2} &= -\frac{N}{2\sigma^2} - \frac{1}{2\sigma^4}(\mathbf{Y} - \boldsymbol{X}\boldsymbol{\beta})'(\mathbf{Y} - \boldsymbol{X}\boldsymbol{\beta}) \\ &= 0. \end{aligned}$$

The equality holds if and only if $\hat{\sigma}^2$=$(\mathbf{Y}$-$\boldsymbol{X}\boldsymbol{\beta})'$ $(\mathbf{Y}$-$\boldsymbol{X}\boldsymbol{\beta})$. The MLE of $\boldsymbol{\beta}$ is then found by minimizing $l(\boldsymbol{\beta}\ ,\ \hat{\sigma}^2)$, i.e. $\hat{\boldsymbol{\beta}}_{MLE}$=$\hat{\boldsymbol{\beta}}_{OLS}$.

To test whether there is a regression relation between the dependent variable $\mathbf{X}$ and $\mathbf{Y}$, we choose between the alternatives: $H_0 : \beta_1 = \ldots = \beta_k = 0$ versus H_a: not all β_i, $i = 1, \ldots, k$ equal zero. One can use the F-test, where

$$F = \frac{R^2/k}{(1-R^2)/(N-k-1)} \sim F_{k,N-k-1}, \text{ under } H_0 \text{ is true}$$

where

$$R^2 = \frac{\hat{\boldsymbol{\beta}} \boldsymbol{X} \boldsymbol{X}' \hat{\boldsymbol{\beta}}' - \overline{\boldsymbol{Y}}^2}{\mathbf{Y}'\mathbf{Y} - N\overline{\boldsymbol{Y}}^2}$$

To test $H_0 : \beta_j = 0$ vs. $H_a : \beta_j \neq 0, j = 1, \ldots, k$, we may use the t-test for t-statistic

$$t_j = \frac{\hat{\beta}_j}{\hat{\sigma} a_{ii}^{1/2}}$$

where a_{ii} is the ith diagonal element of $(\mathbf{X}'\mathbf{X})^{-1}$. The decision rule at level α is: if $\mid t_j \mid \leq t(1 - \alpha/2, N - (k+1))$, do not reject H_0; otherwise reject H_0.

7.2 Most Similar Neighbor Model (MSN)

Moeur and Stage (1995) described the "most similar neighbor" (MSN) inference procedure, in which the value of the most similar neighbor is imputed as the value for the non-sampled plot. The MSN inference is a special case of the k-nearest neighbors (k-NN) method which decides the nearest neighbor in Mahalanobis distances, considering $k = 1$. The similarity is defined by the similarity function

$$D_{ij} = (\mathbf{X}_i - \mathbf{X}_j) G \Lambda^2 G' (\mathbf{X}_i - \mathbf{X}_j)' \tag{7.5}$$

where G are canonical coefficients of the $\mathbf{X}$'s, Λ is a vector of the correlations between the $\mathbf{X}$'s and $\mathbf{Y}$'s from sampled plots, $\mathbf{Y}$ might be any of the forest variables, such as mortality (M), mortality of seedlings (M_s), total numbers of live trees (Lt), Lt of seedlings (Lt_s), total basal area (Tba), Tba of seedlings (Lt_s), and $\mathbf{X}_i = (X_i^{(1)}, \ldots, X_i^{(k)})$ is any of the auxiliary variables, i.e. Aspect, Elevation, Slope, TM bands 1-5 and 7 at plot i. The canonical correlation seeks a linear combination of one set of $\mathbf{Y}$ and a second set of $\mathbf{X}$ such that the correlation is maximimized. And hence a similarity function D_{ij} using such a distance function related to the canonical correlation concept is proposed. First, using the canonical correlation analysis for the N sampled plots, the G and Λ are obtained.

7.3 Spatial Regression Model (SRM) with μ=Xβ

In Chapter 4, we discussed the simple spatial model ("Phase I" model)

(See (4.1)) $$\mathbf{Y} = \boldsymbol{\mu} + B\boldsymbol{\varepsilon},$$

where $\mathbf{Y} = (Y_1, \ldots, Y_N)'$, $\boldsymbol{\mu}=(\mu, \ldots, \mu)'$, $\boldsymbol{\varepsilon} \sim$ MVN($\mathbf{0}$, σ^2 $\mathbf{I}$) and $\boldsymbol{\Gamma} = \mathbf{I} + \rho_1 \mathbf{H} + \rho_2 \mathbf{V} + \rho_3 \mathbf{C}^{(1)} + \rho_4 \mathbf{C}^{(2)} = B'B$.

This section will bring in the auxiliary data in the selected "Phase I" model and then "Phase II" model is written as

$$\begin{aligned} \mathbf{Y} &= \boldsymbol{\mu} + B\boldsymbol{\varepsilon} \\ &\equiv \mathbf{X}\boldsymbol{\beta} + B\boldsymbol{\varepsilon} \end{aligned} \tag{7.6}$$

where $\mathbf{Y} = (Y_1, \ldots, Y_N)'$, $\mu_i = \beta_0 + \sum_{j=1}^{k} \beta_j X_i^{(j)}$, $X_i^{(j)}$ denotes the jth covariate at plot i, $\boldsymbol{\mu}=(\mu_1, \ldots, \mu_N)'$, $\boldsymbol{\varepsilon} \sim$ MVN($\mathbf{0}$, σ^2 $\mathbf{I}$), $\boldsymbol{\Gamma} = \mathbf{I} + \rho_1 \mathbf{H} + \rho_2 \mathbf{V} = B'B$, and $X_i^{(j)}$ is the auxiliary data such as Aspect, Elevation, Slope, TM bands 1-5 and 7 at plot i. Since B is invertible, we assume that $M \equiv (m_{ij}) = \mathbf{I} - B^{-1}$. In individual observation, equation (7.6) is equivalent to

$$Y_i = \beta_0 + \sum_{j=1}^{k} \beta_j X_i^{(j)} + \sum_{l=1}^{N} m_{il}[Y_l - (\beta_0 + \sum_{j=1}^{k} \beta_j X_l^{(j)})] + \varepsilon_i,\ i = 1, \ldots, N, \tag{7.7}$$

where m_{ii} are zeros for $i = 1, \ldots, N$. The value at plot i depends on the mean of its distribution at that plot, plus a weighted sum of the errors at the neighboring plots.

The Gaussian likelihood of $\mathbf{Y} = (Y_1, \ldots, Y_N)'$ is

$$L(\boldsymbol{\beta}, \sigma^2, \rho_1, \rho_2) = (2\pi\sigma^2)^{-N/2} |\ \boldsymbol{\Gamma}\ |^{-1/2} exp\left[\frac{-1}{2\sigma^2}(\mathbf{Y} - \boldsymbol{X\beta})'(\mathbf{Y} - \boldsymbol{X\beta})\right]$$

where $\boldsymbol{\Gamma} = \mathbf{I} + \rho_1 \mathbf{H} + \rho_2 \mathbf{V} + \rho_3 \mathbf{C}^{(1)} + \rho_4 \mathbf{C}^{(2)} = B'B$. Then the log-likelihood of $\mathbf{Y}$ is

$$\begin{aligned} l(\boldsymbol{\beta}, \sigma^2, \rho_i; i = 1, \ldots, 4) &= log\left(L(\boldsymbol{\beta}, \sigma^2, \rho_i; i = 1, \ldots, 4)\right) \\ &= -\frac{N}{2} log(2\pi\sigma^2) - \frac{1}{2} log(|\ \boldsymbol{\Gamma}\ |) - \frac{1}{2\sigma^2}(\mathbf{Y} - \boldsymbol{X\beta})'(\mathbf{Y} - \boldsymbol{X\beta}) \end{aligned}$$

Maximizing with respect to μ and σ^2, we find that

$$\hat{\boldsymbol{\beta}}_{MLE} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \text{ and } \hat{\sigma}^2_{MLE} = \frac{(\mathbf{Y} - \boldsymbol{X\beta})'\boldsymbol{\Gamma}^{-1}(\mathbf{Y} - \boldsymbol{X\beta})}{N} \tag{7.8}$$

The log-likelihood is therefore reduced to

$$l(\boldsymbol{\beta},\, \sigma^2, \rho_1, \rho_2) = -\frac{N}{2}log(2\pi\hat{\sigma}^2) - \frac{1}{2}log(\mid \boldsymbol{\Gamma} \mid) - \frac{N}{2}$$

7.4 Alternative Spatial Zero-Inflated Models with Auxiliary Data

The model (6.6) with p and λ depending on covariates forms the basis of the zero-inflated Poisson (ZIP) regression framework proposed by Lambert (1992). In ZIP regression, the responses $\mathbf{Y} = (Y_1, \ldots, Y_N)'$ are independent and

$$Y_i = \begin{cases} 0 & \text{with probability } 1 - p_i \\ \text{Poisson}(\lambda_i) & \text{with probability } p_i \end{cases}$$

and the parameters $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)'$ and $\boldsymbol{p} = (p_1, \ldots, p_N)'$ satisfy

$$log(\boldsymbol{\lambda}) = \mathbf{X}\boldsymbol{\beta} \tag{7.9}$$

$$logit(\boldsymbol{p}) = log\left(\boldsymbol{p}/(1 - \boldsymbol{p})\right) = \boldsymbol{G\gamma} \tag{7.10}$$

for covariate matrices $\mathbf{X}$ and $\boldsymbol{G}$.

Analogously, we will incorporate the covariates into p and λ's in the model (6.16). Thus a spatial zero-inflated Poisson (SZIP) regression model is given by:

$$\begin{aligned} Y_i = {} & D_0\,(1 - B_i) + (P_i^{(s)} + P_i^{(hr)} + P_i^{(hl)} + P_i^{(vu)} + P_i^{(vd)} \\ & + P_i^{(c_1)} + P_i^{(c_2)} + P_i^{(c_3)} + P_i^{(c_4)} + 1)\, B_i, \end{aligned} \qquad \text{(See (6.16))}$$

and the parameters p and $\boldsymbol{\lambda} = (\lambda_h, \lambda_v, \lambda_{c^{(1)}}, \lambda_{c^{(2)}})'$ satisfy

$$log(\boldsymbol{\lambda}) = \mathbf{X}\boldsymbol{\beta} \tag{7.11}$$

$$logit(p) = log\left(p/(1 - p)\right) = \boldsymbol{G\gamma} \tag{7.12}$$

for covariate matrices $\mathbf{X}$ and $\boldsymbol{G}$. Spatial zero-inflated exponential (SZIE) regression, spatial zero-inflated gamma (SZIG) regression, and spatial zero-inflated negative binomial (SZINB) regression models can be expressed in a similar way to a SZIP regression model.

The moment estimation we used for spatial zero-inflated (SZI) models does not work for spatial ZI models with auxiliary data. In future work, we will examine a better estimation method such as maximum likelihood estimation for SZI models.

7.5 Results

We included a single regressor in the linear regression model (7.1) for Tba with and without the isolated plot observation. Since the results are about the same, we only summarize the results for Tba without the isolated plot ($N = 312$). The results for Tba are summarized in Table 7.1. The predictor variables Elev, Band_1-Band_3, Band_5, Band_7, Cap_1, Cap_2, Cap_3, Ratio_1, and Ratio_2 are significant at level 0.05. We included all predictors except for Tasseled Cap variables in the model (7.1) for Tba and the result is summarized in Table 7.2. In selecting explanatory variables, we used the backward elimination procedure which starts with a complete set, and at each step drops the independent variable that gives the smallest increase in the residual sum of squares. The basic steps are (1) first dropping the variable with the highest p-value, (2) then dropping the second highest one, until only significant variables are left. The reduced model (7.1) of Tba without the isolated plot is (S.D. values in parentheses)

$$\begin{aligned} \hat{Y}_i &= 85.71 + 11.01\,\text{Asp}_S - 0.34\,\text{Band}_4 - 33.23\,\text{Ratio}_2 \\ &\quad (7.08) \qquad (3.91) \qquad (0.05) \qquad (8.99) \end{aligned}$$

$$
\begin{aligned}
R^2 &= 0.1423; \qquad \text{residual S.D.=22.68} \\
-Log(likelihood) &= -1414.61 \\
AIC &= 2837.21 \\
AICC &= 2837.34 \\
Schwartz &= 2852.18
\end{aligned}
$$

Also, we include the location coordinates (s_1, s_2) in the model above. From Table 7.3 (b) and (c), we note that the coefficients for the location coordinates (s_1, s_2) are not significantly different from zero. But we still consider the models as candidate models since the spatial effect was considered.

In this section we examine the "Phase II" model under the explanatory variables we chose for model 7.1(a). The "Phase II" model, Eqn. 7.6 can be rewritten as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\delta} \tag{7.13}$$

where $\boldsymbol{\delta}$ = B $\boldsymbol{\varepsilon}$ is MVN($\mathbf{0}$, $\sigma^2\Gamma$) and $\Gamma = B'B$.

Before we fit the model to the residuals, $\boldsymbol{\delta}$, the sample correlation between different directional neighbors are calculated and listed in Table 7.4. From Table 7.4, we can see that the sample correlations of $\boldsymbol{\delta}$, γ_h and γ_v are very small, one is negative and the other one is positive; and γ_{hv} yields the value of -0.0097. As to the 2-unit-apart sample correlations, γ_h is twice γ_v. Thus, we will consider Scheme (2d), i.e. $\rho_1 = \rho_2$, ρ_3, ρ_4, in the "Phase II" model. The result for Tba from Model 7.6 is summarized in Table 7.3.

In the MSN procedure, when D_{ij} is minimized at $j = l$, the value Y_l is imputed as the prediction for plot i. The first MSN model we consider is the model that includes the predictor variables, Elev, Band_1-Band_3,Band_5, Band_7, Cap_1, Cap_2, Cap_3, Ratio_1, Ratio_2 and Ratio_3. Then we also consider the model

with all variables in the previous MSN model, s_1 and s_2, and the model with an additional variable $s_1 \times s_2$.

Similarly, we include a single regressor in the linear regression model (7.1) for Lt data without the isolated plot observation, i.e. $N = 312$. From Table 7.3, we note the coefficients for the location coordinates are not significant at level $\alpha = .05$. From Table 7.4, we can see that the sample correlations of the residuals for Lt, γ_h and γ_v are about the same. As to the 2-unit-apart sample correlations, γ_h is almost triple γ_v. Thus, we will consider Scheme (2d), i.e. $\rho_1 = \rho_2$, ρ_3, ρ_4, in the "Phase II" model. Also, we consider an alternative one with scheme (2b) and then one less parameter is estimated.

The results for M are summarized in Table 7.10. The predictor variables Elev, Band_2-Band_5, Band_7, and Cap_1, are significant at level 0.05. We included all predictors except for Tasseled Cap variables in the model (7.1) for M and the result is summarized in Table 7.11. The backward elimination procedure was used in the model selection.

$$\begin{aligned} \hat{Y}_i &= 100.37 + 15.95\,\text{Asp}_{sin} + 0.08\,\text{Elev} - 1.89\,\text{Band}_7 - 7.26\,\text{Ratio}_1 \\ &= (18.12) \qquad (5.55) \qquad (0.03) \qquad (0.64) \qquad (3.08) \\ R^2 &= 0.09; \qquad \text{residual S.D.=64.97} \end{aligned}$$

From Table 7.3, we note that the coefficients for the location coordinates are not significant at level $\alpha = .05$. From Table 7.4, we can see that the sample correlations of the residuals for Lt, γ_h and γ_v are about the same. As to the 2-unit-apart sample correlations, γ_h is almost triple γ_v. Thus, we will consider Scheme (2d), i.e. $\rho_1 = \rho_2$, ρ_3, ρ_4, in the "Phase II" model. Also, we consider an alternative one with scheme (2b) because one parameter less has to be estimated.

From Table 7.4, we note that the sample correlations of the residuals for M, are γ_h=0.0989 and γ_v=0.1535. As to the 2-unit-apart sample correlations, γ_h is ten times as much as γ_v, but γ_h is negative. Thus, we will consider Scheme (2d), i.e. $\rho_1 = \rho_2, \rho_3, \rho_4$, in the "Phase II" model. The difference between γ_h and γ_v can not be ignored although there might be bias in the estimated parameters. And hence, an alternative spatial regression model was considered; that is Model 7.6 with scheme (2e).

Table 7.6 summarizes the results of applying a simple regression model with a single predictor and Table 7.7 shows the results of fitting Lt by the general linear model with all predictors variables. Finally, the reduced Model 7.1 for Lt is summarized in Table 7.8.

7.5.1 Model selection

In the model selection we used the leave-one-out Cross-Validation (1CV) for those 33 plots with all 8 neighbors to compare for all candidate models. The results for Tba are summarized in Table (7.5).

We note that the RMSE of Tba data for models (7.1) and (7.6) are apparently smaller than for MSN, but the difference between the two models is slight. When we look at the statistics for other methods AIC, BIC, and AICC, they all suggest model (7.6) is better, i.e. has the smallest values of those three statistics. Compared to the estimated RMSE of Tba prediction based on the simple spatial model, i.e. model (4.1), with correlation scheme (2b), the estimated RMSE of Tba predictions based on model (7.6) is slightly greater (21.92 vs. 19.65), but the estimated AIC, AICC, and BIC based on the selected model (7.6), are respectively 2826.36, 2852.54, and 2826.73, versus those based on the selected model (4.1) without auxiliary data, 2868.10, 2868.21, and 2876.54. For Tba, spatial multi-

variate regression models do not show improvement based on reduced RMSE, but do show slight improvement based on AIC, AICC, and BIC.

The estimated RMSE, AIC, AICC, and BIC for Lt, are summarized in Table 7.9. We can see that RMSEs from models (7.1) and (7.6) are approximately 25% smaller than those from the MSN procedure, Model(7.5). For the selected model without auxiliary information in Chapter 4, the estimated RMSE, AIC, AICC, and BIC, are 62.61, 3495.13, 3495.17, and 3502.21. The estimates for (7.6) in Table 7.9 are slightly less, so the spatial multivariate regression models incorporating auxiliary information do show improvement based on all selected model selection criteria.

For M, the estimated RMSE, AIC, AICC, and BIC, are summarized in Table 7.13. We can see that RMSEs from models (7.1) and (7.6) are approximately 22% smaller than those from the MSN procedure, Model (7.5). In Chapter 4, the simple spatial model with correlation matrix scheme (2a) was suggested and the corresponding estimated RMSE, AIC, AICC, and BIC of M predictions based on such a model without auxiliary information, are 62.21, 3495.13, 3495.17, and 3502.61, respectively. Compared to the estimates based on model (7.6) in Table 7.13, some are reduced; some are increased, but there is little overall change. In general, the selected spatial regression model with auxiliary data offers no real improvement in making M predictions.

In summary, we don't get much improvement in the estimated cross-validated RMSEs by adding the auxiliary information to our models even though we incorporate spatial dependence. The MSN procedure doesn't work any better than the general linear model and spatial regression model; after adding the location coordinates in the MSN procedure, it doesn't decrease our estimated RMSEs. Moeur and Stage included the variogram as a predictor in the MSN inference and they claimed it reduced in the estimated overall RMSE.

Table 7.1: The estimates of the parameters from the linear regression model (7.1) with one predictor for Tba

	$\hat{\beta}_0$	$s.e.(\hat{\beta}_0)$	$\hat{\beta}_i$	$s.e.(\hat{\beta}_i)$	$\hat{\sigma}$	$\hat{R}^2$	p-value
Asp	43.27	2.80	-0.00	0.01	24.41	0.000	0.933
	(0.000)		(0.933)				
Asp_{sin}	43.53	1.39	0.85	1.97	24.40	0.000	0.667
	(0.000)		(0.667)				
Asp_{cos}	43.51	1.38	-1.62	1.94	24.38	0.002	0.405
	(0.000)		(0.405)				
Asp_{RC}	45.15	2.42	-3.30	3.89	24.38	0.002	0.397
	(0.000)		(0.397)				
Asp_{S}	42.73	2.30	1.59	3.94	24.40	0.002	0.687
	(0.000)		(0.687)				
Elev	37.16	3.02	0.02	0.01	24.20	0.020	0.020
	(0.000)		(0.020)				
Slope	41.99	3.22	0.09	0.18	24.40	0.000	0.611
	(0.000)		(0.611)				
Band_1	119.06	19.40	-1.20	0.31	23.83	0.047	0.000
	(0.000)		(0.000)				
Band_2	86.29	9.70	-1.88	0.42	23.66	0.060	0.000
	(0.000)		(0.000)				
Band_3	68.28	6.02	-1.39	0.33	23.74	0.0545	0.000
	(0.000)		(0.000)				
Band_4	66.48	4.41	-0.27	0.05	23.31	0.089	0.000
	(0.000)		(0.000)				
Band_5	64.52	3.82	-0.47	0.08	23.16	0.10	0.000
	(0.000)		(0.000)				
Band_7	57.53	2.95	-1.17	0.22	23.36	0.084	0.000
	(0.000)		(0.000)				
Cap_1	77.66	5.67	-0.32	0.05	23.03	0.11	0.000
	(0.000)		(0.000)				
Cap_2	54.12	2.58	-0.34	0.07	23.54	0.070	0.000
	(0.000)		(0.000)				
Cap_3	38.97	2.33	0.38	0.16	24.19	0.018	0.018
	(0.000)		(0.018)				
Ratio_1	69.09	5.45	-4.24	1.08	23.83	0.047	0.000
	(0.000)		(0.000)				
Ratio_2	53.54	5.14	-18.97	9.33	23.25	0.013	0.043
	(0.000)		(0.043)				
Ratio_3	41.58	6.71	0.47	1.65	24.41	0.000	0.774
	(0.000)		(0.774)				

Note: the p-values in parenthesis are for testing the hypothesis $H_0 : \beta_i = 0$ and the p-values in the last column are for testing the acceptance of the model.

Table 7.2: The model (7.1) with all predictor variables for Tba

Residual Standard Error = 22.5093, Multiple R-Square = 0.1854
N = 312, F-statistic = 4.8274 on 14 and 297 df, p-value = 0

	coef	std.err	t.stat	p.value
β_0	87.4708	35.7877	2.4442	0.0151
Asp	0.0123	0.0194	0.6363	0.5251
Asp_{sin}	7.0637	3.0246	2.3354	0.0202
Asp_{cos}	-2.8510	1.8602	-1.5326	0.1264
Elev	0.0147	0.0100	1.4646	0.1441
Slope	-0.0544	0.1749	-0.3109	0.7561
Band_1	0.0983	0.6474	0.1519	0.8794
Band_2	2.2970	1.2157	1.8895	0.0598
Band_3	-1.5460	1.2048	-1.2832	0.2004
Band_4	-0.5498	0.2969	-1.8518	0.0650
Band_5	0.6241	0.3967	1.5733	0.1167
Band_7	-1.1340	0.7105	-1.5961	0.1115
Ratio_1	-1.9607	3.9968	-0.4906	0.6241
Ratio_2	-47.8559	27.9697	-1.7110	0.0881
Ratio_3	-2.9812	2.3721	-1.2567	0.2098

The t statistic is for testing if the coefficient is zero and the two sided p-value is for the t statistic.

Table 7.3: The model (7.1) with the best predictor variables and location coordinates (s_1, s_2) for Tba

		Model (7.1)(a)		
	coef	std.err	t.stat	p.value
β_0	85.7143	7.0799	12.1066	0.0000
Asp_S	11.0138	3.9052	2.8203	0.0051
Band_4	-0.3446	0.0517	-6.6682	0.0000
Ratio_2	-33.2385	8.9941	-3.6956	0.0003
		Model (7.1)(b)		
	coef	std.err	t.stat	p.value
β_0	82.6222	8.6993	9.4975	0.0000
s_1	0.1301	0.4038	0.3222	0.7475
s_2	0.0603	0.0761	0.7928	0.4285
Asp_S	10.9550	3.9156	2.7978	0.0055
Band_4	-0.3532	0.0520	-6.7944	0.0000
Ratio_2	-32.6283	9.1746	-3.5564	0.0004
		Model (7.1)(c)		
	coef	std.err	t.stat	p.value
β_0	76.5800	11.3330	6.7573	0.0000
s_1	0.6533	0.7472	0.8744	0.3826
s_2	0.2540	0.2448	1.0376	0.3003
$s_1 s_2$	-0.0158	0.0190	-0.8325	0.4058
Asp_S	11.0840	3.9207	2.8271	0.0050
Band_4	-0.3593	0.0525	-6.8407	0.0000
Ratio_2	-31.5830	9.2647	-3.4090	0.0007

The t statistic is for testing if the coefficient is zero and the two sided p-value is for the t statistic.

Table 7.4: Sample Correlations for the residuals, $\boldsymbol{\delta}$, in Model 7.13

	r_h	r_v	r_{hv}	$r_{c(1)}$	$r_{c(2)}$	r_c
Tba	-0.0320	0.0134	-0.0097	0.1810	0.0961	0.1383
Lt	0.1031	0.0989	0.1010	0.0803	0.0230	0.0515
M	0.0989	0.1535	0.1257	0.1078	-0.0793	0.0138

r_h: sample correlation between one-unit-apart (1.7 mile) horizontal neighbors
r_v: sample correlation between one-unit-apart vertical neighbors
r_{hv}: sample correlation between one-unit-apart neighbors
$r_{c(1)}$: sample correlation between one-unit-apart (1.7 $\sqrt{2}$ mile) corner neighbors
$r_{c(2)}$: sample correlation between one-unit-apart (1.7 $\sqrt{2}$ mile) corner neighbors
r_c: sample correlation between one-unit-apart (1.7 $\sqrt{2}$ mile) corner neighbors

Table 7.5: Model selection criteria for all 3 types of competing models for Tba:

		RMSE	AIC	BIC	AICC
1CV	Model (7.1)(a)	20.24	2838.41	2857.13	2838.61
	Model (7.1)(b)	20.12	2841.21	2867.41	2841.57
	Model (7.1)(c)	20.30	2842.48	2872.43	2842.96
	Model (7.5)(a)	28.78	NA	NA	NA
	Model (7.5)(b)	26.77	NA	NA	NA
	Model (7.5)(c)	34.76	NA	NA	NA
	Model (7.6)	21.92	2826.36	2852.54	2826.73

Model (7.1)(a): the model (7.1) with predictors β_0 (intercept), Band_4, and Ratio_2.
Model (7.1)(b): the model (7.1) with predictors β_0 (intercept), Band_4, Ratio_2, and location coordinates s_1, s_2.
Model (7.1)(c): the model (7.1) with predictors β_0 (intercept), Band_4, Ratio_2, location coordinates s_1, s_2, $s_1 \times s_2$.
Model (7.5)(a): the model (7.5), i.e. most similar neighbor model with all predictors
Model (7.5)(b): the model (7.5), i.e. most similar neighbor model with all predictors and location coordinates s_1, s_2.
Model (7.5)(c): the model (7.5), i.e. most similar neighbor model with all predictors, location coordinates s_1, s_2, $s_1 \times s_2$, Model (7.6): teh model (7.6) with predictors β_0 (intercept), Asp_{sin}, Elev, Band_4, and Cap_3.

Table 7.6: The estimates of the parameters from the linear regression model (7.1) with one predictor for Lt

	$\hat{\beta}_0$	$s.e.(\hat{\beta}_0)$	$\hat{\beta}_i$	$s.e.(\hat{\beta}_i)$	$\hat{\sigma}$	$\hat{R}^2$	p-value
Asp	385.80	30.89	-0.10	0.14	269.54	0.002	0.493
	(0.000)		(0.493)				
Asp_{sin}	369.79	15.28	35.58	21.70	268.59	0.009	0.102
	(0.000)		(0.102)				
Asp_{cos}	367.38	15.28	-1.04	21.51	269.75	0.000	0.961
	(0.000)		(0.961)				
Asp_{RC}	369.53	26.77	-4.24	43.01	269.74	0.000	0.921
	(0.000)		(0.921)				
Asp_{S}	334.22	25.31	71.08	43.41	268.59	0.009	0.103
	(0.000)		(0.103)				
Elev	367.50	33.63	-0.00	0.11	269.75	0.000	0.996
	(0.000)		(0.996)				
Slope	335.81	35.53	1.95	1.98	269.33	0.003	0.326
	(0.000)		(0.326)				
Band_1	357.06	219.64	0.16	3.47	269.75	0.000	0.963
	(0.105)		(0.963)				
Band_2	355.18	110.57	0.54	4.81	269.74	0.000	0.911
	(0.001)		(0.911)				
Band_3	432.3	68.35	-3.64	3.73	269.34	0.003	0.330
	(0.000)		(0.330)				
Band_4	273.26	50.73	1.09	0.56	268.12	0.012	0.053
	(0.000)		(0.053)				
Band_5	379.56	44.44	-0.27	0.93	269.71	0.000	0.770
	(0.000)		(0.770)				
Band_7	384.05	34.03	-1.39	2.54	269.62	0.001	0.584
	(0.000)		(0.584)				
Cap_1	304.99	66.34	0.58	0.60	269.34	0.003	0.335
	(0.000)		(0.335)				
Cap_2	313.33	29.3	1.74	0.81	267.75	0.015	0.032
	(0.000)		(0.032)				
Cap_3	314.37	25.7	4.50	1.77	266.97	0.021	0.011
	(0.000)		(0.011)				
Ratio_1	184.37	60.78	37.63	12.11	265.64	0.03	0.002
	(0.003)		(0.002)				
Ratio_2	509.53	56.59	-267.74	102.70	266.84	0.021	0.010
	(0.000)		(0.010)				
Ratio_3	281.48	73.99	21.61	18.22	269.14	0.005	0.236
	(0.000)		(0.236)				

Note: the p-values in parenthesis are for testing the hypothesis $H_0 : \beta_i = 0$ and the p-values in the last column are for testing the acceptance of the model.

Table 7.7: The model (7.1) with all predictor variables for Lt

Residual Standard Error = 265.58, Multiple R-Square = 0.0713
N = 312, F-statistic = 1.6297 on 14 and 297 df, p-value = 0.074

	coef	std.err	t.stat	p.value
β_0	-453.5849	422.2453	-1.0742	0.2836
Asp	0.1077	0.2288	0.4707	0.6382
Asp_{sin}	43.6416	35.6857	1.2229	0.2223
Asp_{cos}	-6.1609	21.9483	-0.2807	0.7791
Elev	-0.0313	0.1182	-0.2649	0.7913
Slope	1.8589	2.0639	0.9007	0.3685
Band_1	6.8879	7.6380	0.9018	0.3679
Band_2	7.1425	14.3434	0.4980	0.6189
Band_3	-7.7924	14.2148	-0.5482	0.5840
Band_4	0.1518	3.5032	0.0433	0.9655
Band_5	-8.2105	4.6803	-1.7543	0.0804
Band_7	13.3605	8.3827	1.5938	0.1120
Ratio_1	64.2024	47.1565	1.3615	0.1744
Ratio_2	104.1342	330.0039	0.3156	0.7526
Ratio_3	38.8062	27.9878	1.3865	0.1666

The t statistic is for testing if the coefficient is zero and the two sided p-value is for the t statistic.

Table 7.8: The model (7.1) with the best predictor variables and location coordinates (s_1, s_2) for Lt

	Model (7.1)(a)			
	coef	std.err	t.stat	p.value
β_0	184.3730	60.7825	3.0333	0.0026
Ratio_1	37.6271	12.1100	3.1071	0.0021
	Model (7.1)(b)			
	coef	std.err	t.stat	p.value
β_0	277.3792	71.8849	3.8587	0.0001
s_1	-14.1999	4.6069	-3.0823	0.0022
s_2	1.3869	0.8716	1.5913	0.1126
Ratio_1	41.6222	12.0854	3.4440	0.0007
	Model (7.1)(c)			
	coef	std.err	t.stat	p.value
β_0	174.0339	100.4176	1.7331	0.0841
s_1	-3.5631	8.5691	-0.4158	0.6778
s_2	5.3021	2.8001	1.8935	0.0592
$s_1 s_2$	-0.3208	0.2181	-1.4710	0.1423
Ratio_1	38.0591	12.3034	3.0934	0.0022

The t statistic is for testing if the coefficient is zero and the two sided p-value is for the t statistic.

Table 7.9: Model selection criteria for all 3 types of competing models for Lt:

		RMSE	AIC	BIC	AICC
1CV	Model (7.1)(a)	234.78	4371.90	4383.12	4371.97
	Model (7.1)(b)	233.59	4366.33	4385.04	4366.52
	Model (7.1)(c)	235.24	4366.12	4388.58	4366.39
	Model (7.5)(a)	319.49	NA	NA	NA
	Model (7.5)(b)	311.05	NA	NA	NA
	Model (7.5)(c)	302.84	NA	NA	NA
	Model (7.6)	240.19	4361.89	4371.21	4361.97

Model (7.1)(a): the model (7.1) with predictor β_0 (intercept) Ratio_1.
Model (7.1)(b): the model (7.1) with predictors β_0 (intercept), Ratio_1, and location coordinates s_1, s_2.
Model (7.1)(c): the model (7.1) with predictors β_0 (intercept), Ratio_1, location coordinates s_1, s_2, $s_1 \times s_2$.
Model (7.5)(a): the model (7.5), i.e. most similar neighbor model with all predictors
Model (7.5)(b): the model (7.5), i.e. most similar neighbor model with all predictors and location coordinates s_1, s_2.
Model (7.5)(c): the model (7.5), i.e. most similar neighbor model with all predictors, location coordinates s_1, s_2, $s_1 \times s_2$.

Table 7.10: The estimates of the parameters from the linear regression model (7.1) with one predictor for M

	$\hat{\beta}_0$	$s.e.(\hat{\beta}_0)$	$\hat{\beta}_i$	$s.e.(\hat{\beta}_i)$	$\hat{\sigma}$	$\hat{R}^2$	p-value
Asp	68.45	7.74	-0.03	0.04	67.56	0.002	0.410
	(0.000)		(0.410)				
Asp_{sin}	63.48	3.83	8.63	5.44	67.36	0.008	0.114
	(0.000)		(0.114)				
Asp_{cos}	62.98	3.83	-4.48	5.39	67.56	0.002	0.406
	(0.000)		(0.406)				
Asp_{RC}	67.74	6.70	-9.48	10.77	67.55	0.002	0.379
	(0.000)		(0.379)				
Asp_{S}	54.98	6.35	16.98	10.89	67.37	0.008	0.120
	(0.000)		(0.120)				
Elev	39.70	8.30	0.09	0.03	66.58	0.031	0.002
	(0.000)		(0.002)				
Slope	63.89	8.92	-0.06	0.50	67.63	0.000	0.901
	(0.000)		(0.901)				
Band_1	155.57	54.81	-1.47	0.87	67.32	0.009	0.091
	(0.005)		(0.091)				
Band_2	111.63	27.58	-2.14	1.20	67.29	0.010	0.075
	(0.000)		(0.075)				
Band_3	96.93	17.05	-1.91	0.93	67.18	0.013	0.041
	(0.000)		(0.041)				
Band_4	90.20	12.69	-0.32	0.14	67.08	0.016	0.025
	(0.000)		(0.025)				
Band_5	90.23	11.02	-0.61	0.23	66.88	0.022	0.009
	(0.000)		(0.009)				
Band_7	85.89	8.41	-1.92	0.63	66.63	0.029	0.002
	(0.000)		(0.002)				
Cap_1	105.65	16.47	-0.40	0.15	66.87	0.022	0.008
	(0.000)		(0.008)				
Cap_2	75.13	7.36	-0.39	0.20	67.22	0.012	0.053
	(0.000)		(0.053)				
Cap_3	54.69	6.49	0.70	0.45	67.37	0.008	0.119
	(0.000)		(0.119)				
Ratio_1	90.41	15.39	-5.66	3.07	67.26	0.011	0.066
	(0.000)		(0.066)				
Ratio_2	77.65	14.32	-27.80	25.98	67.51	0.004	0.285
	(0.000)		(0.285)				
Ratio_3	30.40	18.50	8.18	4.56	67.28	0.010	0.074
	(0.101)		(0.074)				

Note: the p-values in parenthesis are for testing the hypothesis $H_0 : \beta_i = 0$ and the p-values in the last column are for testing the acceptance of the model.

Table 7.11: The model (7.1) with all predictor variables for M

Residual Standard Error = 65.0684, Multiple R-Square = 0.1132
N = 312, F-statistic = 2.7075 on 14 and 297 df, p-value = 9e-04

	coef	std.err	t.stat	p.value
β_0	-31.7014	103.4530	-0.3064	0.7595
Asp	0.0228	0.0561	0.4070	0.6843
Asp_{sin}	19.0961	8.7432	2.1841	0.0297
Asp_{cos}	-6.3016	5.3775	-1.1718	0.2422
Elev	0.0924	0.0290	3.1911	0.0016
Slope	-0.5938	0.5057	-1.1742	0.2413
Band_1	2.3705	1.8714	1.2667	0.2062
Band_2	4.7956	3.5142	1.3646	0.1734
Band_3	-7.4251	3.4827	-2.1320	0.0338
Band_4	0.6408	0.8583	0.7465	0.4559
Band_5	-0.3160	1.1467	-0.2755	0.7831
Band_7	-1.1296	2.0538	-0.5500	0.5827
Ratio_1	-22.8735	11.5537	-1.9798	0.0487
Ratio_2	15.8793	80.8532	0.1964	0.8444
Ratio_3	6.3660	6.8572	0.9284	0.3540

The t statistic is for testing if the coefficient is zero and the two sided p-value is for the t statistic.

Table 7.12: The model (7.1) with the best predictor variables and location coordinates (s_1, s_2) for M

	Model (7.1)(a)			
	coef	std.err	t.stat	p.value
β_0	100.3716	18.1237	5.5382	0.0000
Asp_{sin}	15.9520	5.5519	2.8733	0.0043
Elev	0.0791	0.0270	2.9271	0.0037
Band_7	-1.8865	0.6421	-2.9378	0.0036
Ratio_1	-7.2562	3.0840	-2.3529	0.0193
	Model (7.1)(b)			
	coef	std.err	t.stat	p.value
β_0	118.0681	19.9142	5.9288	0.0000
s_1	-3.1036	1.2624	-2.4584	0.0145
s_2	0.2436	0.2162	1.1266	0.2608
Asp_{sin}	16.3367	5.5189	2.9601	0.0033
Elev	0.1076	0.0307	3.5078	0.0005
Band_7	-2.1074	0.6445	-3.2700	0.0012
Ratio_1	-6.4505	3.0863	-2.0900	0.0374
	Model (7.1)(c)			
	coef	std.err	t.stat	p.value
β_0	128.6107	27.1364	4.7394	0.0000
s_1	-4.1365	2.2023	-1.8782	0.0613
s_2	-0.1342	0.6943	-0.1932	0.8469
$s_1 s_2$	0.0311	0.0543	0.5726	0.5673
Asp_{sin}	16.2637	5.5264	2.9429	0.0035
Elev	0.1067	0.0308	3.4690	0.0006
Band_7	-2.1552	0.6505	-3.3129	0.0010
Ratio_1	-6.0559	3.1657	-1.9130	0.0567

The t statistic is for testing if the coefficient is zero and the two sided p-value is for the t statistic.

Table 7.13: Model selection criteria for all 3 types of competing models for M:

		RMSE	AIC	BIC	AICC
1CV	Model (7.1)(a)	57.00	3496.16	3518.62	3496.43
	Model (7.1)(b)	56.55	3493.97	3523.92	3494.45
	Model (7.1)(c)	56.34	3495.64	3529.33	3496.24
	Model (7.5)(a)	78.09	NA	NA	NA
	Model (7.5)(b)	63.70	NA	NA	NA
	Model (7.5)(c)	72.48	NA	NA	NA
	Model (7.6)	62.19	3486.23	3498.56	3486.43

Model (7.1)(a): the model (7.1) with predictors β_0 (intercept), Asp_{sin}, Elev, Band_7, and Ratio_1.
Model (7.1)(b): the model (7.1) with predictors β_0 (intercept), Asp_{sin}, Elev, Band_7, Ratio_1, and location coordinates s_1, s_2.
Model (7.1)(c): the model (7.1) with predictors β_0 (intercept), Asp_{sin}, Elev, Band_7, Ratio_1, location coordinates s_1, s_2, $s_1 \times s_2$.
Model (7.5)(a): the model (7.5), i.e. most similar neighbor model with all predictors
Model (7.5)(b): the model (7.5), i.e. most similar neighbor model with all predictors and location coordinates s_1, s_2.
Model (7.5)(c): the model (7.5), i.e. most similar neighbor model with all predictors, location coordinates s_1, s_2, $s_1 \times s_2$.

8. PREDICTIONS FOR NON-SAMPLED LOCATIONS INCORPORATING AUXILIARY INFORMATION

To make predictions is one of our objectives for this study. In Chapter 4 we only discussed the predictions (or estimations) for primary sample units (1-ha plots) on the 1.7-mile grid. This chapter investigates predictions of spatially varying response variables at non-sampled plots on the 0.85-mile grid based on the observations at sampled plots, i.e. fill-in predictions or interpolations. Intuitively, a linear combination of the observations is suggested, i.e. a linear prediction, and minimization of mean-squared error (MSE) is used leading to the classical best linear prediction (BLP).

In this chapter, we will use the models described in Chapter 7, i.e. models 7.1, 7.5, and 7.6, to make predictions for 6 selected non-sampled plot locations.

8.1 Predictions and Prediction Intervals Using General Linear Model (GLM)

Definition: (see Graybill 1984, Definition 6.1.1)

General Linear Model. Let $\mathbf{Y}$ be an $N \times 1$ observable vector of random variables; let $\mathbf{X}_*$ be an $N \times k$ matrix ($N > k$) of known fixed numbers; let β be a $k \times 1$ vector of unknown parameters; let ε be an $N \times 1$ unobservable vector of random variables, where $\mathcal{E}[\varepsilon] = \mathbf{0}$ and $\mathrm{Cov}[\varepsilon] = \Sigma$; let these quantities be related by

$$\mathbf{Y} = \mathbf{X}_*\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\mathbf{X}_* = (\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(k)})'$. These specifications define a general linear model. This model is referred to as a general linear sample model.

Theorem: (see Graybill 1984, Theorem 8.2.1)

Consider the general linear model given in the above Definition for the case: ε is distributed normally with mean zero and covariance matrix $\sigma^2\mathbf{I}$, where $\sigma^2 > 0$ is unknown, that is, ε is distributed $N(\mathbf{0}, \sigma^2\mathbf{I})$. Let $\hat{\beta}$ and $\hat{\sigma}^2$ be the UMVU estimators of β and σ^2, respectively,. Let $\mathbf{X}_0$ be a vector in domain D of $\mu(\mathbf{X}_0)$. A $1-\alpha$ prediction interval for $\overline{Y}_0$, the mean of l future observations selected at random from the normal distribution with mean $\mu(\mathbf{X}_0) = \beta'\mathbf{X}_0$ and variance σ^2 is given by

$$\hat{\beta}'\mathbf{X}_0 - h_{\alpha/2}\hat{\sigma} \leq \overline{Y}_0 \leq \hat{\beta}'\mathbf{X}_0 + h_{\alpha/2}\hat{\sigma} \tag{8.1}$$

where

$$h_{\alpha/2} = t_{\alpha/2,N-k-1}\left(\frac{1}{l} + \mathbf{X}_0'S_*^{-1}\mathbf{X}_0\right)^{1/2}, \tag{8.2}$$

and $S_* = \mathbf{X}_*'\mathbf{X}_*$. When $l = 1$, the prediction interval of Eqn. 8.1 yields the prediction interval for an individual future observation.

The general linear model we consider in this thesis is $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ (see 7.1) where $\mathbf{X} = (\mathbf{1}, \mathbf{X}_*)$, and so a $1-\alpha$ prediction interval for $\hat{Y}_{s_0,glm}$ at non-sampled plot location $\boldsymbol{s}_0$, is given by

$$\hat{\beta}'\mathbf{X}_0 - h_{\alpha/2}\hat{\sigma} \leq \hat{Y}_{s_0,glm} \leq \hat{\beta}'\mathbf{X}_0 + h_{\alpha/2}\hat{\sigma} \tag{8.3}$$

where the subscript glm in $\hat{Y}_{s_0,glm}$ indicates the prediction using the general linear model.

$$h_{\alpha/2} = t_{\alpha/2,N-k-1}\left(1 + \mathbf{X}_0'S^{-1}\mathbf{X}_0\right)^{1/2}, \tag{8.4}$$

where $\mathbf{X}_0$ is a vector of covariate observations at location $\boldsymbol{s}_0$, and $S = \mathbf{X}'\mathbf{X}$.

For the simple linear model, $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, $i = 1, \ldots, N$,

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix}$$

$$
\begin{aligned}
S &= \mathbf{X}'\mathbf{X} \\
&= \begin{bmatrix} N & \sum_{i=1}^{N} x_i \\ \sum_{i=1}^{N} x_i & \sum_{i=1}^{N} x_i^2 \end{bmatrix}
\end{aligned}
$$

Thus

$$
\begin{aligned}
& \mathbf{x}_0' S^{-1} \mathbf{x}_0 \\
= & \begin{bmatrix} 1 & x_0 \end{bmatrix} \left(\frac{1}{N\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2} \right) \begin{bmatrix} \sum_{i=1}^{N} x_i^2 & -\sum_{i=1}^{N} x_i \\ -\sum_{i=1}^{N} x_i & N \end{bmatrix} \begin{bmatrix} 1 \\ x_0 \end{bmatrix} \\
= & \frac{1}{N} + \frac{(\overline{x} - x_0)^2}{\sum (x_i - \overline{x})^2}
\end{aligned}
$$

And, hence a $1-\alpha$ prediction interval for $\hat{Y}_{s_0,\,slm}$ at non-sampled plot location $\boldsymbol{s_0}$, is given by

$$
\hat{\beta}_0 + \hat{\beta}_1 x_0 - t_{\alpha/2,N-2}\,\hat{\sigma} A_1 \leq \hat{Y}_{s_0,\,slm} \leq \hat{\beta}_0 + \hat{\beta}_1 x_0 + t_{\alpha/2,N-2}\,\hat{\sigma} A_1 \tag{8.5}
$$

where the subscript slm in $\hat{Y}_{s_0,\,slm}$ indicates the prediction using the simple linear model, and A_1^2 is defined as:

$$
A_1^2 = 1 + \frac{1}{N} + \frac{(\overline{x} - x_0)^2}{\sum (x_i - \overline{x})^2}.
$$

For a general linear model, the standard deviation of the prediction error for the prediction $\hat{Y}_{s_0,\,glm}$ of Y_{s_0} given a specific value of $\mathbf{X}_0$ at non-sampled plot location $\boldsymbol{s_0}$ is

$$
\hat{\sigma}_{Y,\,p,\,glm} = \hat{\sigma}\left(1 + \mathbf{X}_0' S^{-1} \mathbf{X}_0\right)^{1/2} \tag{8.6}
$$

where $\hat{\sigma}^2$ is the estimated standard deviation of the random error ε_i. $\hat{\sigma}_{Y,\,p,\,glm}$ is referred to as the standard error of prediction. The simple linear model is a special case of a general linear regression model and its standard error of prediction is

$$
\hat{\sigma}_{Y,\,p,\,slm} = \hat{\sigma}\sqrt{1 + \frac{1}{N} + \frac{(\overline{x} - x_0)^2}{\sum (x_i - \overline{x})^2}} \tag{8.7}
$$

8.2 Predictions and Prediction Intervals Using the MSN Procedure

The predicted value by the MSN procedure is $\hat{y}_i = y_{j'}$ when D_{ij} is minimized at $j = j'$. The similarity is defined by the similarity function

$$(See 7.5) \qquad D_{ij} = (\mathbf{X}_i - \mathbf{X}_j)G\Lambda^2 G'(\mathbf{X}_i - \mathbf{X}_j)', \ i, j = 1, \ldots, N,$$

where G are canonical coefficients of the $\mathbf{X}$'s, Λ is a vector of the correlations between the $\mathbf{X}$'s and $\mathbf{Y}$'s from sampled plots; $\mathbf{Y}$ might be forest variables, such as mortality (M), mortality of seedlings (M_s), total numbers of live trees (Lt), Lt of seedlings (Lt_s), total basal area (Tba), Tba of seedlings (Lt_s), and $\mathbf{X}_i = (X_i^{(1)}, \ldots, X_i^{(k)})$, i.e. Aspect, Elevation, Slope, TM bands 1-5 and 7 at plot i.

Then a data based predictor from the MSN procedure is denoted as $\hat{Y}_{s_0,\ msn}$ at location s_0. The prediction error of $\hat{Y}_{s_0,\ msn}$ is hard to derive because there is no explicit closed form, so a simulation is conducted to obtain the prediction error.

8.3 Predictions and Prediction Intervals Using Spatial Regression Model (SRM)

The spatial regression model we considered in Chapter 7 is also referred to as "Phase II" model, given by

$$(See 7.6) \qquad \begin{aligned} \mathbf{Y} &= \boldsymbol{\mu} + B\boldsymbol{\varepsilon} \\ &\equiv \mathbf{X}\boldsymbol{\beta} + B\boldsymbol{\varepsilon} \end{aligned}$$

where $\mathbf{Y} = (Y_1, \ldots, Y_N)'$, $\mathbf{X} = (\mathbf{1}, \mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(k)})'$, $\boldsymbol{\Gamma} = \mathbf{I} + \rho_1 \mathbf{H} + \rho_2 \mathbf{V} + \rho_3 \mathbf{C}^{(1)} + \rho_4 \mathbf{C}^{(2)} \equiv B'B$, $\boldsymbol{\varepsilon} \sim \text{MVN}(\mathbf{0}, \sigma^2 \mathbf{I})$. Since this $\boldsymbol{\Gamma}$ does not allow predictions for non-sampled plots on the 0.85 mile grid, we consider an alternative approach and assume that the correlation function is distance based as we studied in Chapter 5. The "Phase II" model, with a distance-related autocorrelation function $\rho(d)$,

can be rewritten as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\delta} \tag{8.8}$$

where $\boldsymbol{\delta}(.)$ is a zero-mean finite-variance error process that is spatially correlated and $\text{Var}[\boldsymbol{\delta}] = \sigma^2\Gamma \equiv \Sigma$ where $\Gamma = (\rho_{ij})_{i,j=1}^{N}$.

Then for known $\boldsymbol{\beta}$ and Γ, an optimal predictor is the simple kriging predictor (Cressie, 1993)

$$\begin{aligned} Y_{s_0,\,srm} &= X_0\boldsymbol{\beta} + \boldsymbol{c}'\Sigma^{-1}(Y - X\boldsymbol{\beta}) && (8.9) \\ \sigma^2_{Y,\,p,\,srm} &= \sigma^2 - \boldsymbol{c}'\Sigma^{-1}\boldsymbol{c} + (X_0 - X'\Sigma^{-1}\boldsymbol{c})'(X'\Sigma^{-1}X)^{-1} \\ &\quad (X_0 - X'\Sigma^{-1}\boldsymbol{c}) && (8.10) \end{aligned}$$

where the subscript srm in $\hat{Y}_{s_0,\,srm}$ indicates the prediction using a spatial regression model, and $\boldsymbol{c} \equiv \sigma^2(\rho_{01}, \ldots, \rho_{0N})'$.

Then a data based predictor is given by (Cressie, 1993)

$$\widehat{Y}_{s_0,\,srm} = X_0\widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{c}}'\widehat{\Sigma}^{-1}(Y - X\widehat{\boldsymbol{\beta}}) \tag{8.11}$$

where the generalized least squares estimator of $\boldsymbol{\beta}$ is

$$\widehat{\boldsymbol{\beta}} = (X'\widehat{\Sigma}^{-1}X)^{-1}X'\widehat{\Sigma}^{-1}Y \tag{8.12}$$

and the estimated mean-squared prediction error is

$$\begin{aligned} \widehat{\sigma}^2_{Y,\,p,\,srm} &= \widehat{\sigma}^2 - \widehat{\boldsymbol{c}}'\widehat{\Sigma}^{-1}\widehat{\boldsymbol{c}} + (X_0 - X'\widehat{\Sigma}^{-1}\widehat{\boldsymbol{c}})'(X'\widehat{\Sigma}^{-1}X)^{-1} \\ &\quad (X_0 - X'\widehat{\Sigma}^{-1}\widehat{\boldsymbol{c}}) && (8.13) \end{aligned}$$

Thus a $1 - \alpha$ prediction interval for $\hat{Y}_{s_0,\,srm}$ at non-sampled plot location $\boldsymbol{s_0}$, is given by

$$\left(\widehat{Y}_{s_0,\,srm} - z_{1-\alpha/2}\widehat{\sigma}_{Y,\,p,\,srm},\ \widehat{Y}_{s_0,\,srm} + z_{1-\alpha/2}\widehat{\sigma}_{Y,\,p,\,srm}\right) \tag{8.14}$$

under the assumption that $Y(\cdot)$ is Gaussian.

To estimate the covariance matrix Σ for $\boldsymbol{\delta}$, we use the least squares method to get the estimated $\widehat{\boldsymbol{\beta}}_{LS}$ of $\boldsymbol{\beta}$, and then calculate $\boldsymbol{\delta}$= $\mathbf{Y}$ - $\mathbf{X}\widehat{\boldsymbol{\beta}}_{LS}$. Then we fit the autocorrelation function to the sample correlation among $\boldsymbol{\delta}$. The autocorrelation functions we consider are listed as follows:

(1) $\rho_{ij}(\boldsymbol{\delta}) = exp(-\theta d_{ij}^{k})$, $\theta, k > 0$,

(2) $\rho_{ij}(\boldsymbol{\delta}) = (1 + \theta d_{ij}^{2})^{(-k)}$, $\theta, k > 0$.

We use least squares and weighted least squares methods in finding the parameters θ and k that minimize

$$\sum_{"d"} N(d) \left\{\rho^{\#}(d) - \rho(d)\right\}^{2} \tag{8.15}$$

for lag distance d and $N(d)$ is the number of pairs of observations with lag distance d. The estimates $\hat{\theta}$ and $\hat{k}$ can be obtained from an iterative, nonlinear estimation routine such as the Gauss-Newton algorithm in the Splus nlminb function. The autocorrelation model fitting for the residuals $\boldsymbol{\delta}$ of Tba on 1-ha plots are portrayed in Figure 8.1. From Figure 8.1, we note that the correlation function (2) doesn't fit well with either weighted or unweighted least squares methods and the fitting curves are overlapped, hence in further considerations, we only use the autocorrelation function (1), $\rho_{ij}(\boldsymbol{\delta}) = exp(-\theta d_{ij}^{k})$, $\theta, k > 0$ for Tba fitting. Figure 8.6 shows autocorrelation model fitting for $\boldsymbol{\delta}$ of Lt. The autocorrelation function (1), $\rho_{ij}(\boldsymbol{\delta}) = exp(-\theta d_{ij}^{k})$, $\theta, k > 0$ fits much better, so this function is used for further discussion for Lt. From Figure 8.9, we note that the correlation function (2) doesn't fit well to $\boldsymbol{\delta}$ of M with either weighted or unweighted least squares methods, so here too, we only use the autocorrelation function (1), $\rho_{ij}(\boldsymbol{\delta}) = exp(-\theta d_{ij}^{k})$, $\theta, k > 0$.

Finally, we plug $\widehat{\Sigma}^{-1}$ into Eqn. (8.12) to get the generalized least squares estimate of $\boldsymbol{\beta}$.

8.4 Bootstrap Methods and Simulations

8.4.1 Bootstrap methods

In this section, we use two different ways of bootstrapping the models described earlier: a simple bootstrap (i.e. bootstrapping pairs (y_i, X_i)) and a semiparametric bootstrap method (i.e. bootstrapping residuals). The bootstrapping pairs method is less sensitive to assumptions than bootstrapping residuals. Which bootstrap method is better? It depends on how well we trust the model.

In bootstrapping residuals, we first select a random sample of bootstrap error terms $\boldsymbol{\varepsilon}^*=(\varepsilon_i^*, \ldots, \varepsilon_N^*)$ from $\{\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_N\}$, and then the bootstrap responses y_i^* are generated according to a specific model. For example, for a simple linear regression model,

$$y_i^* = \boldsymbol{x}_i\hat{\boldsymbol{\beta}} + \varepsilon_i^*, \; i = 1, 2, \ldots, N. \tag{8.16}$$

The predicted value by the MSN procedure is $\hat{y}_i = y_{j'}$ when D_{ij} is minimized at $j = j'$, and so the residual $\hat{\varepsilon}_i$ is obtained by $y - \hat{y}_i$. The MSN procedure focuses on the similarity function to predict via the value at plot j' with the nearest D_{ij} distance to plot i without specifying a certain model. Although the residuals can be obtained in the MSN inference, we still have trouble in back-transforming the residuals to bootstrapped response Y_i^* without the model equation. And hence it is not possible to use the bootstrapped residuals for the MSN procedure. Also, we don't have an explicit form for the prediction error of prediction from the MSN procedure. We will conduct a simulation to investigate the prediction error.

A semiparametric bootstrap method does consider spatial dependence among the data. Suppose that the spatial data follows a spatial regression model, i.e. "Phase II" model

$$(See 7.6) \qquad \mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + B\boldsymbol{\varepsilon}$$

where $\mathbf{Y} = (Y_1, \ldots, Y_N)'$, $\mathbf{X} = (\mathbf{1}, \mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(k)})'$, $\Gamma = (\rho_{ij})_{i,j=1}^{N} \equiv B'B$, $\boldsymbol{\varepsilon} \sim$ MVN($\mathbf{0}$, $\sigma^2\,\mathbf{I}$).

For each bootstrap sample ε^*, transformation back to the y_i is calculated by

$$\mathbf{Y}^* = \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{B}\hat{\boldsymbol{\varepsilon}}.$$

8.4.2 Simulations

Since we don't have an explicit form of prediction errors of the predictions from the MSN procedure, a simulation study is conducted to examine the range of mean squared prediction error of individual prediction from the MSN procedure. The reduced linear model for Tba we obtained in Chapter 7 is $Y_i = 86.31 + 11.05\,\mathrm{Asp}_S - 0.35\,\mathrm{Band}_4 + \varepsilon_i$, where $\boldsymbol{\varepsilon}$=$(\varepsilon_1, \ldots, \varepsilon_{N+6})$ is MVN($\mathbf{0}$, $(22.68)^2\,\mathbf{I}$). The simulation procedure is

(1) Generate $\boldsymbol{\varepsilon}^*$, a realization of size $(N+6)$ from MNV($\mathbf{0}$, $(22.68)^2\,\mathbf{I}$).

(2) Since the covariates are known numbers and we have the estimates, $\hat{\beta}_0 = 85.71$, $\hat{\beta}_1 = 11.01$, $\hat{\beta}_2 = -0.34$, and $\hat{\beta}_3$, the generated realization from model 7.1 is

$$Y_i^* = 85.71 + 11.01\,\mathrm{Asp}_S - 0.34\,\mathrm{Band}_4 + \varepsilon_i^*,\ i = 1, \ldots, N+6$$

(3) Repeat steps 1 and 2 until $b = 1{,}000$ realizations of $\mathbf{Y}^*$ are obtained.

Put the (N+1)-th, ... , (N+6)-th aside, use the prediction rules of the general linear model and the MSN procedure based on the generated values at plot locations $s_1, \ldots, s_N$, where $N = 312$. For each realization, we have a pair of simulated and predicted values, say $(y_i^*, \hat{y}_{i,glm})$, $i = N+1, \ldots, N+6$, i.e. for non-sampled plots 1-6. Figure 8.4 presents the histograms of $(y_i^*, \hat{y}_{i,glm})$. We can see they are very alike.

To compare with the results from the "Phase II" model, i.e. model 7.6, we conduct a simulation by the following procedure:

(1) Generate $\boldsymbol{\varepsilon}^*$, a realization of size $(N+6)$ from MNV$(\mathbf{0}, \mathbf{I})$.

(2) Calculate the values of $\boldsymbol{\delta}^*$ by

$$\boldsymbol{\delta}^* = \hat{\sigma}\hat{B}\boldsymbol{\varepsilon}^*$$

where $\hat{\Gamma} = \hat{B}'\hat{B} = \left(exp(-\hat{\theta}d_{ij}^{\hat{k}})\right)_{i,j=1}^{N+6}$ with $\hat{\theta} = 0.0016$, $\hat{k} = 0.87$ estimated from Tba data.

(3) Since the covariates are known numbers and we have the estimates, $\hat{\beta}_0 = 85.71$, $\hat{\beta}_1 = 11.01$, $\hat{\beta}_2 = -0.34$, and $\hat{\beta}_3$, the realization from model 7.6 is

$$Y_i^* = 85.71 + 11.01\,\mathrm{Asp}_S - 0.34\,\mathrm{Band}_4 + \delta^*,\ i = 1, \ldots, N+6$$

(4) Repeat steps 1 and 3 until $b = 1{,}000$ realizations of $\mathbf{Y}^*$ are obtained.

Similarly, simulations for Lt and M are conducted.

8.5 Results

Table 8.1 summarizes the Tba bootstrapped predictions using $\hat{Y}_{s_0,glm}$ and $\hat{Y}_{s_0,srm}$ at the selected 6 non-sampled plots. The plug-in estimates are obtained by substituting the N observations into the equations of $\hat{Y}_{s_0,glm}$ and $\hat{Y}_{s_0,srm}$; then $\overline{\hat{Y}}_{s_0,glm}$ and $\overline{\hat{Y}}_{s_0,srm}$ are the averages of 1,000 bootstrapped $\hat{Y}_{s_0,glm}$ and $\hat{Y}_{s_0,srm}$, respectively. From Table 8.1, we note than the plug-in estimates $\hat{\sigma}_{Y,p,srm}$ are smaller than $\hat{\sigma}_{Y,p,glm}$, but the plug-in estimates $\hat{Y}_{s_0,glm}$ are not alway smaller than $\hat{Y}_{s_0,srm}$. The plug-in estimates, $\hat{Y}_{s_0,srm}$, for plots 1 and 4 are quite different from $\hat{Y}_{s_0,glm}$. The values from the first term in Eqn. (8.11) for plots 1 and 4 are very close to the plug-in estimates $\hat{Y}_{s_0,glm}$, but the values from the second term of $\hat{Y}_{s_0,srm}$, $\widehat{\Sigma}^{-1}(Y - X\widehat{\boldsymbol{\beta}})$, are -15.39 and -20.59, respectively. For plot 1, it is due to the large weights, 0.33, for the two neighboring plots with the negative residuals -26.07 and -26.04. For plot 4, the general mean $X_0\widehat{\boldsymbol{\beta}}$ is adjusted by the neighboring residuals -6.95 and -44.92 with the weights 0.33.

The estimates of the prediction errors for $\hat{Y}_{s_0,glm}$ and $\hat{Y}_{s_0,srm}$ of $\hat{Y}_{s_0}$ can be rewritten as:

$$\begin{aligned}
\hat{\sigma}^2_{Y,p,glm} &= \hat{\sigma}^2_{srm} + \hat{\sigma}^2_{srm}\mathbf{X}_0' S^{-1}\mathbf{X}_0 \\
&\equiv T_{11} + T_{12} \\
\widehat{\sigma}^2_{Y,p,srm} &= \widehat{\sigma}^2 - \widehat{\boldsymbol{c}}'\widehat{\Sigma}^{-1}\widehat{\boldsymbol{c}} + (X_0 - X'\widehat{\Sigma}^{-1}\widehat{\boldsymbol{c}})'(X'\widehat{\Sigma}^{-1}X)^{-1} \\
&\quad (X_0 - X'\widehat{\Sigma}^{-1}\widehat{\boldsymbol{c}}) \\
&\equiv T_{21} - T_{22} + T_{23}.
\end{aligned}$$

For each plot, the values of T_{11} are very close to those of T_{21} and the values of T_{12} are very close to those of T_{23}, so the additional term T_{22} is the major contributor to the reduced value $\widehat{\sigma}^2_{Y,p,srm}$. The values of T_{22} for the selected 6 non-sampled plot locations are about 169 which is large compared to the values of the first term, 507. Compared to the predictors in Chapter 5, the plug-in mean-squared prediction errors (m.s.p.e.) of $\hat{Y}_{s_0,srm}$ are smaller than those of the predictors in Chapter 5 (338 vs. 587).

When the simple bootstrap method is used, the estimates $\overline{\hat{\sigma}}_{Y,p,glm}$ are slightly bigger than the estimates $\overline{\widehat{\sigma}}_{Y,p,srm}$. The opposite results occur when we use the semiparametric bootstrap method. Figures 8.2 and 8.3 present the histograms of bootstrapped Tba predictions from models 7.1 and 7.6. From Figure 8.2, we note little difference in the histograms of the simple bootstrap and the semiparametric bootstrap methods. That's because the residuals are assumed independent based on the model 7.1 predictions. We can see that the histograms of the Tba bootstrapped predictions at the 6 selected non-sampled plot locations are quite different when the model with spatially correlated errors is assumed, i.e. model 7.6.

Table 8.2 shows the comparison of $\hat{Y}_{s_0,glm}$ and $\hat{Y}_{s_0,msn}$ from the generated model 7.1:

$$Y_i^* = 85.71 + 11.01\,\mathrm{Asp}_S - 0.34\,\mathrm{Band}_4 + \varepsilon_i^*,\ i = 1, \ldots, N+6$$

where $\boldsymbol{\varepsilon}=(\varepsilon_1, \ldots, \varepsilon_{N+6})$ is $\mathrm{MVN}(\mathbf{0}, (22.68)^2\,\mathbf{I})$.

From Table 8.2, we can see that $\widehat{r.m.s.e.}(\widetilde{Y}_{s_0,msn})$'s at each plot are approximately 50% larger than $\widehat{r.m.s.e.}(\widetilde{Y}_{s_0,glm})$. Table 8.3 compares the MSN predictions and spatial regression predictions, based on the generated simple regression model with realizations similar to Tba data. The estimated prediction errors for predictions $\widetilde{Y}_{s_0,srm}$ at plots 2, 3, and 5 are about 50% smaller than those for $\widetilde{Y}_{s_0,msn}$.

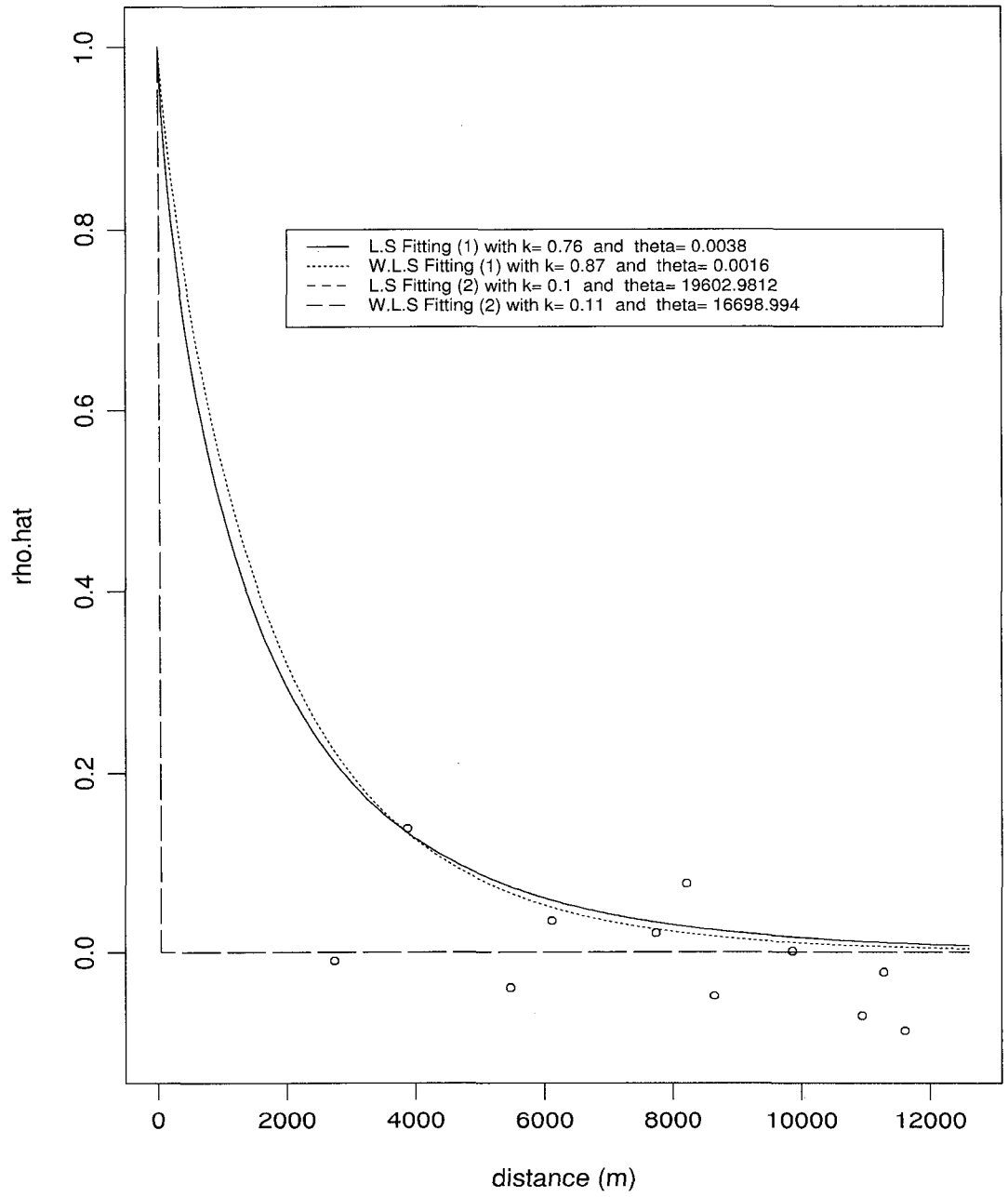
Table 8.3 shows the comparison of $\hat{Y}_{s_0,srm}$ and $\hat{Y}_{s_0,msn}$ from the generated model 7.1:

$$Y_i^* = 86.31 + 11.05\,\mathrm{Asp}_S - 0.35\,\mathrm{Band}_4 + \delta_i, i = 1, \ldots, N, N+1, \ldots, N+6$$

where $\boldsymbol{\delta}=(\delta_1, \ldots, \delta_N, \delta_{N+1}, \ldots, \delta_{N+6})$ is $\mathrm{MVN}(\mathbf{0}, (22.68)^2\,\hat{\Gamma})$, and $\hat{\Gamma} \equiv \hat{B}'\hat{B}=(exp\,(-0.0016 d_{ij}^{0.87}))_{i,j=1}^{N+6}$.

In summary, $\hat{Y}_{s_0,srm}$ is the best predictor when there exists significant spatial dependence among the data based on the smallest mean-squared prediction error (m.s.p.e). In Chapter 7, we mentioned that spatial multivariate regression models incorporating auxiliary information did not show improvement based on estimated overall RMSE. In this chapter, distance-related correlation functions and auxiliary information were included in spatial regression model. Such models did not reduce the estimated overall RMSEs for the whole area, but did reduce m.s.p.e. of the prediction errors for individual plots.

Figure 8.1: Sample correlations and the fitted exponential function for the residuals from Tba data of "Phase II" Model 7.6[1]



[1]

(1) $\rho_{ij}(\boldsymbol{\delta}) = exp(-\theta d_{ij}^{k}),\ \theta, k > 0,$

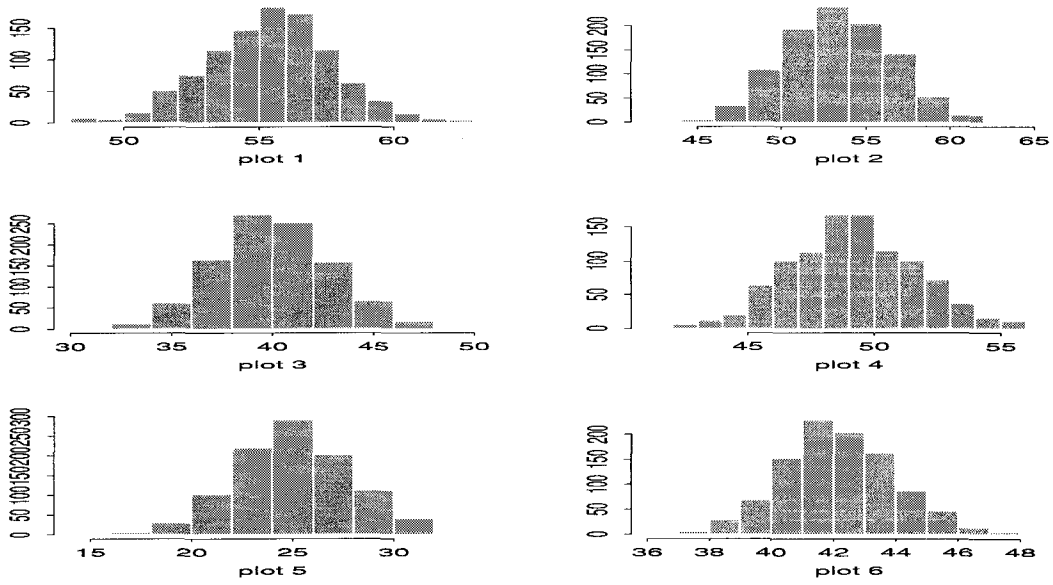
(2) $\rho_{ij}(\boldsymbol{\delta}) = (1 + \theta d_{ij}^{2})^{(-k)},\ \theta, k > 0.$

Table 8.1: The Tba predictions at the selected 6 non-sampled plot locations from 1,000 bootstrapped samples

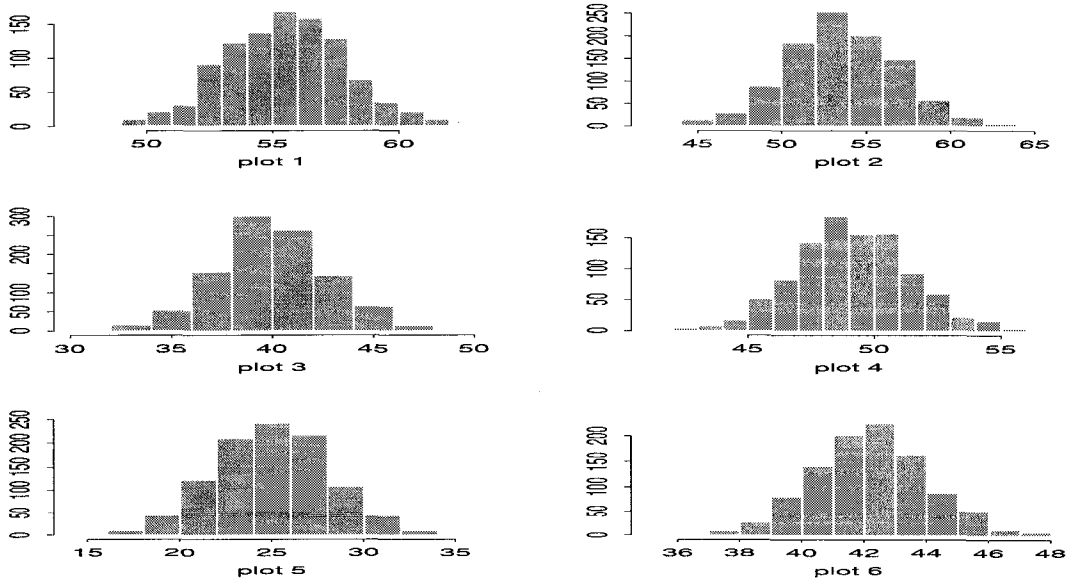
			Average of 1,000 Bootstrapped Estimates			
	Plug-in Estimates[1]		Simple Bootstrap		Semiparametric Bootstrap	
plot	$\widehat{Y}_{s_0,glm}$	$\widehat{Y}_{s_0,srm}$	$\overline{\widehat{Y}}_{s_0,glm}$	$\overline{\widehat{Y}}_{s_0,srm}$	$\overline{\widehat{Y}}_{s_0,glm}$	$\overline{\widehat{Y}}_{s_0,srm}$
1	55.41	39.35	55.38	53.65	55.46	52.68
			(2.32)	(4.22)	(2.37)	(6.03)
2	53.51	60.31	53.37	54.90	53.53	55.51
			(3.17)	(2.11)	(3.29)	(1.54)
3	40.05	36.47	40.03	38.75	39.98	37.69
			(2.80)	(1.65)	(2.70)	(0.74)
4	49.12	28.78	49.18	46.04	49.13	44.17
			(2.48)	(5.42)	(2.28)	(6.27)
5	25.16	25.60	25.08	25.59	24.97	26.08
			(2.83)	(0.61)	(3.07)	(0.64)
6	42.26	50.11	42.15	43.52	42.19	44.58
			(1.76)	(2.16)	(1.82)	(2.60)
	Plug-in Estimates[1]		Simple Bootstrap		Semiparametric Bootstrap	
plot	$\widehat{\sigma}_{Y,p,glm}$	$\widehat{\sigma}_{Y,p,srm}$	$\overline{\widehat{\sigma}}_{Y,p,glm}$	$\overline{\widehat{\sigma}}_{Y,p,srm}$	$\overline{\widehat{\sigma}}_{Y,p,glm}$	$\overline{\widehat{\sigma}}_{Y,p,srm}$
1	22.65	18.54	22.4960	21.77	22.4958	23.36
			(0.8684)	(1.50)	(0.8683)	(1.93)
2	22.75	18.64	22.6016	21.86	22.5984	23.45
			(0.8725)	(1.50)	(0.8723)	(1.91)
3	22.71	18.53	22.5576	21.81	22.5549	23.38
			(0.8700)	(1.52)	(0.8706)	(1.95)
4	22.65	18.47	22.4981	21.74	22.4967	23.33
			(0.8684)	(1.51)	(0.8684)	(1.94)
5	22.73	18.59	22.5777	21.84	22.5747	23.44
			(0.8732)	(1.51)	(0.8714)	(1.93)
6	22.61	18.44	22.4518	21.72	22.4513	23.32
			(0.8671)	(1.51)	(0.8666)	(1.96)

[1] The plug-in estimates for $\widehat{\sigma}_{Y,p,glm}$ and $\widehat{\sigma}_{Y,p,srm}$ are calculated by plugging the N observations into the Eqns. (8.6) and (8.13) and the estimated a_i based on the N observations. Also, the sample standard deviations of $\widehat{Y}_{s_0,glm}$, $\widehat{Y}_{s_0,srm}$, $\widehat{\sigma}_{Y,p,glm}$, and $\widehat{\sigma}_{Y,p,srm}$ are listed in parentheses.

Figure 8.2: 1,000 bootstrapped predictions of Tba from model 7.1 at 6 non-sample plots [1]



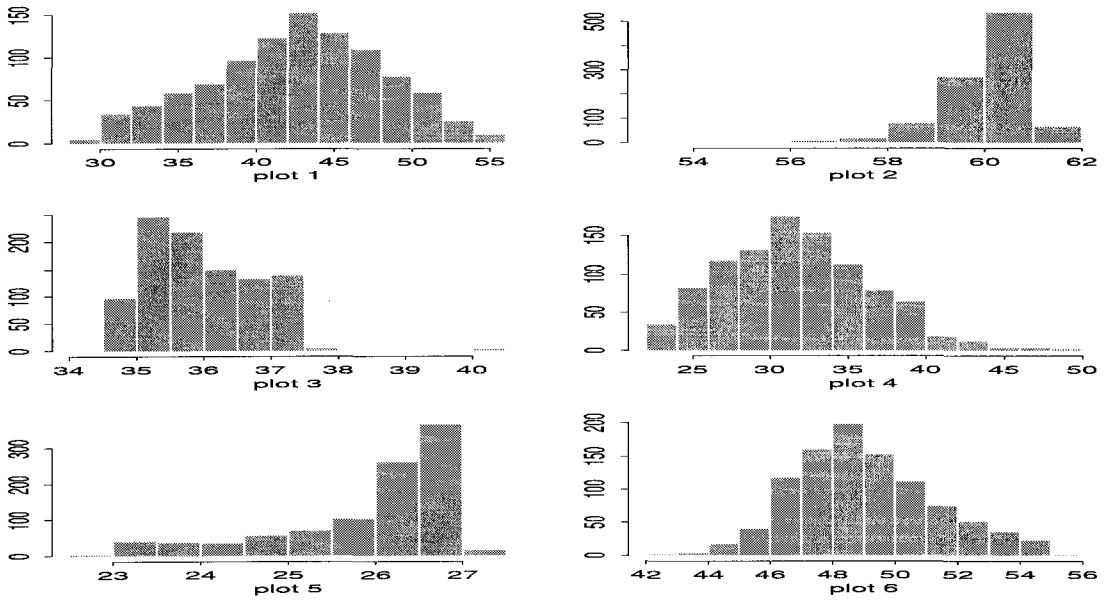
(a) using the simple bootstrap method



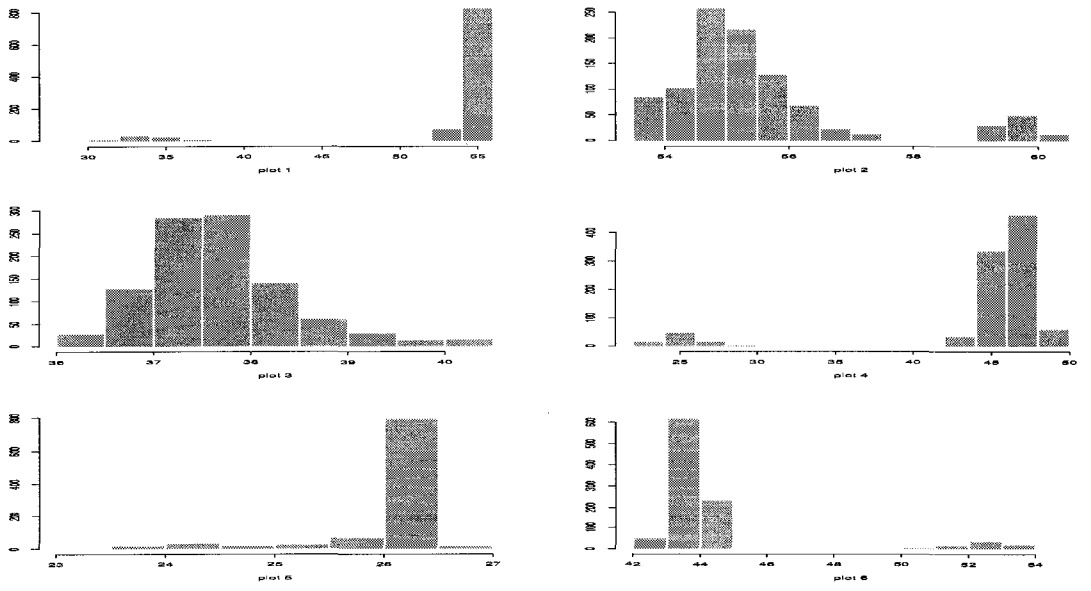
(b) using the semiparametric bootstrap method

[1] The six non-sample plot locations are depicted in Figure 5.4

Figure 8.3: 1,000 bootstrapped predictions of Tba from model 7.6 at 6 non-sample plots [1]



(a) using the simple bootstrap method



(b) using the semiparametric bootstrap method

[1] The six non-sample plot locations are depicted in Figure 5.4

Table 8.2: Comparison of predictors $\widehat{Y}_{s_0,glm}$ and $\widehat{Y}_{s_0,msn}$ for 6 non-sampled plots[1] based on 1,000 Tba realization from general linear model 7.1[2]

plot		$\widehat{Y}_{s_0,glm}$	$\widehat{Y}_{s_0,msn}$	$\hat{\sigma}_{\widehat{Y},p,glm}$	$\hat{\sigma}_{\widehat{Y},p,msn}$
1	mean	55.28	56.52	22.64	NA[3]
	s.e.	2.30	21.88	0.90	NA
2	mean	53.43	52.70	22.74	NA
	s.e.	3.20	22.43	0.90	NA
3	mean	39.94	40.61	22.69	NA
	s.e.	2.77	22.26	0.90	NA
4	mean	48.97	48.14	22.64	NA
	s.e.	2.30	22.36	0.90	NA
5	mean	25.10	28.04	22.71	NA
	s.e.	3.00	22.30	0.90	NA
6	mean	42.18	43.02	22.59	NA
	s.e.	1.82	22.19	0.90	NA

(a). The means and s.e.'s of 1,000 estimates $\widehat{Y}_{s_0,glm}$, $\widehat{Y}_{s_0,msn}$, and $\hat{\sigma}_{\widehat{Y},p,glm}$ in Eqn. (8.6)

plot	$\widehat{r.m.s.e.}(\widehat{Y}_{s_0,glm})$	$\widehat{r.m.s.e.}(\widehat{Y}_{s_0,msn})$
1	22.81	31.55
2	22.48	32.29
3	22.38	31.23
4	23.05	30.84
5	22.40	31.30
6	23.84	32.79

(b). $\widehat{r.m.s.e.}(\hat{Y}_p^i) = \sqrt{\frac{1}{1,000}\sum_{b=1}^{1,000}(\hat{Y}_{p,N+i}^b - Y_{g,N+i}^b)^2}$,

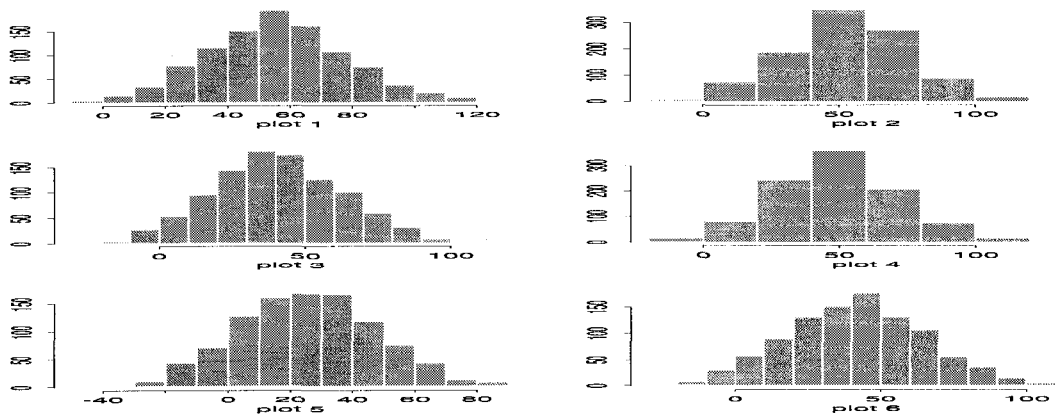
where $\hat{Y}_{p,N+i}^b$, $i = 1,\ldots,6$ denote the prediction at non-sampled plot i using the MSN predictions and the GLM predictions based on the generated values at sampled plots, $1,\ldots,N = 312$ from the b^{th} generated realization of size $(N+6)$, and $Y_{g,N+i}^b$ denotes the b^{th} generated values at non-sampled plot i.

[1] Six non-sample plot locations are depicted in Figure 5.4.

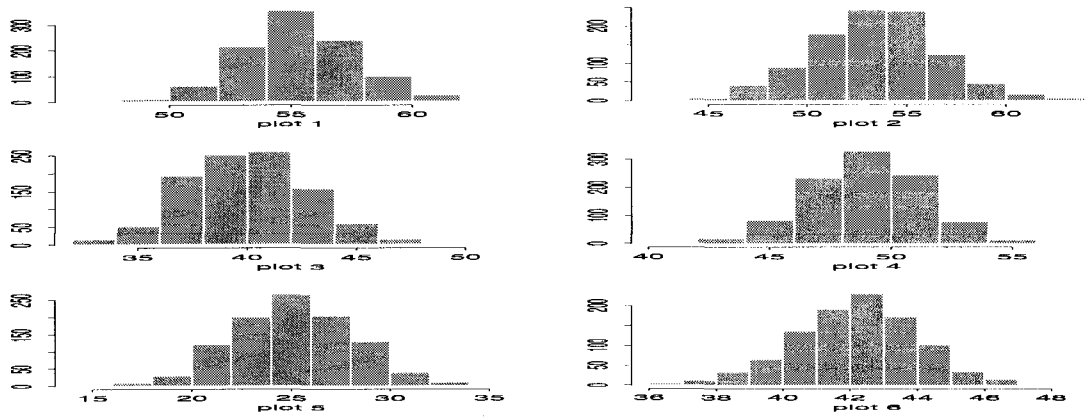
[2] The generated model : $Y_i = 86.31 + 11.05\,\mathrm{Asp}_S - 0.35\,\mathrm{Band}_4 + \varepsilon_i$, $i = 1,\ldots,N,N+1,\ldots,N+6$, where $\boldsymbol{\varepsilon} = (\varepsilon_1,\ldots,\varepsilon_{N+6})$ is $\mathrm{MVN}(\mathbf{0},(22.68)^2\mathbf{I})$.

[3] "NA" indicates the fumula of m.s.p.e. of $\widehat{Y}_{s_0,msn}$ is not extractable.

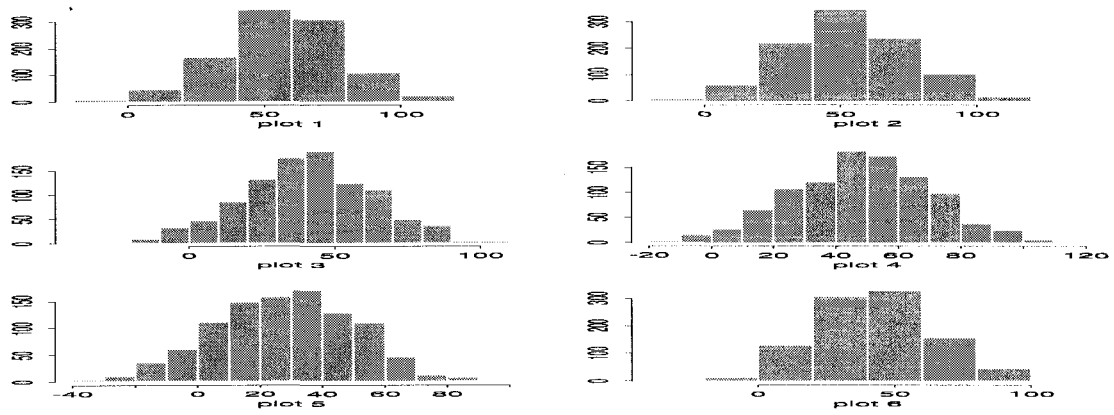
Figure 8.4: 1,000 predictions of Tba from 1,000 generated realizations of model 7.1 [1] at 6 non-sampled plots [2]



(a) the simulated data at 6 non-sampled plots



(b) the Tba predictions using model 7.1 at 6 non-sampled plots



(c) the Tba predictions using MSN at 6 non-sampled plots

[1] The generated model : $Y_i = 86.31 + 11.05\,\mathrm{Asp}_S - 0.35\,\mathrm{Band}_4 + \varepsilon_i$, $i = 1, \ldots, N, N+1, \ldots, N+6$, where $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_{N+6})$ is $\mathrm{MVN}(\mathbf{0}, (22.68)^2\,\mathbf{I})$.

[2] Six non-sample plot locations are depicted in Figure 5.4.

Table 8.3: Comparison of predictors $\widehat{Y}_{s_0,srm}$ and $\widehat{Y}_{s_0,msn}$ at 6 non-sampled plots[1] based on 1,000 Tba realization from spatial regression model 7.6[2]

plot		$\widehat{Y}_{s_0,srm}$	$\widehat{Y}_{s_0,msn}$	$\hat{\sigma}_{\widehat{Y},p,srm}$	$\hat{\sigma}_{\widehat{Y},p,msn}$
1	mean	42.58	55.98	19.01	NA
	s.e	5.61	22.19	1.81	NA
2	mean	59.99	54.76	19.00	NA
	s.e	1.00	22.16	1.81	NA
3	mean	35.97	40.57	18.99	NA
	s.e	0.82	22.04	1.81	NA
4	mean	31.76	48.96	19.00	NA
	s.e	4.79	22.24	1.81	NA
5	mean	25.93	27.78	19.00	NA
	s.e	1.02	22.07	1.81	NA
6	mean	49.01	46.43	19.00	NA
	s.e	2.27	23.13	1.81	NA

(a). The means and s.e.'s of 1,000 estimates $\widehat{Y}_{s_0,srm}$, $\widehat{Y}_{s_0,msn}$, and $\hat{\sigma}_{\widehat{Y},p,srm}$ in Eqn. (8.13)

plot	$\widehat{r.m.s.e.}(\widehat{Y}_{s_0,srm})$	$\widehat{r.m.s.e.}(\widetilde{Y}_{s_0,msn})$
1	22.94	29.27
2	19.15	28.03
3	18.53	27.36
4	26.04	28.40
5	18.00	28.13
6	20.66	25.27

(b). $\widehat{r.m.s.e.}(\hat{Y}_p^i) = \sqrt{\frac{1}{1{,}000}\sum_{b=1}^{1{,}000}(\hat{Y}_{p,N+i}^b - Y_{g,N+i}^b)^2}$,

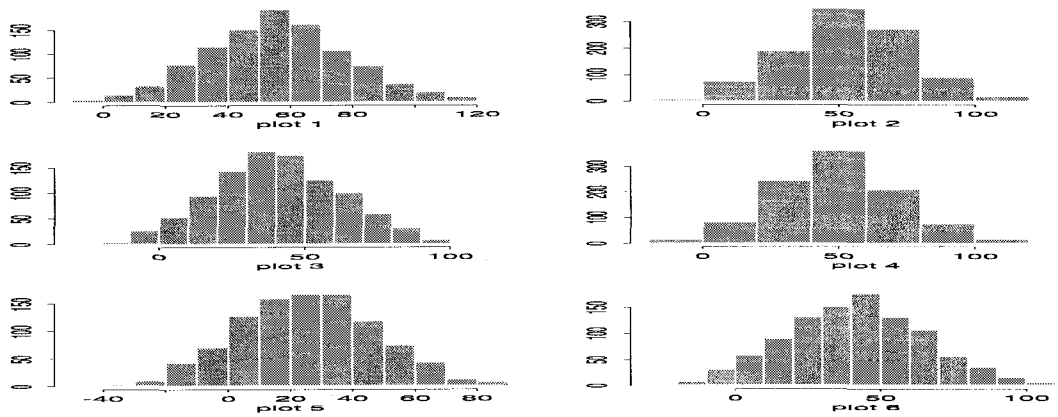
where $\hat{Y}_{p,N+i}^b$, $i = 1,\ldots,6$ denote the prediction at non-sampled plot i using the MSN predictions and the SRM predictions based on the generated values at sampled plots, $1,\ldots,N = 312$ from the b^{th} generated realization of size $(N+6)$, and $Y_{g,N+i}^b$ denotes the b^{th} generated values at non-sampled plot i.

[1] Six non-sample plot locations are depicted in Figure 5.4.

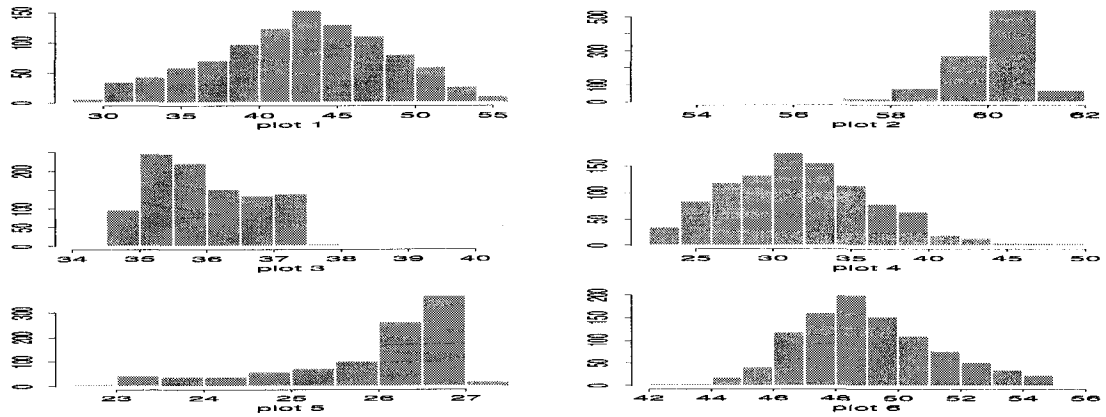
[2] The generated model : $Y_i = 86.31 + 11.05\,\mathrm{Asp}_S - 0.35\,\mathrm{Band}_4 + \delta_i$, $i = 1,\ldots,N, N+1,\ldots,N+6$, where $\boldsymbol{\delta} = (\delta_1,\ldots,\delta_{N+6})$ is $\mathrm{MVN}(\mathbf{0}, (22.68)^2\hat{\Gamma})$, and $\hat{\Gamma} = \hat{B}'\hat{B} = \left(exp(-0.0016 d_{ij}^{0.87})\right)_{i,j=1}^{N+6}$ [3]

"NA" indicates the fumula of m.s.p.e. of $\widehat{Y}_{s_0,msn}$ is not extractable.

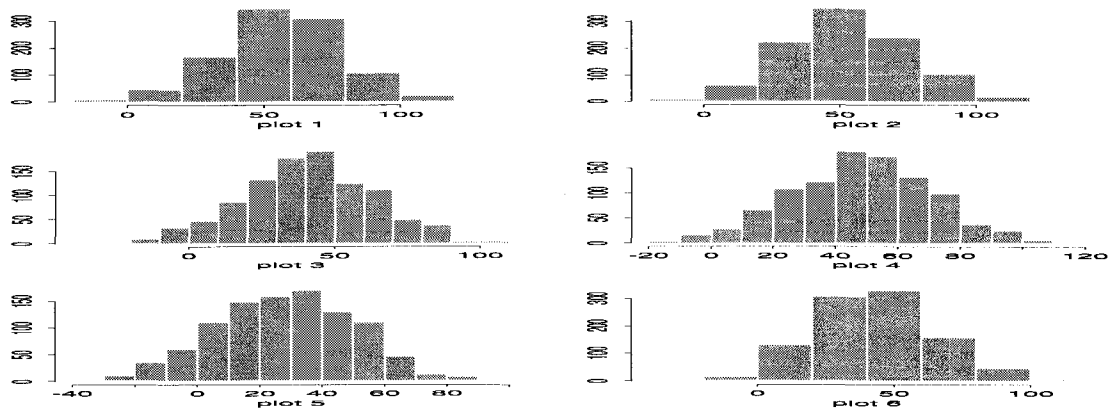
Figure 8.5: 1,000 predictions of Tba from 1,000 generated realizations of model 7.6 [1] at 6 non-sampled plots [2]



(a) the simulated data at 6 non-sampled plots



(b) the Tba predictions using model 7.6 at 6 non-sampled plots

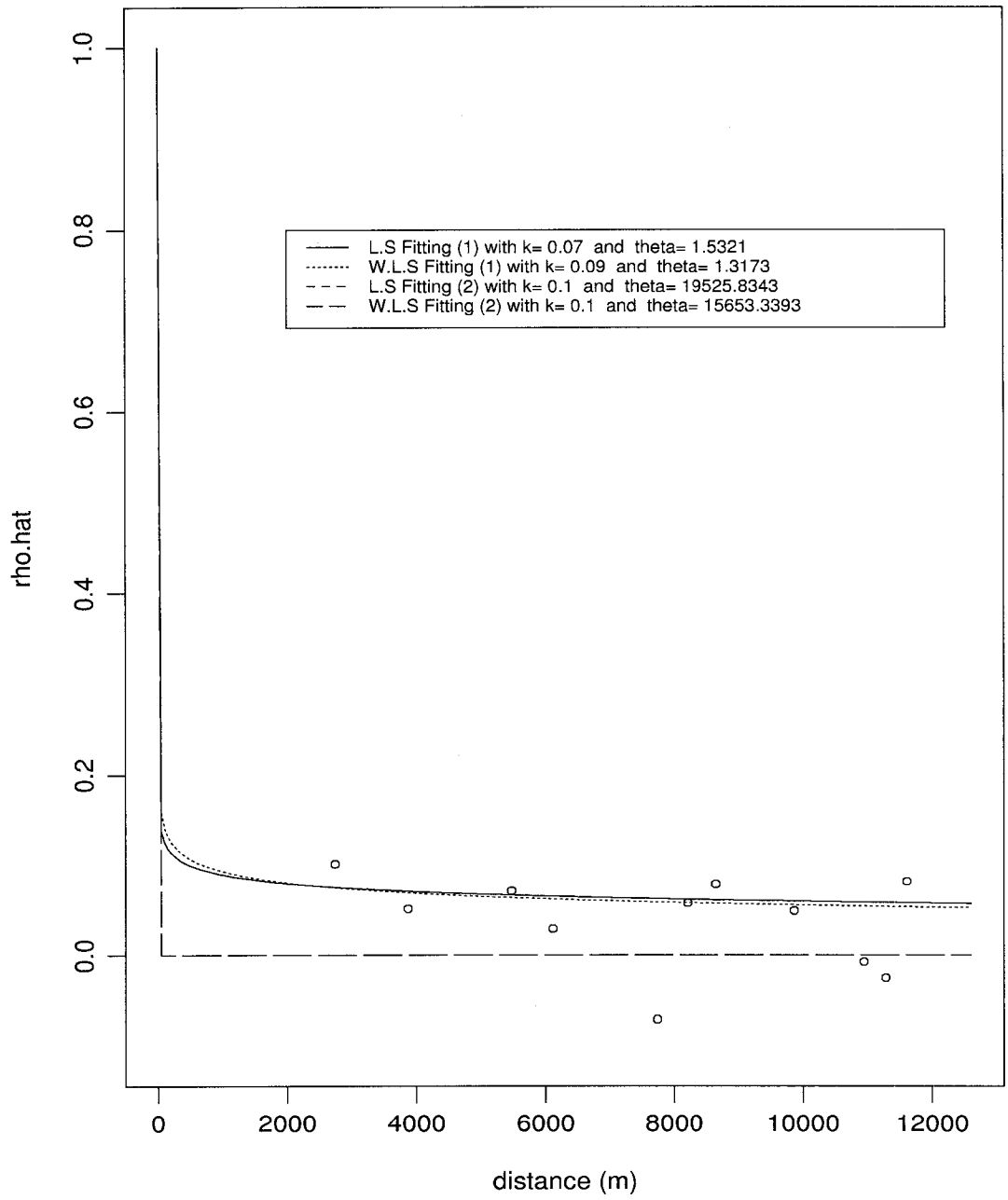


(c) the Tba predictions using MSN at 6 non-sampled plots

[1] The generated model : $Y_i = 86.31 + 11.05\,\mathrm{Asp}_S - 0.35\,\mathrm{Band}_4 + \delta_i$, $i = 1, \ldots, N, N+1, \ldots, N+6$, where $\boldsymbol{\delta} = (\delta, \ldots, \delta_{N+6})$ is $\mathrm{MVN}(\mathbf{0}, \hat{\sigma}^2 \hat{\Gamma})$.

[2] Six non-sample plot locations are depicted in Figure 5.4.

Figure 8.6: Sample correlations and the fitted exponential function for the residuals from Lt data of "Phase II" Model 7.6[1]



[1]

(1) $\rho_{ij}(\boldsymbol{\delta}) = exp(-\theta d_{ij}^{k}),\ \theta, k > 0,$

(2) $\rho_{ij}(\boldsymbol{\delta}) = (1 + \theta d_{ij}^{2})^{(-k)},\ \theta, k > 0.$

Table 8.4: The Lt predictions at the selected 6 non-sampled plot locations from 1,000 bootstrapped samples

			Average of 1,000 Bootstrapped Estimates			
	Plug-in Estimates[1]		Simple Bootstrap		Semiparametric Bootstrap	
plot	$\widehat{Y}_{s_0,glm}$	$\widehat{Y}_{s_0,srm}$	$\overline{\widehat{Y}}_{s_0,glm}$	$\overline{\widehat{Y}}_{s_0,srm}$	$\overline{\widehat{Y}}_{s_0,glm}$	$\overline{\widehat{Y}}_{s_0,srm}$
1	311.12	310.64	310.58	319.21	311.96	311.11
			(21.15)	(24.03)	(24.45)	(0.53)
2	325.47	331.70	325.31	332.90	326.35	333.78
			(17.99)	(12.43)	(21.00)	(1.89)
3	447.76	438.15	450.80	436.10	448.89	436.54
			(35.22)	(22.07)	(30.15)	(1.60)
4	372.51	368.64	373.58	370.41	373.48	368.90
			(16.08)	(3.55)	(15.46)	(0.23)
5	469.90	460.12	473.51	454.30	471.07	456.74
			(42.24)	(30.26)	(36.59)	(3.12)
6	388.00	356.07	389.48	375.18	389.01	352.42
			(18.75)	(16.68)	(16.65)	(4.17)
	Plug-in Estimates[1]		Simple Bootstrap		Semiparametric Bootstrap	
plot	$\widehat{\sigma}_{Y,p,glm}$	$\widehat{\sigma}_{Y,p,srm}$	$\overline{\widehat{\sigma}}_{Y,p,glm}$	$\overline{\widehat{\sigma}}_{Y,p,srm}$	$\overline{\widehat{\sigma}}_{Y,p,glm}$	$\overline{\widehat{\sigma}}_{Y,p,srm}$
1	265.83	258.52	265.59	257.15	265.62	262.53
			(13.38)	(19.35)	(13.42)	0.66
2	265.56	258.14	265.31	256.94	265.35	262.10
			(13.37)	(19.14)	(13.40)	0.66
3	266.47	258.62	266.23	257.54	266.26	262.55
			(13.39)	(19.48)	(13.45)	0.67
4	265.22	257.74	264.98	256.47	265.02	261.71
			(13.35)	(19.37)	(13.39)	0.66
5	267.25	259.87	267.01	258.50	267.04	263.84
			(13.43)	(19.45)	(13.49)	0.67
6	265.30	258.25	265.05	256.63	265.09	262.19
			(13.35)	(19.45)	(13.39)	0.68

[1] The plug-in estimates for $\widehat{\sigma}_{Y,p,glm}$ and $\widehat{\sigma}_{Y,p,srm}$ are calculated by plugging the N observations into the Eqns. (8.6) and (8.13) and the estimated a_i based on the N observations. Also, the sample standard deviations of $\widehat{Y}_{s_0,glm}$, $\widehat{Y}_{s_0,srm}$, $\widehat{\sigma}_{Y,p,glm}$, and $\widehat{\sigma}_{Y,p,srm}$ are listed in parentheses.

Table 8.5: Comparison of predictors $\widehat{Y}_{s_0,glm}$ and $\widehat{Y}_{s_0,msn}$ at 6 non-sampled plots[1] based on 1,000 Lt realization from general linear model 7.1[2]

plot		$\widehat{Y}_{s_0,glm}$	$\widehat{Y}_{s_0,msn}$	$\hat{\sigma}_{\widehat{Y},p,glm}$	$\hat{\sigma}_{\widehat{Y},p,msn}$
1	mean	310.64	308.73	265.68	NA[3]
	s.e.	22.97	264.89	10.52	NA
2	mean	324.97	331.80	265.41	NA
	s.e.	19.77	261.63	16.50	NA
3	mean	447.06	445.80	266.32	NA
	s.e.	29.00	252.01	10.54	NA
4	mean	371.93	360.83	265.07	NA
	s.e.	14.91	272.19	10.49	NA
5	mean	469.15	463.04	267.10	NA
	s.e.	35.08	256.54	10.57	NA
6	mean	387.39	377.78	265.15	NA
	s.e.	16.13	262.15	10.49	NA

(a). The means and s.e.'s of 1,000 estimates $\widehat{Y}_{s_0,glm}$, $\widehat{Y}_{s_0,msn}$, and $\hat{\sigma}_{\widehat{Y},p,glm}$ in Eqn. (8.6)

plot	$\widehat{r.m.s.e.}(\widehat{Y}_{s_0,glm})$	$\widehat{r.m.s.e.}(\widetilde{Y}_{s_0,msn})$
1	265.88	374.72
2	261.49	369.76
3	262.28	367.33
4	270.66	380.46
5	260.37	356.57
6	278.88	382.90

$$(b).\widehat{r.m.s.e.}(\hat{Y}_p^i) = \sqrt{\frac{1}{1{,}000}\sum_{b=1}^{1{,}000}(\hat{Y}^b_{p,N+i} - Y^b_{g,N+i})^2}$$

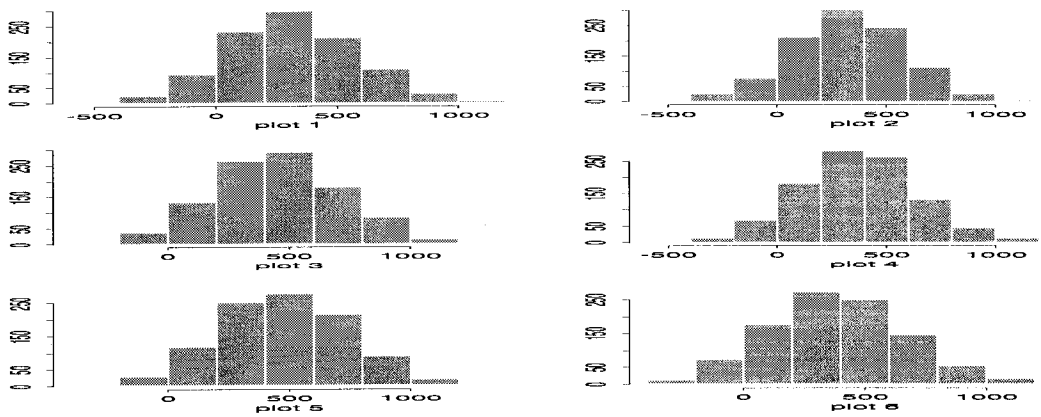
where $\hat{Y}^b_{p,N+i}$, $i = 1,\ldots,6$ denote the prediction at non-sampled plot i using the MSN predictions and the GLM predictions based on the generated values at sampled plots, $1,\ldots,N = 312$ from the b^{th} generated realization of size $(N+6)$, and $Y^b_{g,N+i}$ denotes the b^{th} generated values at non-sampled plot i.

[1] Six non-sample plot locations are depicted in Figure 5.4.

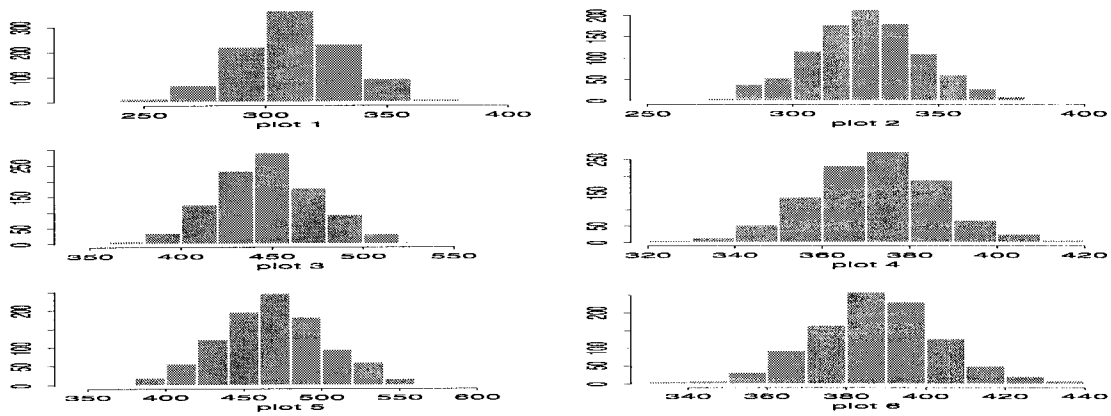
[2] The generated model : $Y_i = 184.37 + 37.63\,\text{Ratio}_i + \varepsilon_i$, $i = 1,\ldots,N, N+1,\ldots,N+6$, where $\boldsymbol{\varepsilon}=(\varepsilon_1,\ldots,\varepsilon_{N+6})$ is $\text{MVN}(\mathbf{0},(265.64)^2\,\mathbf{I})$.

[3] "NA" indicates the fumula of m.s.p.e. of $\widehat{Y}_{s_0,msn}$ is not extractable.

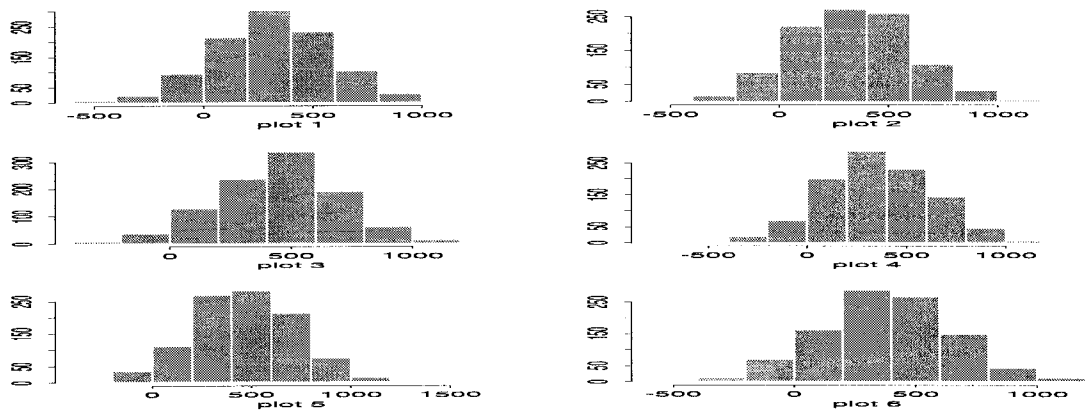
Figure 8.7: 1,000 predictions of Lt from 1,000 generated realizations of model 7.1 [1] at 6 non-sampled plots [2]



(a) the simulated Lt data at 6 non-sampled plots



(b) the Lt predictions using model 7.1 at 6 non-sampled plots



(c) the Lt predictions using MSN at 6 non-sampled plots

[1] The generated model : $Y_i = 184.37 + 37.63\,\text{Ratio}_1 + \varepsilon_i$, $i = 1, \ldots, N, N+1, \ldots, N+6$, where $\boldsymbol{\varepsilon}=(\varepsilon_1, \ldots, \varepsilon_{N+6})$ is $\text{MVN}(\mathbf{0}, (265.64)^2\,\mathbf{I})$.

[2] Six non-sample plot locations are depicted in Figure 5.4.

Table 8.6: Comparison of predictors $\widehat{Y}_{s_0,srm}$ and $\widehat{Y}_{s_0,msn}$ at 6 non-sampled plots[1] based on 1,000 Lt realization from spatial regression model 7.6[2]

plot		$\widehat{Y}_{s_0,srm}$	$\widehat{Y}_{s_0,msn}$	$\hat{\sigma}_{\widehat{Y},p,srm}$	$\hat{\sigma}_{\widehat{Y},p,msn}$
1	mean	317.40	304.79	252.85	NA[3]
	s.e.	17.20	289.36	13.48	NA
2	mean	334.44	325.18	252.71	NA
	s.e.	11.54	262.48	13.47	NA
3	mean	435.37	429.71	252.60	NA
	s.e.	17.56	250.76	13.45	NA
4	mean	370.35	381.46	252.75	NA
	s.e.	2.72	261.02	13.46	NA
5	mean	452.75	413.16	252.80	NA
	s.e.	25.27	264.87	13.48	NA
6	mean	368.21	410.84	252.85	NA
	s.e.	19.81	271.81	13.50	NA

(a). The means and s.e.'s of 1,000 estimates $\widehat{Y}_{s_0,srm}$, $\widehat{Y}_{s_0,msn}$, and $\hat{\sigma}_{\widehat{Y},p,srm}$ in Eqn. (8.13)

plot	$\widehat{r.m.s.e.}(\widehat{Y}_{s_0,srm})$	$\widehat{r.m.s.e.}(\widehat{Y}_{s_0,msn})$
1	256.46	372.08
2	252.22	369.85
3	252.65	343.94
4	261.91	370.49
5	252.51	365.91
6	271.99	375.62

$$(b).\widehat{r.m.s.e.}(\hat{Y}_p^i) = \sqrt{\frac{1}{1{,}000}\sum_{b=1}^{1{,}000}(\hat{Y}_{p,N+i}^b - Y_{g,N+i}^b)^2}$$

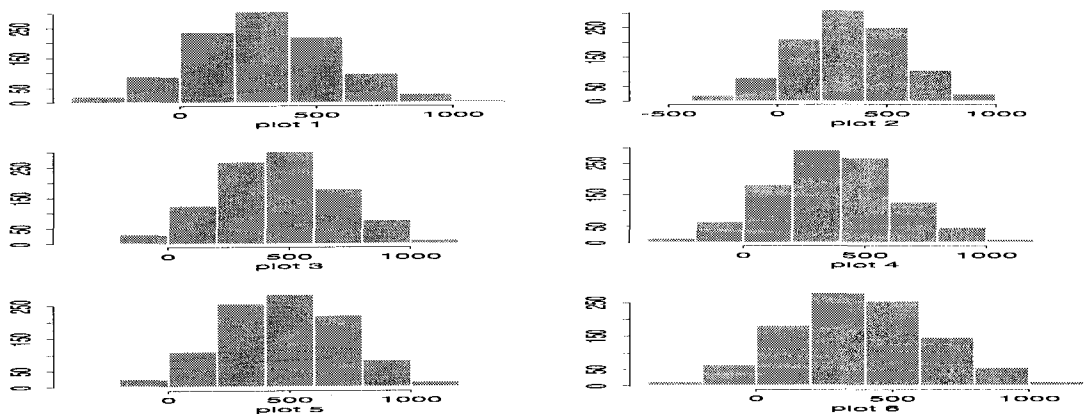
where $\hat{Y}_{p,N+i}^b$, $i = 1,\ldots,6$ denote the prediction at non-sampled plot i using the MSN predictions and the SRM predictions based on the generated values at sampled plots, $1,\ldots,N = 312$ from the b^{th} generated realization of size $(N+6)$, and $Y_{g,N+i}^b$ denotes the b^{th} generated values at non-sampled plot i.

[1] Six non-sample plot locations are depicted in Figure 5.4.

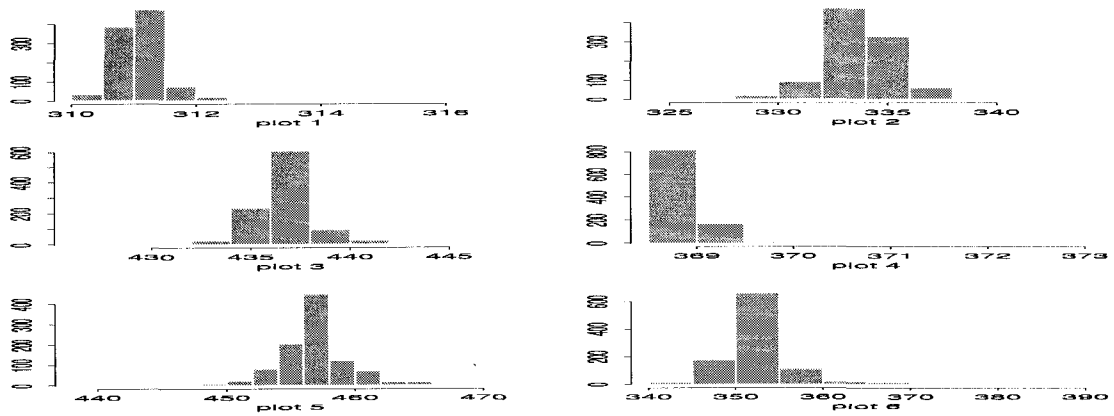
[2] The generated model : $Y_i = 184.37 + 37.63\,\text{Ratio}_1 + \delta_i$, $i = 1,\ldots,N, N+1,\ldots,N+6$, where $\boldsymbol{\delta}=(\delta_1,\ldots,\delta_{N+6})$ is $\text{MVN}(\mathbf{0},(265.64)^2\hat{\Gamma})$, and $\hat{\Gamma} = \hat{B}'\hat{B}=\left(exp(-1.3173d_{ij}^{0.09})\right)_{i,j=1}^{N+6}$

[3] "NA" indicates the fumula of m.s.p.e. of $\widehat{Y}_{s_0,msn}$ is not extractable.

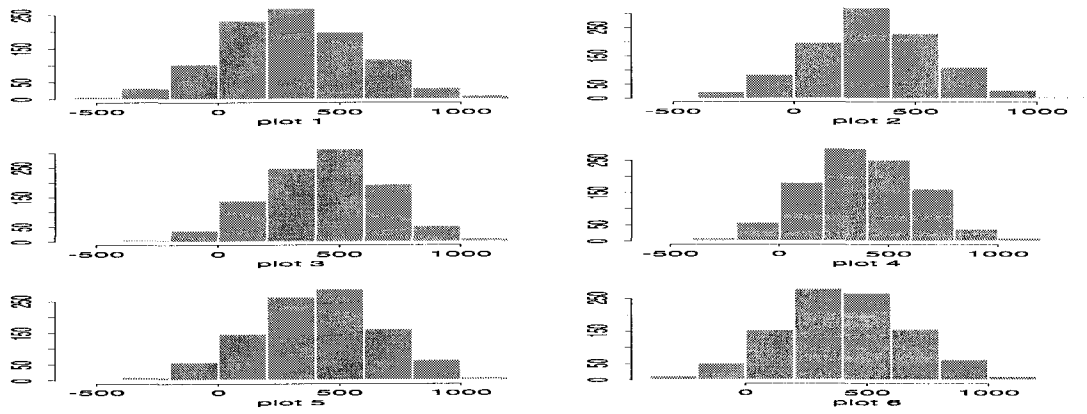
Figure 8.8: 1,000 predictions of Lt from 1,000 generated realizations of model 7.6 [1]at 6 non-sampled plots [2]



(a) the simulated data at 6 non-sampled plots



(b) the Lt predictions using model 7.6 at 6 non-sampled plots

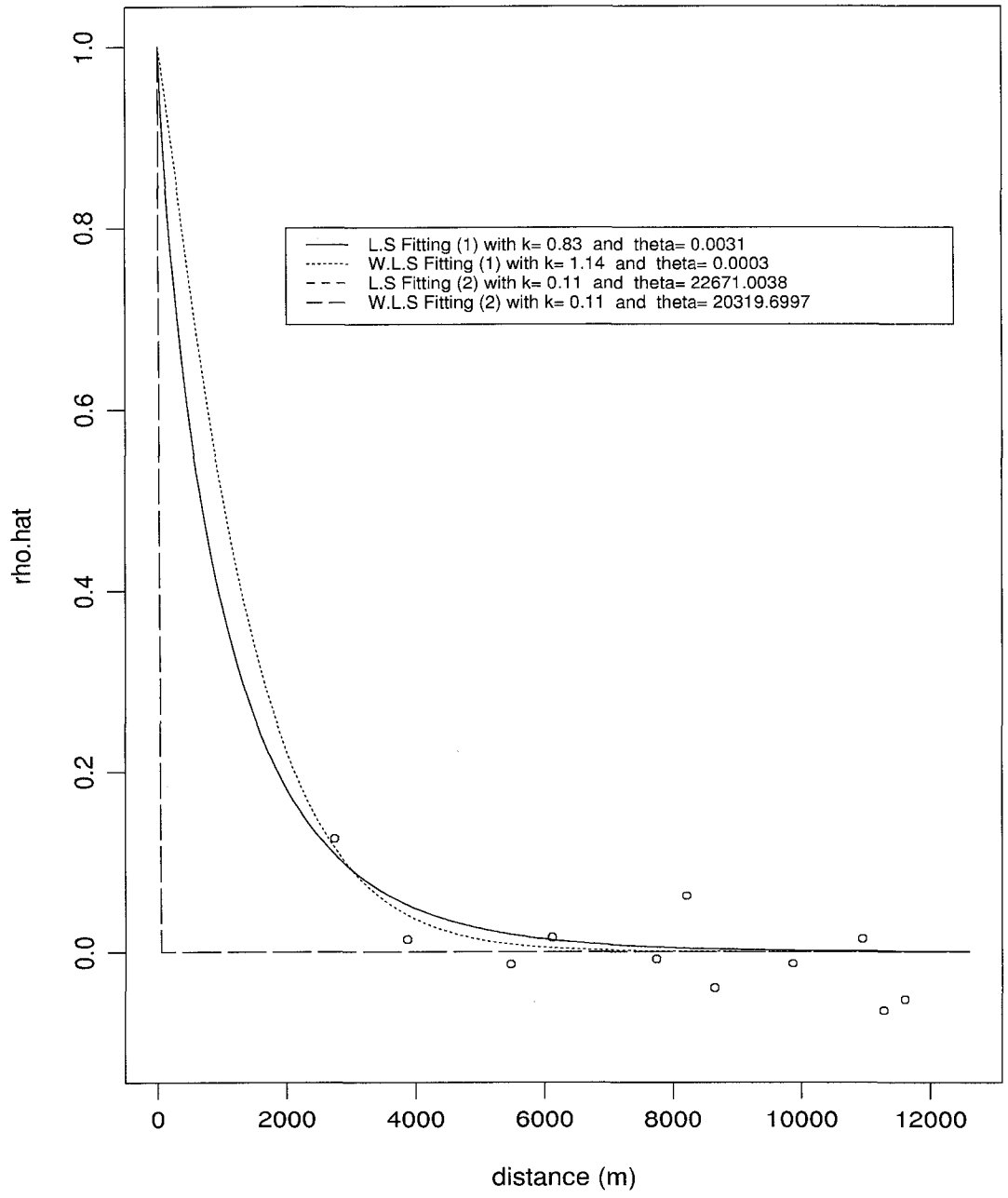


(c) the Lt predictions using MSN at 6 non-sampled plots

[1] The generated model : $Y_i = 184.37 + 37.63\,\mathrm{Ratio}_1 + \delta_i$, $i = 1, \ldots, N, N+1, \ldots, N+6$, where $\boldsymbol{\delta}=(\delta, \ldots, \delta_{N+6})$ is $\mathrm{MVN}(\mathbf{0}, \hat{\sigma}^2\hat{\Gamma})$.

[2] Six non-sample plots are depicted in Figure 5.4.

Figure 8.9: Sample correlations and the fitted exponential function for the residuals from M data of "Phase II" Model 7.6[1]



[1]

(1) $\rho_{ij}(\boldsymbol{\delta}) = exp(-\theta d_{ij}^{k}),\ \theta, k > 0,$

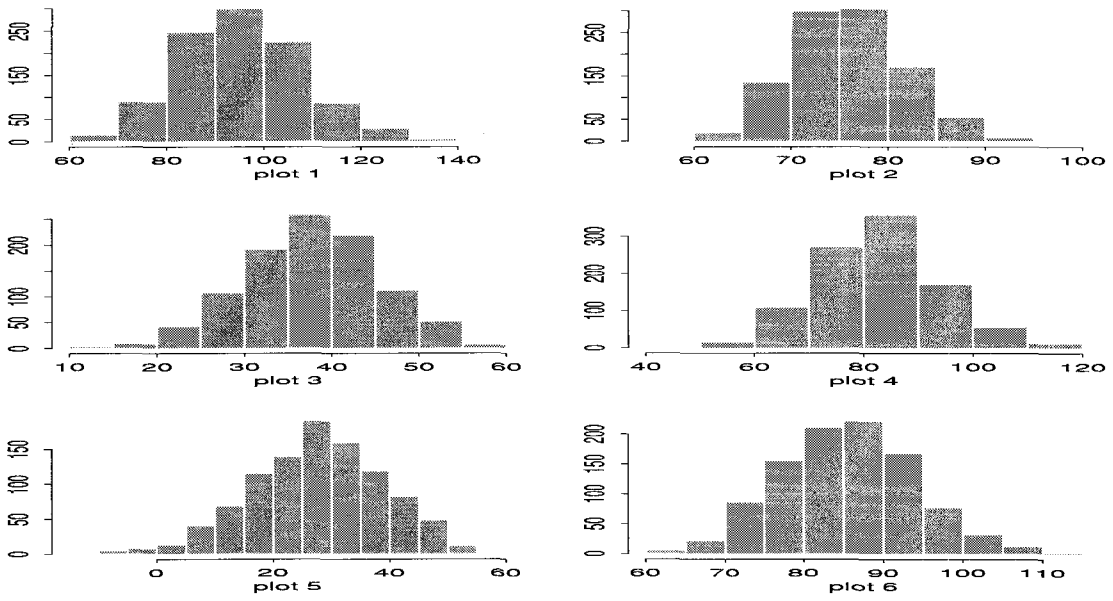
(2) $\rho_{ij}(\boldsymbol{\delta}) = (1 + \theta d_{ij}^{2})^{(-k)},\ \theta, k > 0.$

Table 8.7: The M predictions at the selected 6 non-sampled plot locations from 1,000 bootstrapped samples

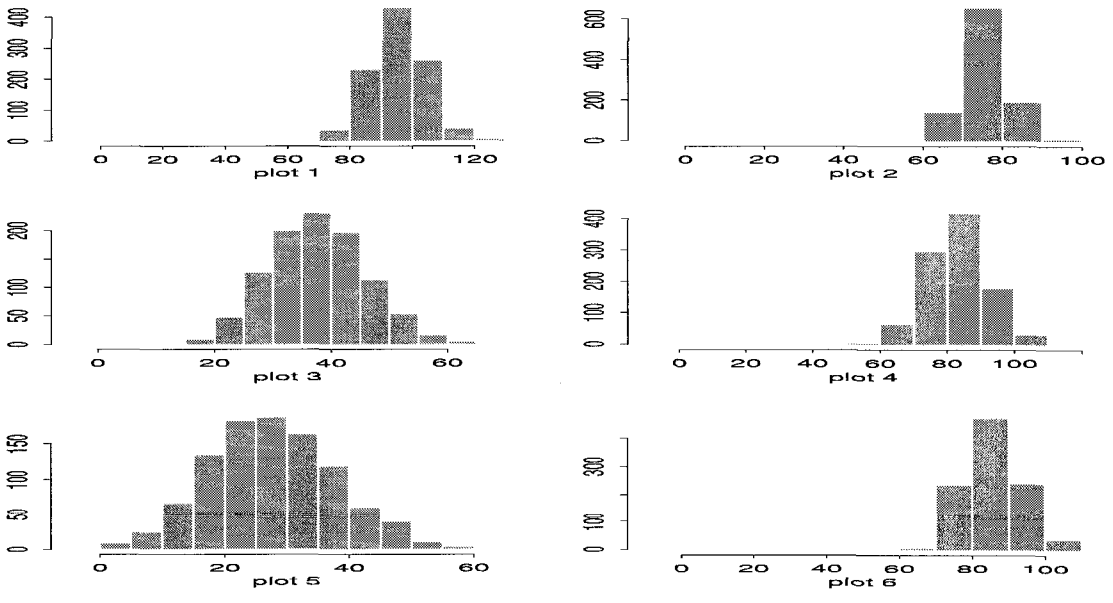
			Average of 1,000 Bootstrapped Estimates			
	Plug-in Estimates[1]		Simple Bootstrap		Semiparametric Bootstrap	
plot	$\widehat{Y}_{s_0,glm}$	$\widehat{Y}_{s_0,srm}$	$\overline{\widehat{Y}}_{s_0,glm}$	$\overline{\widehat{Y}}_{s_0,srm}$	$\overline{\widehat{Y}}_{s_0,glm}$	$\overline{\widehat{Y}}_{s_0,srm}$
1	95.35	53.43	95.45	90.30	95.42	53.93
			(12.71)	(11.18)	(9.32)	(4.95)
2	75.67	70.26	75.91	75.35	75.66	70.95
			(5.82)	(0.96)	(5.92)	(1.40)
3	37.96	43.55	37.63	39.10	37.61	44.92
			(7.78)	(2.95)	(8.47)	(0.36)
4	83.05	52.31	82.90	77.63	82.89	51.61
			(11.44)	(9.35)	(9.23)	(2.27)
5	28.02	45.87	27.59	33.43	27.76	47.94
			(11.32)	(8.07)	(10.18)	(0.47)
6	85.48	126.97	85.42	88.80	85.46	123.68
			(8.55)	(8.14)	(8.01)	(6.93)
	Plug-in Estimates[1]		Simple Bootstrap		Semiparametric Bootstrap	
plot	$\widehat{\sigma}_{Y,p,glm}$	$\widehat{\sigma}_{Y,p,srm}$	$\overline{\widehat{\sigma}}_{Y,p,glm}$	$\overline{\widehat{\sigma}}_{Y,p,srm}$	$\overline{\widehat{\sigma}}_{Y,p,glm}$	$\overline{\widehat{\sigma}}_{Y,p,srm}$
1	65.09	53.90	64.11	61.89	64.15	55.76
			(6.45)	(7.36)	(6.47)	(2.35)
2	64.68	54.07	63.71	61.54	63.76	55.91
			(6.41)	(7.27)	(6.43)	(2.24)
3	64.98	53.63	64.02	61.71	64.05	55.46
			(6.45)	(7.37)	(6.46)	(2.37)
4	65.03	53.52	64.06	61.77	64.10	55.37
			(6.45)	(7.39)	(6.47)	(2.38)
5	65.25	54.08	64.28	62.04	64.32	55.93
			(6.47)	(7.38)	(6.49)	(2.33)
6	64.91	53.19	63.94	61.63	63.98	55.03
			(6.44)	(7.40)	(6.45)	(2.40)

[1] The plug-in estimates for $\widehat{\sigma}_{Y,p,glm}$ and $\widehat{\sigma}_{Y,p,srm}$ are calculated by plugging the N observations into the Eqns. (8.6) and (8.13) and the estimated a_i based on the N observations. Also, the sample standard deviations of $\widehat{Y}_{s_0,glm}$, $\widehat{Y}_{s_0,srm}$, $\widehat{\sigma}_{Y,p,glm}$, and $\widehat{\sigma}_{Y,p,srm}$ are listed in parentheses.

Figure 8.10: 1,000 bootstrapping predictions of M from model 7.1 at 6 non-sample plots [1]



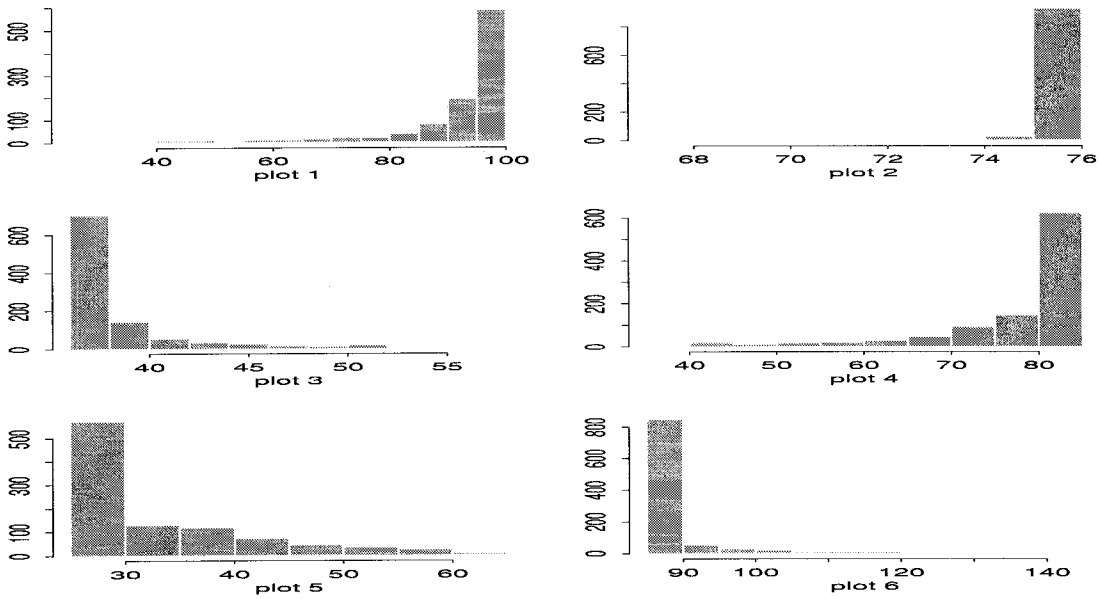
(a) using the simple bootstrap method



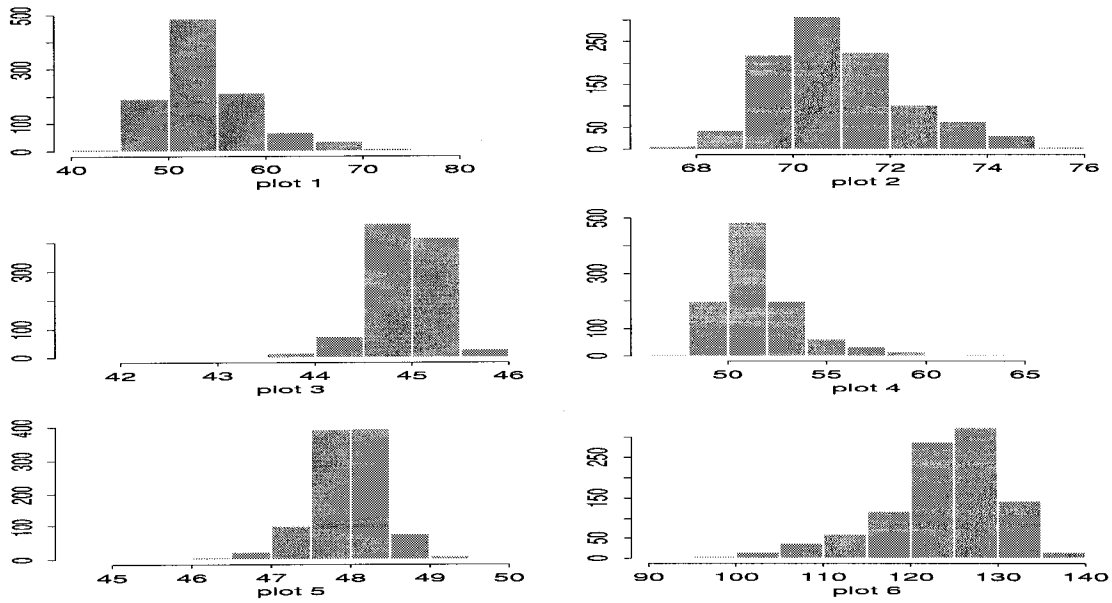
(b) using the semiparametric bootstrap method

[1] Six non-sample plot locations are depicted in Figure 5.4

Figure 8.11: 1,000 bootstrapping predictions of M from model 7.6 at 6 non-sample plots [1]



(a) using the simple bootstrap method



(b) using the semiparametric bootstrap method

[1] Six non-sample plot locations are depicted in Figure 5.4

Table 8.8: Comparison of predictors $\widehat{Y}_{s_0,glm}$ and $\widehat{Y}_{s_0,msn}$ at 6 non-sampled plots[1] based on 1,000 M realization from general linear model 7.1[2]

plot		$\widehat{Y}_{s_0,glm}$	$\widehat{Y}_{s_0,msn}$	$\hat{\sigma}_{\widehat{Y},p,glm}$	$\hat{\sigma}_{\widehat{Y},p,msn}$
1	mean	95.35	95.50	65.04	NA[3]
	s.e.	8.96	66.17	2.58	NA
2	mean	75.74	75.55	64.64	NA
	s.e.	5.51	63.99	2.57	NA
3	mean	37.86	51.32	64.94	NA
	s.e.	8.30	63.05	2.58	NA
4	mean	82.89	76.16	64.98	NA
	s.e.	8.67	65.31	2.58	NA
5	mean	28.09	33.77	65.21	NA
	s.e.	9.93	67.13	2.59	NA
6	mean	85.65	97.28	64.86	NA
	s.e.	7.90	65.95	2.58	NA

(a). The means and s.e.'s of 1,000 estimates $\widehat{Y}_{s_0,glm}$, $\widehat{Y}_{s_0,msn}$, and $\hat{\sigma}_{\widehat{Y},p,glm}$ in Eqn. (8.6)

plot	$\widehat{r.m.s.e.}(\widehat{Y}_{s_0,glm})$	$\widehat{r.m.s.e.}(\widetilde{Y}_{s_0,msn})$
1	65.78	92.21
2	63.92	90.11
3	64.43	88.51
4	66.63	92.29
5	63.96	92.88
6	68.53	94.27

$$(b).\widehat{r.m.s.e.}(\hat{Y}_p^i) = \sqrt{\frac{1}{1{,}000}\sum_{b=1}^{1{,}000}(\hat{Y}_{p,N+i}^b - Y_{g,N+i}^b)^2}$$

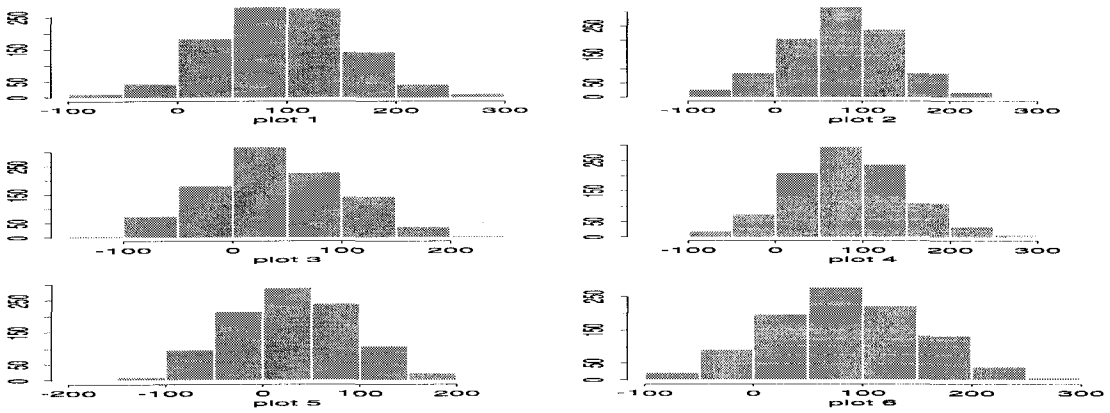
where $\hat{Y}_{p,N+i}^b$, $i = 1,\ldots,6$ denote the prediction at non-sampled plot i using the MSN predictions and the GLM predictions based on the generated values at sampled plots, $1,\ldots,N = 312$ from the b^{th} generated realization of size $(N+6)$, and $Y_{g,N+i}^b$ denotes the b^{th} generated values at non-sampled plot i.

[1] Six non-sample plot locations are depicted in Figure 5.4.

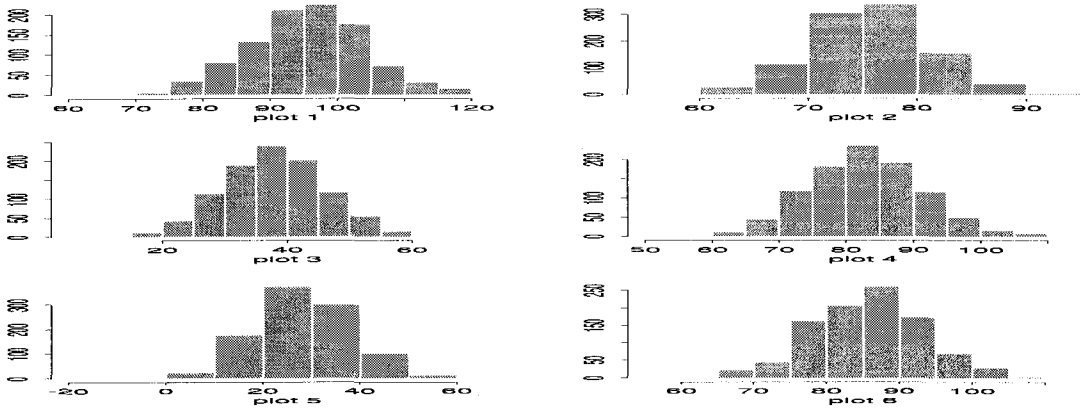
[2] The generated model : $Y_i = 100.37 + 0.08\,\text{Elev} - 1.87,\text{Band}_7 + -7.26\,\text{Ratio}_1 + \varepsilon_i$, $i = 1,\ldots,N,N+1,\ldots,N+6$, where $\boldsymbol{\varepsilon}=(\varepsilon_1,\ldots,\varepsilon_{N+6})$ is $\text{MVN}(\mathbf{0},(64.97)^2\,\mathbf{I})$.

[3] "NA" indicates the fumula of m.s.p.e. of $\widehat{Y}_{s_0,msn}$ is not extractable.

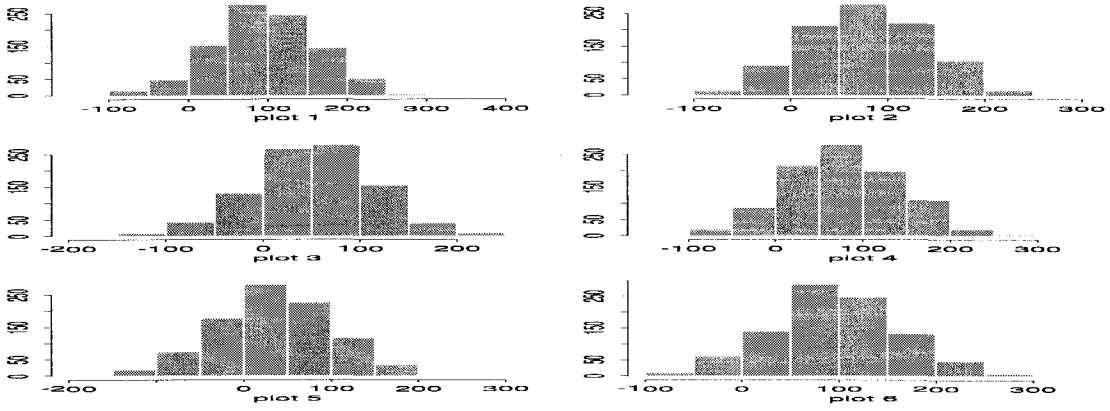
Figure 8.12: 1,000 predictions of M from 1,000 generated realizations of model 7.1 [1] at 6 non-sampled plots [2]



(a) the simulated data at 6 non-sampled plots



(b) the M predictions using model 7.1 at 6 non-sampled plots



(c) the M predictions using MSN at 6 non-sampled plots

[1] The generated model : $Y_i = 100.37 + 0.08\,\text{Elev} - 1.87, \text{Band}_7 + -7.26\,\text{Ratio}_1 + \varepsilon_i,\ i = 1, \ldots, N, N+1, \ldots, N+6$, where $\boldsymbol{\varepsilon}=(\varepsilon_1, \ldots, \varepsilon_{N+6})$ is $\text{MVN}(\mathbf{0}, (64.97)^2\,\mathbf{I})$.

[2] Six non-sample plot locations are depicted in Figure 5.4.

Table 8.9: Comparison of predictors $\widehat{Y}_{s_0,srm}$ and $\widehat{Y}_{s_0,msn}$ at 6 non-sampled plots[1] based on 1,000 M realization from spatial regression model 7.6[2]

plot		$\widehat{Y}_{s_0,srm}$	$\widehat{Y}_{s_0,msn}$	$\hat{\sigma}_{\widehat{Y},p,srm}$	$\hat{\sigma}_{\widehat{Y},p,msn}$
1	mean	80.31	101.47	60.30	NA[3]
	s.e.	13.50	64.07	4.39	NA
2	mean	74.75	76.89	60.26	NA
	s.e.	1.73	68.56	4.38	NA
3	mean	41.21	44.94	60.23	NA
	s.e.	2.76	62.22	4.38	NA
4	mean	68.02	70.84	60.28	NA
	s.e.	9.75	65.20	4.38	NA
5	mean	41.99	41.07	60.28	NA
	s.e.	6.93	63.15	4.38	NA
6	mean	95.80	83.40	60.29	NA
	s.e.	12.52	66.72	4.40	NA

(a). The means and s.e.'s of 1,000 estimates $\widehat{Y}_{s_0,srm}$, $\widehat{Y}_{s_0,msn}$, and $\hat{\sigma}_{\widehat{Y},p,srm}$ in Eqn. (8.13)

plot	$\widehat{r.m.s.e.}(\widehat{Y}_{s_0,srm})$	$\widehat{r.m.s.e.}(\widehat{Y}_{s_0,msn})$
1	56.20	82.29
2	52.07	87.20
3	51.91	77.97
4	56.89	85.33
5	54.26	82.00
6	57.96	81.69

$$\text{(b)}.\widehat{r.m.s.e.}(\hat{Y}_p^i) = \sqrt{\frac{1}{1{,}000}\sum_{b=1}^{1{,}000}(\hat{Y}_{p,N+i}^b - Y_{g,N+i}^b)^2}$$

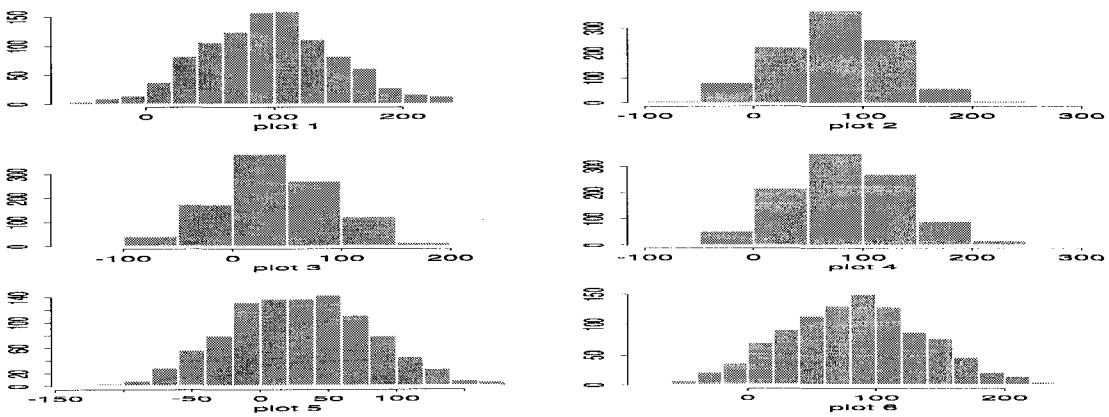
where $\hat{Y}_{p,N+i}^b$, $i = 1,\ldots,6$ denote the prediction at non-sampled plot i using the MSN predictions and the SRM predictions based on the generated values at sampled plots, $1,\ldots,N = 312$ from the b^{th} generated realization of size $(N+6)$, and $Y_{g,N+i}^b$ denotes the b^{th} generated values at non-sampled plot i.

[1] Six non-sample plot locations are depicted in Figure 5.4.

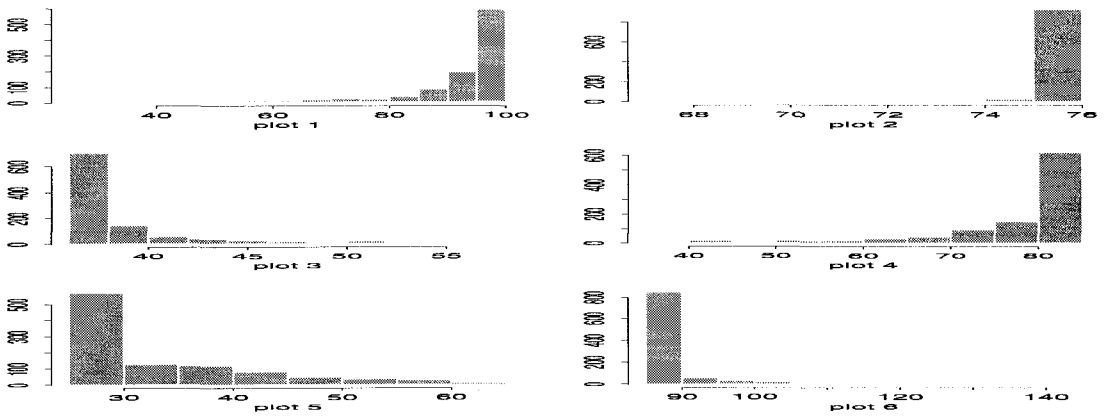
[2] The generated model : $Y_i = 100.37 + 0.08\,\text{Elev} - 1.87,\text{Band}_7 + -7.26\,\text{Ratio}_1 + \varepsilon_i$, $i = 1,\ldots,N,N+1,\ldots,N+6$, where $\boldsymbol{\varepsilon}=(\varepsilon_1,\ldots,\varepsilon_{N+6})$ is $\text{MVN}(\mathbf{0},\hat{\sigma}^2\hat{\Gamma})$, and $\hat{\Gamma} = \hat{B}'\hat{B} = (exp(-0.0003d_{ij}^{1.14}))_{i,j=1}^{N+6}$

[3] "NA" indicates the fumula of m.s.p.e. of $\widehat{Y}_{s_0,msn}$ is not extractable.

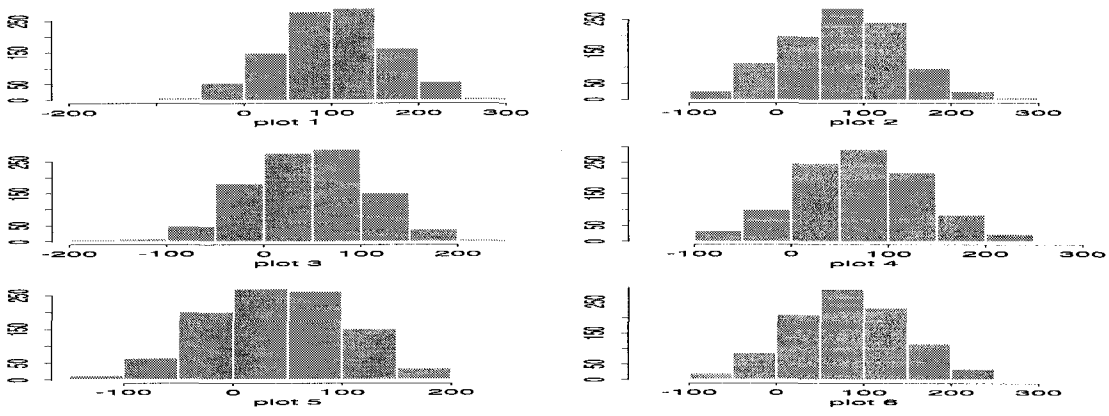
Figure 8.13: 1,000 predictions of from 1,000 generated M realizations of model 7.6 [1] at 6 non-sampled plots [2]



(a) the simulated data at 6 non-sampled plots



(b) the M predictions using model 7.6 at 6 non-sampled plots



(c) the M predictions using MSN at 6 non-sampled plots

[1] The generated model : $Y_i = 86.31 + 11.05\,\mathrm{Asp}_S - 0.35\,\mathrm{Band}_4 + \delta_i$, $i = 1, \ldots, N, N+1, \ldots, N+6$, where $\boldsymbol{\delta} = (\delta, \ldots, \delta_{N+6})$ is $\mathbf{MVN}(\mathbf{0}, \sigma^2 \hat{\Gamma})$.

[2] Six non-sample plot locations are depicted in Figure 5.4.

9. CONCLUSIONS AND FURTHER RESEARCH

We selected the response variables M, M_s, Lt, Lt_s, Tba, and Tba_s for this study because of their economic and ecological importance. This chapter starts with a summary of the methods described in Chapters 4-8 and then makes specific conclusions regarding those methods. Lastly, some further research is suggested.

The overall results are likely to be quite disappointing to practitioners of small area estimation in forestry science. There is little spatial dependence among the response variables at the 1.7-mile grid scale and consequently such dependence is of little use in making predictions at non-sampled plots. Also, there is little useful correlation between our response variables and the many available auxiliary variables, including those from satellite imagery. Auxiliary data contributed little to improving predictions in our study. A need for models that are capable of accommodating a significant number of zeros in some of our response variables was recognized and such models were introduced; but they too were unable to improve predictions. On the positive side, at the subplot scale of the data, there was sufficient spatial dependence to be useful in reducing prediction errors.

9.1 Discussion

As mentioned in Chapter 1, the making of predictions is one of the objectives for this study. In the following subsections, we will discuss the prediction rules used in this thesis.

9.1.1 Predictions at unobserved plot locations

There are two types of plots we desire to predict for: one is for unobserved or non-sampled plots on the 1.7-mile grid and the other is for non-sampled plots on the 0.85-mile grid.

For predictions, there are three types of predictors in this paper:

1. predict with the sample mean $\overline{Y}$,
2. predict with linear combinations of neighboring observations,
3. predict with non-linear combinations of neighboring observations, naming back-transformation predictions.

Type 2 above can be predictions based on models with or without auxiliary information.

Chapter 4 introduced models with different correlation schemes based on lattice data with or without holes. Such non-distance related models based on the 1.7-mile plot data don't allow for prediction of the response variable for non-sampled plots on the 0.85-mile grid. They can only make predictions for individual non-sampled plots on the 1.7-mile grid. All responses are clearly nonnormal except possibly for Tba. Appropriate transformations to normality are sometimes plausible and are considered, but with many zeros transformations to normality are not possible. Models that accommodate the many zeros are introduced in Chapter 6.

In Chapter 5, linear combinations of the observations and distance-based correlation functions were used to study predictions. Only correlation functions that were positive and decreased with increasing distance were studied. The direct predictor $\widetilde{Y}_{s_0}$ (see Eqn. 5.20) is based on fitting a specific function to the sample correlations. The direct predictor $\widetilde{Y}_{s_0}$ doesn't assume that $Y(.)$ follows a Gaussian process.

Since most of our data are not normally distributed, several transformations to normality were tried. The 3-parameter lognormal transformation worked reasonably well for mortality (M), number of live trees (Lt), and total basal area (Tba) data. We fit distance-related autocorrelation functions to the untransformed data (the Y scale) and the transformed data (the Z scale). For correlation function fitting, we consider the direct fitting to the Y scale and indirect fitting to the Z scale. Then the predictions for non-sampled plots can be obtained by the predictions based on the Y scale or back-transforming the predictions based on the Z scale. The predictor is no longer a linear prediction after back-transformations on the Z scale. Back-transformation type prediction requires specification of the function that does the transforming.

For plot data, we used 1-fold cross validation for 33 plots that had all 8 first and second order neighbors using the predictors described. These predictions were compared to the one using the sample mean of the remaining $(N-1) = 311$ observations to predict the individual response at the withheld plot. Among the predictors, direct, indirect, "plain" back-transformed, "approximately" unbiased back-transformed, and "unbiased" back-transformed predictors, the smallest prediction error belonged to the direct predictor, $\widetilde{Y}_{s_0}$. But there is only slight improvement in the overall estimated RMSEs using those predictors over using just the sample mean as the predictor. Those predictions don't work well because there is essentially weak or no spatial dependence at the scale of the data. We also used the prediction rules to evaluate model selection for subplot based data and then the estimated cross-validated RMSEs using the predictors in this chapter were much small than the one using the sample mean. The conclusion is that plots need to be sampled at much shorter distance than 1.7 mile or even shorter than 0.85 mile to have useful spatial information. Maybe subplots should be used as primary sampling units instead of the plots currently used for the Current Veg-

etation Survey (CVS) in the Siuslaw Forest. And there is no need to sample all five of the subplots in each plot to get useful spatial information. A more efficient way than the current grid sampling strategy to get useful spatial information is to sample some proportion of the plots on shorter grid distance than sampling most of the plots on the 1.7 mile grid. An alternative technique for data collection is to sample 3 subplots out of 5 subplots in each plot on a shorter grid such as 0.85 mile or 0.44 mile. In so doing, we can sample a number of subplots on the 0.85-mile grid system and get much more useful spatial information that is needed.

The back-transformation approach leads to an increase in mean squared prediction error (m.s.p.e.) of prediction with or without bias correction; especially for a process with large variance. Since there is no model assumption for the direct predictor $\widetilde{Y}_{s_0}$, it is unnecessary to make transformations. For prediction purposes, the simple predictor $\widetilde{Y}_{s_0}$ is suggested. For illustration, we used the predictor to get predictions at 6 selected non-sampled plots on the 0.85-mile grid. The results for each plot appear similar for Tba and Lt data, but quite different for M data. The M predictions ranged from 14.16 to 136.13 by the direct predictor $\widetilde{Y}_{s_0}$.

Another goal of this study is to determine reliable approaches for error estimation in the predictions obtained by the methods above. We will discuss the bootstrap methods used to achieve it in the following subsection.

9.1.2 Prediction error estimation for an individual plot prediction

Prediction error estimation for individual plot prediction is categorized as follows:

(1) If there is an closed form for the prediction error of an individual prediction, and

(a) if the model is correctly specified and there exists strong spatial dependence among the data, the semiparametric bootstrap method is suggested, or,

(b) if not, the simple bootstrap method will be ok and it is not as time-consuming as the semiparametric bootstrap method.

(2) If there is no closed form for the prediction error of an individual prediction, conduct an appropriate simulation which can mimic the realization of the observed data and then estimate the prediction error by the estimator

$$\widehat{r.m.s.e.}(\hat{Y}_p^i) = \sqrt{\frac{1}{1,000}\sum_{b=1}^{1,000}(\hat{Y}^b_{p,N+i} - Y^b_{g,N+i})^2}$$

where $\hat{Y}^b_{p,N+i}$ denote the prediction at non-sampled plot i, $i = 1, \ldots, N'$, N' is the number of future plots we would like to predict, using those three prediction rules based on the generated values at sampled plots, $1, \ldots, N$ from the b^{th} generated realization of size $(N + N')$, and $Y^b_{g,N+i}$ denotes the b^{th} generated values at non-sampled plot i.

There is an closed form for prediction error of the predictor $\widetilde{Y}_{s_0}$. Due to weak spatial dependence for the plot based data, there is not much difference between the bootstrapped mean squared prediction errors of prediction at the 6 non-sampled plots from simple bootstrapping and semiparametric bootstrapping.

To deal with lots of zero outcomes as in the seedlings data, in Chapter 6 spatial zero-inflated models that accommodate spatial dependence were proposed. We considered a spatial zero-inflated Poisson (SZIP) model and a spatial zero-inflated Negative Binomial (SZINB) model for discrete data; and a spatial zero-inflated Exponential (SZIE) model and spatial zero-inflated Gamma (SZIG) model for data that can be considered continuous except for the zero values. The prediction

rules described in this chapter can only make predictions for plots on the 1.7 mile grid; in other words, they can only make extrapolations for plots on the 1.7 mile grid in the area in the neighborhood of the Siuslaw National Forest. They don't allow predictions for plots on more intensive grids.

In Chapter 7, we discussed the following procedures to incorporate the auxiliary data: (1) general linear model, (2) spatial regression model with $\boldsymbol{\mu}=\mathbf{X}\ \boldsymbol{\beta}$ where $\mathbf{X}$ is an $N \times (k+1)$ matrix and $\boldsymbol{\beta}$ is a $(k+1)$-dimensional vector of regression parameters, (3) spatial models with similar neighbors which is defined by the similarity function formed by the auxiliary data. We don't get much improvement in the estimated cross-validated RMSEs by adding the auxiliary information to our models over the simple spatial models we used in Chapter 4; for M data, 62.19 vs. 62.61; for Lt data, 240.19 vs. 242.08; for Tba data, it is even worse, 21.92 vs. 19.65. In Chapter 8 we used the predictor based on the models introduced in this chapter to make predictions for 6 selected plots on the 0.85 mile grid. The prediction interval at level α is constructed for individual plot prediction at a plot on the 0.85 mile grid (or an even more intensive grid). There is no closed form for the prediction error for prediction using the MSN procedure, so several simulations designed to yield realizations similar to Tba, Lt, and M data were conducted. The $\widehat{r.m.s.e.}(\hat{Y}_p^i)$'s for MSN inference are much larger than the general linear model, Model 7.1, and the spatial regression model, Model 7.6. That is because the MSN procedure constrains the predictions by maximum and minimum data values; and Models 7.1 and 7.6 don't.

The reason we use ancillary data such as aspect, elevation, slope, and TM bands is because such data is available for non-sampled plots and non-sampled subplots and for prediction purpose we need to use the information available for non-sampled plots. With limitation of auxiliary information to non-sampled plots, the topographic data such as aspect, elevation, and slope, doesn't contain

available information to make predictions for individual response at non-sampled plots. The TM band data only helps a little bit.

Chapter 8 compared the predictions based general linear models(GLM), spatial regression models (SRM), and the most similar neighbor (MSN) procedure. Distance-related correlation functions and auxiliary information were included in spatial regression model. Such models did not reduce estimated overall RMSEs, but did reduce m.s.p.e. of the prediction errors for individual plots. $\hat{Y}_{s_0,srm}$ is the best predictor when there exists significant spatial dependence among the data based on the smallest mean-squared prediction error (m.s.p.e.). Hence an appropriate model selection criterion is needed for the purpose of individual predictions based on spatial models. Compared to the predictors in Chapter 5, the plug-in m.s.p.e.s of $\hat{Y}_{s_0,srm}$ are smaller than those of the predictors in Chapter 5 (338 vs. 587).

9.2 Conclusions

(1) for plot based data the predictions described in Chapter 5 do not work well even if the model is correctly specified. The predictions based on plot data don't work well because adjacent plots contain little or no information (i.e. there is essentially no spatial dependence at the plot scale of the data). Sufficient spatial dependence at the subplot level was found to be useful in predictions.

(2) Compared to the results from Chapter 5, we note that for Tba (which has a low percentage of zero outcomes) the estimated MSE based on the best predictor, 400.15 is much smaller than the smallest one in Chapter 5, 578.65. The estimated RMSEs of Lt_s, M_s, and Tba_s using the prediction rules from SZI models are not much improved over just using the sample mean. Since there are a high percentage of zeros in seedlings data, the sample sizes we

had were too small to obtain decent estimates of the parameters in the SZI models. The directional mean schemes we considered for spatial zero-inflated models are constant mean based 1.7-mile grid data. They only can predict the non-sampled plots or missing plots on the 1.7-mile grid, and not allow for the predictions at non-sampled plots on the 0.85-mile grid.

(3) In this thesis, we derived the prediction errors for back-transformed prediction and also conducted several simulations to see the reliability of prediction errors. For a process with large variance, the prediction errors for "plain" back-transformed and "unbiased" back-transformed work well, but the one for "approximately" unbiased back-transformed predictions are way off from the prediction errors we got from simulations.

(4) The predictions using the models adding the auxiliary information into the models discussed in Chapter 4 didn't improve the RMSE much. The predictions via spatial regression models with distance-related correlation did reduce mean-squared prediction errors for individual plots. Auxiliary information can useful if the model is specified correctly, e.g. considering the combination of the neighboring observation related to distance-related correlation functions.

Main contributions of this study are: (1) we derived the predictions errors for predictions based on transformed and non-transformed data for non-sampled locations using only field sample plot or subplot information, (2) derived the prediction errors for predictions based on spatial multivariate regression models with distance-related correlation functions, (3) used spatial zero-inflated models to handle the numerous zeros in the data, (4) compared the normal approximate and bootstrap-t prediction intervals for the predictors with coefficients based on

distance-related correlation functions with and without auxiliary information, and (5) provided reliable bootstrap methods for different situations.

After exploring many models, we found that for prediction purposes, a simple model that assumes a a spatial correlation structure without any distributional assumptions worked as well as any other. The specific spatial correlation structure was selected by minimizing an overall mean squared error. Our recommended predictor is a linear combination of measurement at neighboring sites. The coefficients in the linear combinations are functions of the estimated neighboring correlations. These simple predictors have an analytical formula for the prediction errors and these too can be estimated using the estimated correlations. These predictors work for any of the response variables Tba, Lt, and M.

We did find that certain auxiliary data was useful in reducing prediction errors for Tba. To give some idea as to the size, of the errors involved for this variable we illustrate with the results of our study of predicting the value at each of six sites on the 0.85-mile grid. One would be predicting a value over a range of about twenty to about sixty. If the sample mean was used as a predictor and independence among sites assumed, the prediction error formula is $[\hat{\sigma}^2(1 + \frac{1}{N})]$, which is estimated to be 24.38. The estimated prediction errors using our simple predictors at the six sites are, respectively, 24.23, 24.22, 24.21, 24.24, 24.11, and 24.12, which indicate little improvement over the nominal 24.38. On the other hand, the corresponding estimated prediction errors for the same six sites, when one incorporates aspect and band 4 (representing active vegetation) as auxiliary variables, are, respectively, 18.54, 18.64, 18.53, 18.47, 18.59 and 18.44, which represents a nearly twenty-five percent improvement.

For Lt one would predict a value between above three hundred and five hundred. The nominal $[\hat{\sigma}^2(1 + \frac{1}{N})] = 269.76$. Our respective estimated prediction errors at the six sites are 264.81, 264.56, 264.52, 264.70, 264.53 and 264.55 with-

out auxiliary data, and 258.52, 258.14, 258.62, 257.74, 259.87 and 258.25 with auxiliary data. The overall reduction in prediction errors is less than five percent.

For M one would predict a value between about forty and a hundred and thirty. The nominal estimated prediction errors range from 52.71 to 52.76 without auxiliary data. The estimated prediction errors using auxiliary data ranged from 53.19 to 54.08., so for M, auxiliary data provides no improvement in prediction.

9.3 Future Research

This dissertation has revealed many more questions that could direct future work. Reliable model selection criterion is a fertile area for spatial models. Bayesian inference is widely used in spatial models to capture the uncertainty of the spatial dependence parameters and could be further investigated.

9.3.1 Spatial model cross-validation

McQuarrie and Tsai (1998, p.255-256) introduced univariate autoregressive cross-validation. They pointed out a key assumption of standard cross-validation is that the observation withheld is independent of the remaining $(N-1)$ observations. This assumption will fail for an AR(p) model. Hence, they proposed a univariate autoregressive cross-validation. In the AR(p), they assume that there exists a constant l such that y_i and y_j are approximately independent for $\mid i-j \mid > l$. The univariate autoregressive cross-validation could withhold $\pm l$ additional observations around the observation withheld, i.e. when withholding y_t we should withhold the block $y_{t-1}, \ldots, y_t, \ldots, y_{t+1}$.

We could explore spatial model cross-validation by the following procedure; that is withholding the observation y_i, plus additional observations falling within neighborhood $N_h(i)$, where $N_h(i)$ is the set of the distance for all observations y_j smaller than the neighborhood distance $\boldsymbol{h}$. Each time, we only using the

remaining $N- \mid N_h(i) \mid -1$ observations to estimate the parameters in the model. Then predict the withheld observation by the prediction rule. The procedure is repeated until we obtain the predictions for all N observation values.

9.3.2 Spatial zero-inflated (SZI) regression models

As mentioned in Chapter 6, a better parameter estimation method is needed for spatial zero-inflated (ZI) models. It should be worth while to examine maximum likelihood estimation for spatial ZI models. We also pointed out that the prediction rules described in Chapter 6 for spatial zero-inflated (SZI) models don't allow for predictions at plots on the 0.85 mile grid or on a more intensive grid. And hence, an alternative approach for prediction using SZI models is needed; possibly assuming a distance-related mean function.

Another issue about SZI models arose in Chapter 7 and that is to bring the auxiliary information into the parameter structure of SZI models, like Lambert (1992) did for zero-inflated models. A spatial zero-inflated regression model is more complicated than a zero-inflated regression model due to spatial dependence. But it is a fertile area in economical and medical problems since the phenomenon of zero-inflation happens for various reasons in those fields. Winkelmann (2000) discussed some other kind of inflated models, including one-inflated, and zero-and-one inflated models. Spatial one-inflated models and spatial zero-and-one inflated models are candidates for future work.

9.3.3 Spatial models with location perturbations

Initially we had hoped to study spatial models with location perturbations, but this requires data not currently available on location errors. In the following subsection, the concept of perturbation models and tentative results are discussed.

The experimenter hoped to observe $\{Y_{(s_1,s_2)}\}$ intending to make inferences regarding some model for the process $\{Y_{(s_1,s_2)}\}$, but observes the process $\{X_{S(s_1,s_2)}\}$

at location $\{S(s_1, s_2)\}$ instead of (s_1, s_2). A simple location perturbation model could be the following: define $\{S(s_1, s_2)\}$ by

$$\{S(s_1, s_2)\} = \begin{cases} (s_1, s_2) & \text{w.p. } p^2 \\ (s_1 \pm 1, s_2) & \text{w.p. } p(1-p)/2 \\ (s_1, s_2 \pm 1) & \text{w.p. } p(1-p)/2 \\ (s_1 \pm 1, s_2 \pm 1) & \text{w.p. } (\frac{1-p}{2})^2 \end{cases}$$

where $0 < p < 1$. For example, assuming that p is $1/2$, $X_{(1,1)}$ is $Y_{(1,1)}$ with probability $1/4$, $X_{(1,1)}$=$Y_{(0,1)}$ or $Y_{(2,1)}$ with probability $1/8$ (each horizontal directional shift), and $X_{(1,1)}$=$Y_{(1,0)}$ or $Y_{(1,2)}$ with probability $1/8$ (each vertical directional shift), and $X_{(1,1)}$=$Y_{(0,0)}$, $Y_{(0,2)}$, $Y_{(2,0)}$, or $Y_{(2,2)}$ with probability $1/16$ (each four corner directional shifts). Then the observed $X_{(1,1)}$ is the desired $Y_{(1,1)}$ only 25 % of the time. The question is,what loss in efficiency in estimating the parameters of the $\{Y_{(s_1,s_2)}\}$ results from observing the estimated $\{X_{(s_1,s_2)}\}$ process?

Possible delineations of perturbation models for $\{S(s_1, s_2)\}$ are

(i) $S(s_1, s_2) = (s_1, s_2) +$ "noise" where "noise" is referred to as $\mathbf{Z} = (Z_1(\mathbf{s}), Z_2(\mathbf{s}))$, $\mathbf{s} = (s_1, s_2)$.

(ii) construct $\{S(s_1, s_2)\}$ as a random field.

Possible constraining properties that $\{S(s_1, s_2)\}$ might possess are:

1. $\mathcal{E}\,[S(s_1, s_2)] = (s_1, s_2) = \mathbf{s}$,

2. stationary increments; that is, the distribution of $S(\mathbf{t} + \mathbf{h}) - S(\mathbf{t})$ depends on $\mathbf{h}$.

9.3.4 Bayesian framework for Phase 1 model

This section outlines some Bayesian type ideas compatible with models we used. In the Bayesian approach, the populations parameters $\boldsymbol{\Theta}$ are assumed to have a probability distribution and the information from the observed data is

used to modify a prior distribution for the parameters into a posterior probability distribution by Bayes Theorem.

$$p(\boldsymbol{\Theta} \mid \boldsymbol{y}) = \frac{p(\boldsymbol{y} \mid \boldsymbol{\Theta})\pi(\boldsymbol{\Theta})}{p(\boldsymbol{y})}, \qquad \text{if } p(\boldsymbol{y}) \neq 0.$$

The Phase I model was defined as

$$\mathbf{Y} = \boldsymbol{\mu} + B\boldsymbol{\varepsilon}, \qquad \text{(See (4.1))}$$

where $\mathbf{Y} = (Y_1, \ldots, Y_N)'$, $\boldsymbol{\mu}=(\mu_1, \ldots, \mu_N)'$, $\boldsymbol{\varepsilon} \sim \text{MVN}(\mathbf{0}, \sigma^2\ \mathbf{I})$. Therefore, $\mathbf{Y} \sim \text{MVN}(\boldsymbol{\mu}, \sigma^2\ \boldsymbol{\Gamma})$. Assume that $\boldsymbol{y}$ is the observation vector of $\mathbf{Y}$. Thus the likelihood function of $\boldsymbol{y}$ is

$$L(\boldsymbol{y} \mid \mu, \sigma^2, \rho_i; i = 1, \ldots, 4) = (2\pi\sigma^2)^{-N/2} \mid \boldsymbol{\Gamma} \mid^{-1/2} exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{y} - \mu\mathbf{1})'\boldsymbol{\Gamma}^{-1}(\boldsymbol{y} - \mu\mathbf{1})\right]$$

where $\boldsymbol{\Gamma}= \mathbf{I} + \rho_1\, \mathbf{H} + \rho_2\, \mathbf{V} + \rho_3\, \mathbf{C}^{(1)} + \rho_4\, \mathbf{C}^{(2)}$.

Application of Bayesian methods require the postulation of prior distributions discussed next.

9.3.4.1 Prior distributions

Prior distributions could be derived from earlier empirical work or from the opinions of subject-area experts. Such prior specification is called informative prior. In many studies, we wish to endow the prior distribution with little informative content. This can be achieved by the employment of a "diffuse' or "noninformative" prior distribution for $\boldsymbol{\Theta}$. We could try several classes of priors including informative and noninformative prior distribution.

In determining a class of priors, we could assume that spatial dependence parameters $\boldsymbol{\rho}=(\rho_1, \ldots, \rho_k)$ are independent of the parameter, μ, and the residual variance, σ^2, so that the prior densities satisfy

$$\pi(\mu, \sigma^2, \rho_1, \ldots, \rho_k) = \pi(\mu \mid \sigma^2)\pi(\sigma^2)\pi(\rho_1, \ldots, \rho_k)$$

i.e. (μ, σ) is independent of $\boldsymbol{\rho}=(\rho_1, \ldots, \rho_k)$.

For Lt data, one could deal with the model we selected in the chapter 4, i.e. k=2, including informative and noninformative prior distributions. One could deal with the following three sets of priors: (1) $\pi(\mu \mid \sigma^2)$, $\pi(log(\sigma))$, and $\pi(\rho_1, \rho_2)$ are independently and uniform distributed, (2) $\pi(\mu \mid \sigma^2) \sim N(\mu_0, \sigma^2)$, $\pi(\sigma^2) \sim IG(a, b)$, and $\pi(\rho_1, \rho_2) \sim uniform(-1, 1)$, and (3) $\pi(\mu \mid \sigma^2) \sim lognorm(\mu_0, \sigma^2)$. Once prior distributions are given, we can derive posterior distributions.

9.3.4.2 Posterior distributions

For the Phase I model, the posterior distribution of $\boldsymbol{\Theta}$ given the data $\boldsymbol{y}$ is

$$\begin{aligned} p(\boldsymbol{\Theta} \mid \boldsymbol{y}) &= p(\mu, \sigma, \rho_1, \rho_2 \mid \boldsymbol{y}) \\ &\propto p(\boldsymbol{y} \mid \mu, \sigma, \rho_1, \rho_2)\pi(\mu \mid \sigma^2)\pi(\sigma^2)\pi(\rho_1, \rho_2) \end{aligned} \tag{9.1}$$

The Bayesian predictive distribution for y_j at non-sampled plot location, $p(y_j \mid \boldsymbol{y})$, is given by

$$\begin{aligned} p(y_j \mid \boldsymbol{y}) &= \int p(y_j \mid \boldsymbol{\Theta})p(\boldsymbol{\Theta} \mid \boldsymbol{y})d\boldsymbol{\Theta}, \\ &\propto \int p(y_j \mid \mu, \sigma, \rho_1, \rho_2)\pi(\mu \mid \sigma^2)\pi(\sigma^2)\pi(\rho_1, \rho_2). \end{aligned} \tag{9.2}$$

where $\boldsymbol{\Gamma}= \mathbf{I}+ \rho_1 \mathbf{H}+\rho_2 \mathbf{V}$. The posterior distribution is analytically intractable due to the parametrization of $\boldsymbol{\Gamma}$ and the normalizing constant in the prior distribution.

Since the posterior distribution is analytically intractable, we could not easily obtain maximum posterior estimates of these parameters. We could however approximate the predictive distribution using Gibbs sampling. The Gibbs sampling estimation procedure uses the full conditional distributions in an iterative procedure to obtain estimates of the parameters for the intractable posterior distribution.

9.3.5 Spatial-temporal models

Forest survey data is often collected by plot locations periodically. For the Siuslaw Forest, the second occasion grid survey is done by project year and remeasurement panel. A remeasurement panel consists of approximately 25% of the first occasion sample units (the data we used in this thesis). Originally, a different panel was to be remeasured every 2 years on each Forest resulting in the entire Forest being totally remeasured over an 8 year period. Beginning with the 1999 projects, however, the remeasurement cycle was adjusted to a panel every 3 years with the total coverage completed every 12 years. Since our response variables such as Total basal area, mortality, number of live tress, do change with time, it may be worthwhile to examine spatial-temporal models for data over time.

In summary, the follow up research we recommend is to develop a spatial model-validation, to study spatial zero-inflated models with distance based mean function, spatial zero-inflated regression models, spatial models with location perturbations, and Bayesian inference for the models we developed in the current study.

References

1. Akaike, H. (1978). A Bayesian analysis of the minumum AIC procedure. Annals of the Institute of Statistical Mathematics 30, Part A, 9-14.

2. Brimicombe, Allan (1998). A fuzzy coordinate system for locational uncertainty in space and time. Innovations in GIS 5: Selected Papers from the Fourth National Conference on GIS Research UK (GISRUK), Taylor & Francis Ltd., pp.141-150.

3. Carlin, B.P. and Louis, T.A. (1998). Bayes and empirical Bayes methods for data analysis, 1st ed., Chapman & Hall.

4. Cohen, A. C. (1960). On a class of pseudo-contagious distributions, Univeristy of Georgia, Institute of Statistics, Technical Report No. 11.

5. Cook, William H. (2000). Forest/Non-forest stratification in georgia with Landsat Thematic Mapper data. In: McRoberts, R. E. Reams, G. A. and Van Deusen, P. C. (Eds.), Proceedings of the First Annual Forest Inventory and Analysis Symposium. Nov. 2-3, 1999 San Antonio, TX. USDA For. Serv. Gen. Tech. Rep. GTR NC-213. pp. 28-30.

6. Crist, E.P., and Cicone, R.C. (1984). A plysically-based transformation of Thematic Mapper data – the TM tasseled cap. IEEE Transactions on Geoscience and Remote Sensing. Vol. GE-22, No. 3, pp. 256-263.

7. Dalenius, T. (1987). Discussion. In Platek et al. cited below, pp.257-263.

8. Donahue, Rafe M.J. (1992). Dissertation. Estimation for nearest 1-neighbor processes. Colorado State University, Dept. of Statistics.

9. Fong, Daniel Y.T. (1995). A note on information loss in analyzing a mixture model of count data. Commun. Statist. - Theory Meth., 24(12), 3197-3209.

10. Fong, Daniel Y.T. and Yip, Paul (1993). An EM algorithm for a mixture model of count data. Statistcs & Probability Letters 17, 53-60.

11. Franco-Lopez, H. (1999). Dissertation. Updating forest monitoring systems estimates. The University of Minnesota.

12. Garcia-Soidan, P.H. and Hall, P. (1997). On sample reuse methods for spatial data. Biometrics 53, 273-281.

13. Gilks, W. R., Richardson, S., Spiegelhalter D. J. (1996). Markov Chain Monte Carlo in Practice. Chapman & Hall: London.

14. Ghosh, M. and Rao, J.N.K. (1994). Small area estimation: an appraisal. Stat. Science Vol 9, No 1: 55-93.

15. Graybill, Franklin A. (1984). Theory and Application of the Linear Model. 2nd printing. Duxbury Press.

16. Hall, P. (1985). Resampling a coverage pattern. Stochastic Processes and their Applications 20, 231-246.

17. Hall, P. and Jing, B. (1996). On sample reuse methods for dependent data. J. R. Statist. Soc. B 58, No 4, pp. 727-737.

18. Haining, Robert (1990). Spatial data analysis in the social and environmental science. Cambridge University Press, New York.

19. Isaaks, A.G. and Srivastava, R.M. (1989). An Introduction to Applied Geostatistics. Oxford University Press, New York.

20. Johnson, Norman L, and Kotz Samuel (1969) Distributions in Statistics: Discrete Distributions, Boston: Houghton Mifflin.

21. Johnson, Norman L, Kotz Samuel, and Balakrishnan, N.. (1997). Discrete Multivariate Distributions, New York: John Wiley & Sons, Inc.

22. Johnson, Norman L, Kotz Samuel, and Balakrishnan, N.. (1998). Continuous Univariate Distributions, Vol. 1, 2nd ed., New York: John Wiley & Sons, Inc.

23. Kalton, G. (1987). Discussion. In Platek et al. cited below, pp 264-266.

24. Kangas, A. (1996). Small-area estimates using model-based methods. Can. J. For. Res. 26: 758-766.

25. Lahiri, Kaiser, Cressie, and Hsu (1999). Prediction of Spatial Cumulative Distribution Functions Using subsampling.(with discussion) J. Amer. Stat. Assoc. Vol. 94, No. 445, 86-107; Rejoinder, 107-110.

26. Lambert, Diane (1992). Zero-Inflated Poisson Regression, With an Application to Defects in Manufacturing. Technometrics, Vol. 34, 1-14.

27. Lillesand, Thomas M. and Kiefer Ralph W. (2000). Remote sensing and image interpretation, 4th ed. John Wiley & Sons, Inc.

28. Lin, Jin-Mann S. (1996). Thesis. Time series with perturbed time index. Colorado State University, Dept. of Statistics.

29. Malec, J., Sedransk, J., Moriarity, C.L. and LeClere, F.B. (1997). Small area inference for binary variables in the National Health Interview Survey. J. Amer. Stat. Assoc. Vol. 92, No. 439, 815-835

30. McLachlan, Geoffrey and Peel, David. (2000). Finite Mixture Models. John Wiley & Sons, Inc.

31. McQuarrie, Allan D.R. and Tsai, Chih-Ling (1998). Regression and Time Series Model Selection. World Scientific Publishing Co. Pte. Llt.

32. Moeur, M. and Stage, A.R. (1995). Most similar neighbor: an improved sampling inference procedure for natural resource planning. Forest Science 41: 337-359.

33. Moeur, M. and Hershey, R.R. (1998). Preserving spatial and attribute correlation in the interpolation of forest inventory data. 3rd International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, May 20-22, 1998, Ouebec City, Quebec, Canada. Ann Arbor Press.

34. Moisen, G.G. and Edwards, T.C. Jr. (1999). Use of generalized linear models and digital data in a Forest Inventory of Northern Utah. J. Agric,Biol. and Env. Stat. 4: 372-390.

35. Pfeffermann, D. (1993). The role of sampling weights when modeling survey data. Intern. Stat. Rev. 61: 317-337. HA11 I505 not useful at this time.

36. Pandey, K. N. (1965). On generalized inflated Poisson distribution, Journal of Scientific Research, Benares Hindu University, **15**, 157-162.

37. Platek, R; Rao, J.N.K, Sarndal, C. E. and Singh, M. (eds) (1987). Small area statistics: an international symposium. J.Wiley and Sons, NY.

38. Politis, Dimitris N., Romano, Joseph J., and Wolf, Michael. (1999). Subsampling, New York: Springer.

39. Ripley, B.D. (1996). Pattern recognition and neural networks, New York: Cambridge University Press.

40. Roberts, D. W., and Cooper, S. V. (1989). "Concepts and Techniques of Vegetation Mapping." in Land Classifications Based on Vegetation: Applications for Resource Management, eds. D. Ferguson, P. Morgan, and F. D. Johnson, General Technical Report INT-257, U.S. Department of Agriculture, Forest Service, Intermountain Research Station, Ogden, UT, pp. 90-96.

41. Schaible, W.L. (1992). Use of small area statistics in US Federal Federal Programs. In: Small Area Statistics and Survey Designs. G.Kalton, J.Kordos and R.Platek, eds.) 1:95-114. Central Stat. Office, Warsaw.

42. Schreuder, H.T., Gregoire, T.G. and Wood, G.B. (1993). Sampling methods for multiresources forest inventory. J.Wiley and Sons, N.Y. pp. 446.

43. Schwarz, G. (1978). Estimating the dimension of a model. Annals of Statistics 6, 461-464.

44. Shimizu, K. and Iwase, K. (1987). Unbiased estimation of the autocovariance function in a stationary generalized lognormal process. Communications in Statistics. Theory and Methods, **16**, 2145-2154.

45. Singh, S. N. (1963). A note on inflated Poisson distribution, Journal of the Indian Statistical Association, **1**, 140-144.

46. Smith, A.F.M. and Gelfand, A.E. (1992). Bayesian Statistics Without Tears: A Sampling-Resampling Perspective. The American Statistician, Vol. 46, No. 2.

47. Tomppo, E., Goulding, C. and Katila, M. (1999). Adapting Finnish Multi-Source Forest Inventory Techniques to the New Zealand Preharvesting Inventory. Scand. J. For. Res. 14: 182-192.

48. Yip, Paul (1988). Inference about the mean of a Poisson distribution in the presenece of a nuisance parameter. Austral. J. Statist., 30(3), 299-306.

49. Yip, Paul (1991). Conditional inference on a mixture model for the analysis of count data. Commun. Statist. - Theory Meth., 20(7), 2045-2057.