INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA

UMI®
800-521-0600

NOTE TO USERS

This reproduction is the best copy available.

UMI

DISSERTATION

THE AUTOLOGISTIC MODEL WITH COVARIATES FOR SAMPLE DATA AND ROBUST SAMPLING DESIGNS USING PREDICTED PROBABILITY OF PRESENCE

Submitted by

Molly Leecaster

Department of Statistics

In partial fulfillment of the requirements

for the degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 1999

COLORADO STATE UNIVERSITY

October 6, 1999

WE HEREBY RECOMMEND THAT THE DISSERTATION THE AUTOLOGISTIC MODEL WITH COVARIATES FOR SAMPLE DATA AND ROBUST SAMPLING DESIGNS USING PREDICTED PROBABILITY OF PRESENCE PREPARED UNDER OUR SUPERVISION BY MOLLY LEECASTER BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

Committee Member

Committee Member

Committee Member

Co-adviser

Adviser

Department Head

ABSTRACT

THE AUTOLOGISTIC MODEL WITH COVARIATES FOR SAMPLE DATA AND ROBUST SAMPLING DESIGNS USING PREDICTED PROBABILITY OF PRESENCE

We extend the autologistic model for spatially correlated binary response lattice data by including covariates related to presence and using sample data from a subset of sites instead of complete data. We assume the binary response is presence or absence of a species. The model output is the predicted probability of presence for sites over the lattice. We present a Bayesian framework and develop a Gibbs sampling estimation procedure. We demonstrate using three examples that the autologistic model with covariates for sample data reproduces the truth more accurately than either the autologistic model, which uses only spatial relation information, or the logistic regression model, which uses only covariate information.

We further extend our model to account for possibly imperfect detection in the sample observations. We assume there are covariates which are related to detectability of the species and modify the likelihood function to a logistic regression form. We expand the Bayesian set-up and Gibbs sampling estimation procedure to include the modified likelihood function. With two examples we demonstrate that also using sighting-related covariates achieves good reproduction of the truth.

We investigate the use of the predicted probability in development of robust sampling designs whose goal is maximization of detection. The model output can

be used to determine inclusion probabilities for unequal probability samples, define homogeneous partitions for stratified designs and be included in a ratio estimator of number of occupied sites. We present and compare four unequal probability sample design estimators, two strata construction methods, two within strata sampling methods, seven allocation strategies, and a fixed top stratum design. These comparisons are made with respect to the expected number of observed presence sites and the variance of the estimator of occupied sites. The fixed top stratum design, which samples all sites in the strata with the highest predicted probability, achieves the greatest detection and also attains a small variance for the estimator. This work serves to advance statistical methods used to map and detect rare species. These methods are applicable as well to mapping, or image analysis, and sample designs for lattice data.

Molly Leecaster
Department of Statistics
Colorado State University
Fort Collins, Colorado 80523
Fall 1999

ACKNOWLEDGEMENTS

I would like to thank my advisers, Jennifer Hoeting and David Bowden, for their support, advice, and guidance over the last four years. I would like to thank Raymond Czaplewski for introducing the project and supporting our work along the way. This work was funded by USDA-USFS contract number PNW 95-0766. I would also like to thank Matt Calder for the development of and assistance with his S-plus compiler.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION

The United States Forest Service (USFS) manages public land that is both home to many species of plants and animals and ripe for logging. Management decisions regarding this land depend, in part, on the abundance of the native species of plants and animals. For many species the abundance and range is not known. The first step in management of public lands is to determine which species need special treatment, and therefore their habitat special management.

The USFS has a goal to develop methodology which provides improved information on species presence. We took on this goal as the project for our current work. Obtaining improved information on species presence involves sampling. The spatial nature of environmental data leads us to use a mapping tool to develop the best sampling program. So our project became two-fold. First we develop methodology which will utilize all of the available information about a species of interest in order to produce a map of likelihood of presence of the species. Second, we investigate sampling designs which are based on these maps, with the goal of detection of the species of interest.

Development of the autologistic model with covariates for sample data achieves the first goal of development of a map. We extend the autologistic model, which is a spatial model for binary data, to improve prediction. The autologistic model with covariates for sample data utilizes observed data collected from a sample of

sites over the area of interest, habitat information measured over the entire area of interest, a measure of search intensity, and the assumption that the species form clusters to provide predicted likelihoods of species presence.

The second goal, investigation of sampling designs to increase detection of the species, is attained by rigorous investigation of designs which utilize the map of predicted probability of presence produced through the first goal. The information from the map can be used in at least three ways to improve detection of species. The first approach we consider is to use these predicted probabilities of presence in selecting sites to sample in an unequal probability sampling strategy. The second way we use the map information is to stratify the population based on the predicted probabilities. The population will be stratified into segments which are homogeneous with respect to predicted probabilities. A final use we make of the predicted probabilities is through the ratio estimator of number of sites with species presence. This estimator may be especially useful when using a sampling design which does not incorporate these predicted probabilities in its design, such as simple random sampling.

The work we present has been motivated by the USFS project, but is applicable to a variety of fields. The autologistic model with covariates for sample data which we develop is applicable to the many other applications which this type of model has been used such as agricultural research (Gumpertz *et al.*, 1997), forestry (Preisler, 1993), archaeology (Besag *et al.*, 1991), and biological range mapping (Heikkinen and Högmander, 1994; Högmander and Møller, 1995; Augustin *et al.*, 1996).

In Chapter 2 we introduce the basic autologistic model and some extensions. This model is the basis of our autologistic model with covariates for sample data.

In Chapter 3 we derive the extensions to the basic model. We introduce the information we use in the extended model and provide a Bayesian framework. Within this framework we propose methodology for parameter estimation. In Chapter 4 we demonstrate the utility of the extended model through three simulated examples. The examples serve to compare utility of the model with respect to different environmental factors as well as to compare our model to two existing, more limiting models, logistic regression and the basic autologistic model. Our model initially uses a simplistic function for species detection. We extend this in Chapter 5 where we propose a more comprehensive approach to model search intensity over the area of interest. Chapter 6 is devoted to the investigation of sampling designs and estimators of detection. We compare different ways to use the map information in sampling designs by comparing estimates of detection of presence sites and variance for various properties of the population in simulated setups and through the examples formulated in Chapter 4. We also compare the utility of the three models, the autologistic model with covariates for sample data, logistic regression, and the autologistic model with respect to sampling designs. In Chapter 7 we summarize our conclusions and propose areas for future work.

Chapter 2

AUTOLOGISTIC MODEL

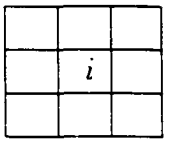
The basic autologistic model is a spatial model for binary response. The model has the same form as a logistic regression model, but the model covariate is a function of the 'neighboring' responses. The spatial dependence is defined by a locally dependent Markov random field. In image analysis applications the binary response for each pixel is black or white. In our application the response at each site (or pixel) is presence or absence of the species of interest. The input to the basic autologistic model is a response at every site which is observed with error. The likelihood function of the data is the functional form of this observation mechanism. The output of the basic autologistic model is the predicted probability of positive response for each site, i.e. black, or presence. In this chapter we will introduce notation, assumptions, the basic autologistic model and extensions.

2.1 Notation

The area of interest, for example an image or a particular forest, is divided into N sites on a lattice structure. We assume for our examples a square grid structure of appropriate scale over the area of interest. Each of the N sites on the lattice will be the square area inside the gridlines and the sites will be contiguous.

For each site there is a true state of the binary response. We denote this unknown truth for the i^{th} site by $x_i, i = 1, \ldots, N$. Black, or presence, at site i is

Figure 2.1: A Second Order Neighborhood.



indicated by $x_i = 1$, and white, or absence, at site i is indicated by $x_i = 0$. The truth over the whole lattice is denoted by a vector $\underline{x}$ of length N.

There is also an observed response at each site which is denoted for site i by y_i. An observed value of $y_i = 1$ indicates observed black at site i or observed presence of the species at site i. Observed absence, or white, at site i is denoted by $y_i = 0$. The observed values for the whole lattice are denoted by a vector $\underline{y}$ of length N.

Spatial dependence between sites is assumed for a 'neighborhood' of each site. The neighborhood of site i, δ_i, can be defined in a variety of ways. A first order neighborhood of site i includes the four sites north, south, east, and west of site i. A second order neighborhood would include the first order neighbors plus the four corner sites, as shown in Figure 2.1. Any defined neighborhood must meet the neighborhood condition: if site i is a neighbor of site j, then site j must be a neighbor of site i,

$$i \in \delta_j \iff j \in \delta_i.$$

The spatial correlation between sites is assumed to be limited to the defined neighborhood. We summarize the information in the neighborhood of site i using a spatial covariate, s_i. This value can be defined in a variety of ways. For example s_i could be the number of neighbors with 'black', or 'presence', status. We will introduce versions of the spatial covariate, s, as they appear in the basic model, an extension, and in our autologistic model with covariates for sample data.

In summary, the notation for the autologistic model is defined:

$$
\begin{aligned}
N &= \text{number of sites on the lattice over the area of interest} \\
y_i &= \begin{cases} 1 & \text{observed black, or species presence, at site } i \\ 0 & \text{observed white, or species absence, at site } i \end{cases} \\
x_i &= \begin{cases} 1 & \text{true black, or species presence, at site } i \\ 0 & \text{true white, or species absence, at site } i \end{cases} \\
\underline{x}_{-i} &= \text{true status for all sites except site } i \\
\delta_i &= \text{neighborhood of site } i \\
\underline{x}_{\delta_i} &= \text{true status for neighbors of site } i \\
s_i &= \text{spatial covariate for site } i \text{ some function of } \underline{x}_{\delta_i}.
\end{aligned}
$$

In what follows probability density functions will be represented by p, probabilities will be denoted by Pr, likelihood functions by f, and prior distributions by π.

2.2 Assumptions

Three assumptions underlie the autologistic model (Besag, 1972). Cressie (1993) summarizes the assumptions and development for the autologistic as well as other spatial auto-models such as the auto-Poisson.

The first assumption is that the likelihood function of the observations is conditionally independent given the truth at each site.

$$f\left(\underline{y} \mid \underline{x}\right) = \prod_{i=1}^{N} p(y_i \mid x_i).$$

The probability of observing a black or white pixel at pixel i depends only on the true color at pixel i, not colors of the other pixels. In our application, the probability that we observe a species at a site depends only on whether the species is actually present at the site, independent of presence/absence in other sites. This

product form for the likelihood function allows for mathematical convenience, but also provides a valid foundation for independence of site by site detectability in our application. Using this form for the likelihood function we can specify detectability, the probability of observing a species on a sampled site given that it truly exists there (see Chapter 5).

The second assumption defines a locally dependent Markov random field. It is assumed that the distribution of *true* black, or presence, x_i, at each site given all other sites, $\underline{x}_{-i}$, is the probability of x_i given the truth at neighboring sites, $\underline{x}_{s_i}$.

$$Pr(x_i \mid \underline{x}_{-i}) \equiv Pr_i\left(x_i \mid \underline{x}_{s_i}\right).$$

This limits the spatial dependence to the neighborhood of the site. Sites outside of the neighborhood of site i are assumed to be independent of x_i given $\underline{x}_{s_i}$.

The final assumption is the positivity condition. If $\zeta = \{\underline{y} : Pr\left(\underline{y}\right) > 0\}$ and $\zeta_i = \{y_i : Pr(y_i) > 0\}$ for $i = 1, \ldots, N$, then the positivity condition is satisfied if $\zeta = \zeta_1 \otimes \cdots \otimes \zeta_N$, where $\zeta = \zeta_1 \otimes \cdots \otimes \zeta_N$ is the Cartesian product of all sets $\zeta_1, \ldots, \zeta_N$. The positivity condition requires that any configuration of presences and absences over those sites is possible. There are no restrictions on possible images as would be the case for contagious disease where a site with presence cannot occur without presence in a neighboring site.

2.3 Basic Autologistic Model

The basic autologistic model was formulated by Besag (1972) for images which were measured with error. We introduce the data, spatial dependence assumption and characterization of this model.

The observations are recorded values of black or white at each site. The reliability of the observations is not known, except that there may be error. The

autologistic model estimates the probability of presence for each pixel using the presence/absence of the neighboring pixels. The basic autologistic model updates the image to an image closer to the truth using observed values and the assumption of neighborhood dependence. In other words, the model is used to 'clean up' images that are 'dirty'.

The spatial covariate, s_i, establishes the extent of spatial dependence assumed when applying the autologistic model to images. The neighborhood must be specified in accordance with the neighborhood condition defined above. A first order neighborhood is often a reasonable assumption. For image analysis, a common spatial covariate is defined as the number of black pixels in the neighborhood.

$$s_i = \sum_{j \in \delta_i} x_j. \tag{2.1}$$

so $s_i = \{0, \ldots, \#\text{ of neighbors}\}$. Larger values of s_i correspond to "blacker" neighborhoods.

Using this spatial covariate, the resulting autologistic model is formulated as,

$$Pr\left(x_i = 1 \mid \underline{x}_{\delta_i}, \beta_1\right) = \frac{\exp(s_i \beta_1)}{1 + \exp(s_i \beta_1)}. \tag{2.2}$$

where β_1 is the spatial coefficient. Both of these coefficients need to be specified or estimated. The model has the form of a logistic regression model, but the covariate is a function of neighboring values of the response, hence the name *auto*logistic.

The joint distribution for $\underline{x}$ over the whole lattice has the form

$$p(\underline{x} \mid \beta) \propto \exp(\beta \underline{s}), \tag{2.3}$$

where $\underline{s}$ is defined in Equation 2.1. There is a denominator which is the sum over all possible combinations of black and white pixels and is the normalizing function

for the probability. This sum generally has too many terms for computation to be tractable. Since the joint distribution, $p(\underline{x})$, is analytically intractable except for very small lattices, maximum likelihood estimation for the autologistic model is not feasible.

Many approaches to estimation for the autologistic model have been developed. Besag (1986; 1974; 1989) develops estimation procedures based on iterated conditional modes, coding, and a Bayesian setup, respectively. In all of these approaches, the coefficient β_1 is specified *a priori*. In the iterated conditional modes approach (Besag, 1986) each site is updated based on information from the rest of the lattice, along a raster scan using the conditional likelihood. A raster scan simply starts at one corner, say the bottom left and visits sites sequentially left to right, up successive rows. The coding technique (Besag, 1974) works with partial sets of data. For instance, using a second order neighborhood, there would be eight neighbors for each site and eight sets of partial data. The joint distribution of sites with the same relative neighborhood position (e.g., the top right corner site of the neighborhood) given the rest of the sites is used to obtain estimates. The relative position is shifted and the next joint distribution is used. The resulting estimate is a combination of these joint distributions. Besag (1989) adopts a Bayesian estimation approach for the autologistic model applying a Gibbs sampler (Geman and Geman, 1984). Possolo (1986) suggests a logit approach to estimation as an alternative to pseudolikelihood estimation. Gray *et al.* (1992) combine these approaches for estimation of noisy binary images.

2.4 Extensions to the Autologistic Model

Several authors have extended the autologistic model to accommodate various applications. We'll briefly discuss these applications and then summarize the extension which formed the basis of our work.

2.4.1 Applications of the Autologistic Model

Besag, York, and Mollie (1991) apply the autologistic model to both archeology and epidemiology data. They use the Gibbs sampler to obtain Bayes estimates of the coefficients.

Gumpertz, Graham and Ristaino (1997) model spatial patterns of disease in agricultural fields using the autologistic model. They add covariate terms in the model. In their model, they have complete data for the response and covariates as opposed to a sample of observed responses. They estimate parameters using maximum pseudolikelihood (Besag, 1974). The pseudolikelihood is the likelihood which results by assuming conditional independence of the true presence/absence at sites given the other sites.

Preisler (1993) modeled destruction due to bark beetles using an autologistic model. He also employed maximum pseudolikelihood estimation.

Huffer and Wu (1995) use an autologistic model with covariates to model the distribution of several plant species in Florida using Markov chain Monte Carlo (MCMC) maximum likelihood method of Geyer and Thompson (1992). They also investigate the distributional properties of their MCMC estimates. These authors (Wu and Huffer, 1997) compare three methods of estimation: coding, maximum pseudolikelihood, and their MCMC method. Their MCMC method produces smaller

standard errors and mean square error, but the bias in their estimates is consistently larger than the bias for the other two methods.

Augustin, *et al.* (1998) develop an autologistic model to model the spatial distribution of red deer in Scotland. They develop a Gibbs sampler to estimate true presence/absence, but they do not put prior distributions on the parameters in the autologistic model. Instead they use the pseudolikelihood to estimate these parameters.

2.4.2 Autologistic Model with Search Intensity

Heikkinen and Högmander (1994, HH hereafter) use the autologistic model for biogeographic range estimates. HH extend the work of estimating biogeographic distributions introduced by Högmander and Møller (1995) who use the autologistic model to estimate distribution maps of several bird species in central Finland. Högmander and Møller extended the autologistic model to account for heterogeneous research activity, and they use maximum marginal posterior estimation and the iterated conditional modes algorithm for estimation. HH also use the level of search employed on each site to extend the autologistic model in predicting probability of presence but use a Bayesian setup for estimation.

HH apply their methodology to model presence/absence of a species of toad. The observations are made with error corresponding to the level of search employed. If the site was intensely searched and no toads were found, then an observed absence at that site is deemed more likely to be correct. If the site was searched only slightly, then an observed absence would not be so believable. This added information enters the model through an altered form of the likelihood function, $f(\underline{y} \mid \underline{x})$.

To introduce their extended model, further notation is necessary. HH define the spatial covariate, s_i, as the number of neighbors with presence minus the number of neighbors with absence.

$$s_i = \sum_{j \in \delta_i} \left(\mathrm{I}_{[x_j=1]} - \mathrm{I}_{[x_j=0]} \right).$$

This form of the spatial covariate has a range of integers from minus the number of neighbors to the number of neighbors. Negative values indicate a 'whiter' neighborhood, and positive values a 'blacker' neighborhood.

HH categorize search intensity into three levels corresponding to high, medium and low search intensity. The levels are denoted by $\epsilon = 0, 1,$ or 2 and g_{e_i} is the search intensity at the i^{th} site. The probability of observed absence given the species is present, is given by,

$$p(y_i = 0 \mid x_i) = \begin{cases} 1, & \text{if } x_i = 0, \\ g_{e_i}, & \text{if } x_i = 1. \end{cases}$$

So the distribution of g_ϵ is the probability of "missing" the species at the specified search level, ϵ.

HH adopt a fully Bayesian approach for estimation. I will present the likelihood function, priors, and posterior as HH have formulated them. The likelihood function for the lattice is, by assumption, the product of the likelihood function for each site.

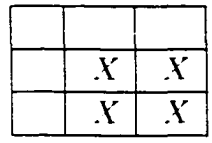
$$f\left(\underline{y} \mid \underline{x}, \underline{g}\right) = \prod_{i=1}^{N} p\left(y_i \mid x_i, g_{e_i}\right).$$

The parameters in this likelihood function are the true presence/absence, $\underline{x}$, and the the detectability, g_{e_i}, $e_i = 0, 1, 2$.

From the locally dependent Markov random field assumption it follows that the prior distribution on $\underline{x}$ is

$$\pi(\underline{x} \mid \beta) \propto \exp(\underline{s}\beta). \tag{2.4}$$

Figure 2.2: Specification of β_{MAX} for Autologistic Model



This likelihood function is intractable due to the normalizing constant which is the sum of this exponential function over all possible configurations of species presence/absence. The prior distribution for the detectability parameters, g_r, is defined to provide the natural ordering, $0 \leq g_2 \leq g_1 \leq g_0 \leq 1$. The probability of missing the species increases as search level decreases. The hyperprior for β is chosen as uniform over the range 0 to β_{MAX}. The value for β_{MAX} is determined by putting a minimum probability on finding a 'corner' of presence sites. If the true neighborhood has a corner of species presence, shown as X's in Figure 2.2, then the probability of predicted presence at the center site is set to be greater than or equal to some specified value. For Figure 2.2, $s_i = 4 - 5 = -1$, so β_{MAX} can be solved for.

The full posterior distribution for the parameters of the autologistic model of HH is

$$p\left(\underline{x}, \beta, \underline{g} \mid \underline{y}\right) \propto f\left(\underline{y} \mid \underline{x}, \underline{g}\right) \pi\left(\underline{x} \mid \beta\right) \pi\left(\underline{g}\right) \pi\left(\beta\right).$$

This posterior distribution is intractable due to the prior on $\underline{x}$.

In HH, estimation is accomplished via a 'Gibbs-Hastings' sampler. For a Gibbs sampler, all full conditional distributions are needed. For the autologistic model of HH, the full conditional distribution for β is intractable, so a pseudolikelihood approximation and Metropolis-Hastings step is used. This approach will be described for our extended model in 3.3.3.

HH obtain results consistent with the sightings and search level. They present posterior distributions with little variability for the estimated parameters. With this work HH have opened the door for further extensions of the autologistic model.

The extensions of HH are not general enough to fill our needs. We need to extend the detectability to involve sites that were not sampled. We also include habitat covariates.

Chapter 3

AUTOLOGISTIC MODEL WITH COVARIATES FOR SAMPLE DATA

Our goal for the USFS project is to produce a map of predicted probability of presence for each species over an area of interest. The available information we have is sighting information at some sites, sampling status for all sites (each site is sampled or not sampled), covariate information (which is assumed to be related to species presence) at all sites, and the assumption that the species form clusters. The extensions to the autologistic model presented in the preceeding chapter do not fully meet our needs. Unsampled sites have no presence/absence information. This is different from the models thus far presented since there was an assumption made that observations are available, albiet with error, at every site on the grid.

We introduce the autologistic model with covariates for sample data which incorporates sample observed presence/absence data, and covariates, and we use a Bayesian setup for estimation. The chapter begins with an introduction of notation to be used. The Bayesian framework is then presented by defining and deriving the likelihood function, priors and posterior. We then detail the estimation procedure which involves a Gibbs-Hastings sampling scheme.

3.1 Notation

In this section, we introduce the notation for the autologistic model with covariates for sample data. Additional clarification is provided in subsequent sections where needed. We defined notation for the basic autologistic model in Chapter 2. We use similar notation here, but some of the notation takes on slightly different meaning for the autologistic model with covariates for sample data. Here we present only terms whose meaning has been altered or new notation. All notation is summarized in Table 3.1. We will use the language of species presence/absence for the binary response, although the autologistic model with covariates for sample data is applicable to situations appropriate for the basic autologistic model such as image analysis.

We assume the data have been observed on a finite regular lattice. The sites are defined as before as the area between gridlines. We denote the number of sites in this lattice, N. A setup for hexagons or other shapes is possible as well.

3.1.1 Observed Data

Observed presence or absence is available at only some sites since some of the sites were not sampled. We define y_i to be 1 if the species was observed at site i, and redefine y_i to be 0 if either the species was not observed at site i or site i was not sampled. A value of $y_i = 0$ at site i may represent observed absence or lack of information. The observed values for the lattice are denoted by the vector $\underline{y} = (y_1, \ldots, y_N)$.

In addition to the observed presence/absence values, $\underline{y}$, we define a variable $\underline{a}$ representing the sampling status for all sites. We define $a_i = 1$ if site i was sampled and $a_i = 0$ if site i was not sampled. If site i was not sampled, $a_i = 0$, the species cannot be observed and necessarily $y_i = 0$. If the species was observed at site i,

$y_i = 1$, then the site must have been sampled, $a_i = 1$. The vector $\underline{a} = (a_1, \ldots, a_N)$ represents the sampling status for the lattice of interest.

The covariates, habitat variables in our application, are represented by the $N \times (p+1)$ matrix Z. The number of covariates measured at each site is denoted by p. The first column of Z is a column of ones which corresponds to an intercept term in the autologistic model with covariates for sample data (Equation 3.2, described below).

3.1.2 Neighborhood Characterization

We use a spatial covariate to characterize the presence or absence in the neighborhood. We define the spatial covariate for site i as the sum of the predicted probability of presence over sites in the neighborhood of site i, δ_i.

$$s_i = \sum_{j \in \delta_i} Pr(x_j = 1). \tag{3.1}$$

where $Pr(x_i = 1)$ is the predicted probability that the species is present in site i. Apriori, $P(x_i = 1)$ is equal to either 0 or 1. It is typical to adjust the spatial covariate for edge effects. The sites on the edge of the lattice have fewer neighbors than the interior sites and therefore have less information contained therein. When we perform estimation, we use only interior sites. Thus these edge sites, which have limited information, do not affect estimation. We discuss this estimation procedure in detail in Section 3.3.3.2.

Additional parameters, $\underline{\theta}$ and β, will be defined in Section 3.2.2 The other notation remains as defined in Chapter 2. The notation for the autologistic model with covariates for sample data is summarized in Table 3.1.

Table 3.1: Notation for the Autologistic Model with Covariates for Sample Data

N	number of sites in the area of interest
p	the number of covariates
a_i	$\begin{cases} 1 & \text{site } i \text{ was sampled} \\ 0 & \text{site } i \text{ was not sampled} \end{cases}$
y_i	$\begin{cases} 1 & \text{observed species presence at site } i \\ 0 & \text{observed species absence at site } i, \text{ or site } i \text{ was not sampled} \end{cases}$
x_i	$\begin{cases} 1 & \text{true species presence at site } i \\ 0 & \text{true species absence at site } i \end{cases}$
$\underline{x}_{-i}$	$\{x_j : j \neq i\}$ = true species presence/ absence for all sites except site i
δ_i	neighborhood of site i
$\underline{x}_{\delta_i}$	$\{x_j : j \in \delta_i\}$ = true presence/absence for neighbors of site i
$\underline{z}_i$	$1 \times (p+1)$ vector of covariates for site i
$\underline{\theta}$	$1 \times (p+1)$ vector of covariate coefficients
s_i	spatial covariate for site i
β	spatial covariate coefficient

3.2 Bayesian Formulation

The autologistic model with covariates for sample data is defined

$$\Pr(x_i = 1 \mid \underline{x}_{-i}, \underline{\theta}, \beta) = \frac{\exp\{\underline{z}_i\underline{\theta}^T + \beta s(x_i)\}}{1 + \exp\{\underline{z}_i\underline{\theta}^T + \beta s(x_i)\}}. \tag{3.2}$$

where $\underline{z}_i$ is a vector of covariates for the ith site with the first element equal to 1. $\underline{\theta}$ is a vector of parameters for the covariates, and β is the parameter associated with the spatial covariate, $s(x_i)$. The spatial covariate for site i, $s(x_i)$, is equal to the total number of sites where the species was present in the neighborhood of site i.

Since $s(x_i)$ is the *auto*covariate, we have chosen to treat this term separately from the other covariates for ease of understanding.

We consider a Bayesian setup for the autologistic model with covariates for sample data for two reasons. First, the Bayesian setup allows the incorporation of existing information through the prior distributions. The incorporation of information is important for ongoing monitoring programs since new information is always being gained. The USFS project involves development of sampling plans which are to be the basis for monitoring programs. The USFS will monitor many of the rarely seen species by sampling at regular time intervals. For a species being monitored, a sampling plan would be developed or modified for each time interval of monitoring. Information obtained in one sample will be used in the model to predict probability of presence. These predicted probabilities of presence will be used to increase the efficiency of the sampling plan for the following sampling period. This monitoring procedure lends itself to defining prior distributions based on information from the previous sample. In this way, information gained from the current sampling can readily be utilized for the next sampling plan inside a Bayesian framework. The logistics for updating prior distributions would be addressed after the initial sample data is collected. The posterior distributions from the first sample will be used as the prior distributions for the map of predicted probabilities used to design the second sample.

The second reason to consider a Bayesian framework is for ease in estimation of parameters. The likelihood for the true scene, $p(\underline{x})$, is intractable, as described below. Straightforward maximum likelihood estimation cannot be implemented in this case, so an alternative estimation procedure is necessary. Other options for estimation, including maximum pseudolikelihood estimation, were discussed in Chapter 2.

We introduce a Bayesian framework for estimation of parameters for the autologistic model with covariates for sample data by first introducing the likelihood function, prior distributions and the posterior distribution.

3.2.1 Likelihood Function

The likelihood function is the joint density of the data given the parameters. For the autologistic model with covariates for sample data we define the likelihood function of the joint density of $\underline{y}$, given the underlying true presence/absence $\underline{x}$, as $f\left(\underline{y} \mid \underline{x}\right)$. By assumption (Chapter 2), this joint density is the product of the conditional densities at each site.

$$f\left(\underline{y} \mid \underline{x}\right) = \prod_{i=1}^{N} p(y_i \mid x_i).$$

since each site is conditionally independent, given the truth at that site. We use sample observed presence/absence data instead of observations at every site, so the likelihood function for the autologistic model with covariates for sample data will be modified from the likelihood function in the basic autologistic model.

We consider the likelihood function to be a function of detectability. Probability of detection is the likelihood function of observing the species given that the species is present. As in Heikkinen and Högmander (1994, HH hereafter), we formulate the likelihood function in terms of nondetectability. Nondetectability is defined as the likelihood of not observing the species given that the species is present, or restated, the likelihood of 'missing' the species in our search. For now we concentrate on the sample data setup with two levels of search intensity; each site is sampled or is not sampled. We assume that sampling is done with a high level of rigor so that nondetection probability is zero for sampled sites and one for unsampled sites.

Extensions to this likelihood function setup to account for imperfect detection are described in Chapter 5.

Our likelihood function limits the possible *true* scenes to be consistent with the observed sample. We define possible true scenes to be only those configurations of presence/absence over the lattice for which the unsampled sites, $a_i = 0$, can have either $x_i = 1$ or $x_i = 0$, but sampled, $a_i = 1$, sites are fixed at $x_i = y_i$. The modified likelihood function allows updating for unsampled sites, but fixes presence/absence for sampled sites. Thus we define the likelihood function for the autologistic model with covariates for sample data as the product of $p(y_i, a_i \mid x_i)$ over all sites where

$$Pr(y_i = 0, a_i \mid x_i) = \begin{cases} 1, & \text{if } x_i = 0 \\ 1 - a_i, & \text{if } x_i = 1. \end{cases} \tag{3.3}$$

So we necessarily observe $y_i = 0$ if the species is truly absent or if we do not search the site. Similarly,

$$Pr(y_i = 1, a_i \mid x_i) = \begin{cases} 0, & \text{if } x_i = 0 \\ a_i, & \text{if } x_i = 1. \end{cases} \tag{3.4}$$

The likelihood function defined in this way limits possible true scenes to those consistent with the observed data.

3.2.2 Prior Distributions

In a Bayesian setup, the parameters in the likelihood function, $\underline{x}$ in our case, are considered to be random variables with distributions. The distributions for these parameters are prior distributions. Prior distributions reflect beliefs about the parameter values, prior to observing the data.

The covariates, Z, are related to the true presence/absence, $\underline{x}$. We introduce a prior distribution $\pi(\underline{x})$, which is a function of the new covariates as well as the

spatial covariate. By assumption, the true presence/absence on the lattice of interest is represented by a locally dependent Markov random field (LDMRF, Chapter 2). The prior distribution on $\underline{x}$ is,

$$\pi(\underline{x} \mid \beta, \underline{\theta}) \propto \exp\left\{\left(\sum_{i=1}^{N} z_i x_i\right)\underline{\theta}^T + \beta \sum_{i<j} x_i x_j I_{[i\sim j]}\right\}. \tag{3.5}$$

where $I_{[i\sim j]} = 1$ when site i and j are neighbors and 0 otherwise. Note that this is the *a priori* definition of s_i (Eqn 3.1). As in the basic autologistic model, the form of the prior distribution is of a logistic regression function where the spatial covariate is a function of the neighborhood responses. The prior distribution on $\underline{x}$ is analytically intractable since the sum in the denominator is over all possible configurations of presence/absence on the lattice.

In a fully Bayesian hierarchical framework, the parameters of the prior distribution, in our case $\underline{\theta}$ and β, are in turn considered random variables with distributions called hyperprior distributions. We next present the form of the hyperprior distributions we use in the autologistic model with covariates for sample data.

The hyperprior distributions for $\underline{\theta}$, and β are specified as follows. We assume

$$\underline{\theta} \sim N(\underline{0}, \Sigma),$$

a multivariate normal distribution with mean $\underline{0}$, and a diagonal covariance matrix, Σ, with $\left(\tau_0^2, \tau_1^2, \ldots, \tau_p^2\right)$ on the diagonal. Here $(\tau_0^2, \tau_1^2, \ldots, \tau_p^2)$ are hyperparameters to be chosen. The covariate coefficients, $\underline{\theta}$, are assumed to be independent *a priori*. This assumption is not reasonable when columns of Z constitute a set of dummy variables for a categorical variable. Raftery, Madigan, and Hoeting (1997) suggest an appropriate hyperprior distribution for this situation in the context of linear regression.

The hyperprior distribution we assume for β, the spatial parameter, is gamma, $\Gamma(\psi, \alpha)$, where

$$\pi(\beta) = \frac{\alpha^{-\psi}}{\Gamma(\psi)} \beta^{\psi-1} \exp(-\beta/\alpha). \tag{3.6}$$

and ψ and α are hyperparameters to be chosen. This hyperprior distribution constrains the spatial parameter to be greater than 0. For populations that form clusters, this is certainly a reasonable constraint. If the population of interest is not necessarily clustered, or is repulsive to nearby presence, then a different hyperprior distribution would need to be specified to allow negative values of the coefficient for the spatial covariate β.

3.2.3 Posterior Distribution

The posterior distribution is the conditional distribution of all of the parameters given the data. By Bayes theorem, the posterior distribution is the normalized likelihood function times the prior distribution(s). For the autologistic model with covariates for sample data, the posterior distribution is

$$p\left(\underline{x}, \beta, \underline{\theta} \mid \underline{y}, \underline{a}\right) \propto f\left(\underline{y}, \underline{a} \mid \underline{x}\right) \pi(\underline{x} \mid \beta, \underline{\theta}) \pi(\beta) \pi(\underline{\theta}). \tag{3.7}$$

This posterior distribution is analytically intractable due to the normalizing constant in the prior distribution for $\underline{x}$, $\pi(\underline{x} \mid \beta, \underline{\theta})$.

3.3 Estimation

The parameters in the autologistic model with covariates for sample data to be estimated are $\underline{x}$, β, and $\underline{\theta}$. Since the posterior distribution is analytically intractable, we cannot easily obtain maximum posterior estimates of these parameters. We mention two techniques for performing this estimation. Exact Sampling

(Propp and Wilson, 1996) is a Markov chain Monte Carlo technique which could be used for estimation of the presence/absence parameters $\underline{x}$. The methodology uses the idea of coupling from the past. Heuristically, the image is started at the two extremes, all presence and all absence, which are updated in reverse until they couple, or achieve the same image. At the point it can be assumed that the chain has converged and all subsequent iterations are used in estimation. In practice coupling is difficult to achieve so we adopt an alternative estimation technique, Gibbs sampling (Geman and Geman, 1984). The Gibbs sampling estimation procedure uses the full conditional distributions in an iterative procedure to obtain estimates of the parameters from the intractable posterior distribution. We describe this Gibbs sampling methodology, derive the full conditional distributions, describe a modification of the Gibbs sampling for full conditional distributions which are intractable, and discuss convergence criteria.

3.3.1 Gibbs Sampling

For the autologistic and related models, many authors have used a Bayesian framework and MCMC methodologies for estimation (Besag *et al.*, 1991; Huffer and Wu, 1995; Augustin *et al.*, 1998; Heikkinen and Högmander, 1994). HH utilize a fully Bayesian hierarchical framework and a Gibbs-Hastings sampling procedure for parameter estimation. We extend this approach to obtain parameter estimates for the autologistic model with covariates for sample data.

The algorithm proceeds by sampling iteratively from the full conditional distributions for the parameters. A full conditional distribution for a parameter is the conditional distribution for that parameter given all other parameters and the data. One iteration of the Gibbs sampler involves sampling a value for each parameter from the corresponding full conditional distribution. The full conditional

distributions are updated following the sampling for each parameter, thus utilizing the updated information in a continuous cycle. This procedure is iterated until the Gibbs sampler converges to the appropriate stationary distribution. The collection of sampled values can be used to estimate the distribution of each parameter. The means of these distributions are used as the parameter estimates.

There are $N - n + 2 + p$ parameters to be estimated in the autologistic model with covariates for sample data: $N - n$ presence/absence parameters for unsampled sites, $\underline{x}$; p coefficients, $\theta_1, \ldots, \theta_p$, corresponding to the covariates, 1 intercept term, θ_0; and 1 coefficient for the spatial covariate, β. The Gibbs sampling estimation procedure requires a full conditional distribution for each parameter.

We derive the full conditional distributions and describe the sampling procedure for each parameter in the autologistic model with covariates for sample data. We then summarize the whole procedure which involves specifying inputs, sampling values, and monitoring convergence.

3.3.2 Full Conditional Distributions for True Presence/Absence

The full conditional distribution for $\underline{x}$ is formulated using the posterior distribution in Equation 3.7, and the assumed mutual independence of $\underline{y}$, β, and $\underline{\theta}$.

$$\begin{aligned} p\left(\underline{x} \mid \underline{y}, \underline{a}, \underline{\theta}, \beta\right) &= \frac{p\left(\underline{x}, \beta, \underline{\theta} \mid \underline{y}, \underline{a}\right)}{p\left(\beta, \underline{\theta} \mid \underline{y}, \underline{a}\right)} \\ &\propto \frac{f\left(\underline{y}, \underline{a} \mid \underline{x}\right) \pi\left(\underline{x} \mid \beta, \underline{\theta}\right) \pi(\beta) \pi\left(\underline{\theta}\right)}{\pi(\beta)\pi(\underline{\theta})} \\ &= f\left(\underline{y}, \underline{a} \mid \underline{x}\right) \pi\left(\underline{x} \mid \beta, \underline{\theta}\right) \\ &\propto \left(\prod_{i=1}^{N} p(y_i, a_i \mid x_i)\right) \exp\left\{\left(\sum_{i=1}^{n} z_i x_i\right) \underline{\theta}^T + s_i \beta\right\}. \end{aligned} \tag{3.8}$$

where the likelihood function, $p(y_i, a_i \mid x_i)$, is given in Equations 3.3 and 3.4. This distribution is intractable due to the sum in the denominator over all possible configurations of presence/absence over the lattice.

The Gibbs sampling algorithm uses the full conditional distribution for *each* parameter, so we are interested in the full conditional distribution for $x_i, i = 1, 2, \ldots, N$, which is formulated as

$$p\left(x_i \mid \underline{y}, \underline{a}, \underline{x}_{-i}, \beta, \underline{\theta}\right) \propto p(y_i, a_i \mid x_i)\, \pi\left(x_i \mid \underline{x}_{\delta_i}, \beta, \underline{\theta}\right),$$

the likelihood function for the observation at site i given the truth at site i multiplied by the conditional prior distribution of the truth at site i given the truth at all other sites and the model parameters. The conditional prior distribution for each x_i given $\underline{x}_{-i}$ is

$$\pi\left(x_i \mid \underline{x}_{\delta_i}, \beta, \underline{\theta}\right) = \frac{\exp\left(x_i\left(\underline{z}_i\underline{\theta}^T + s_i\beta\right)\right)}{1 + \exp\left(\underline{z}_i\underline{\theta}^T + s_i\beta\right)}. \tag{3.9}$$

the logistic function of the spatial covariate and habitat covariates which is obtained from the prior distribution for $\underline{x}$ in Equation 3.5. Note that the full conditional distribution for the truth at each site, x_i is analytically tractable since the sum in the denominator, over all possible x_i, is the sum over $x_i = 0$ and $x_i = 1$.

We use the likelihood function in Equations 3.3 and 3.4 and the conditional prior distribution for x_i in Equation 3.9 to obtain the full conditional distribution for x_i:

$$\begin{aligned} p\left(x_i = 1 \mid \underline{y}, \underline{a}, \underline{x}_{-i}, \beta, \underline{\theta}\right) &= p(y_i, a_i \mid x_i = 1)\, \pi\left(x_i = 1 \mid \underline{x}_{\delta_i}, \beta, \underline{\theta}\right) \\ &\propto \begin{cases} \exp(\underline{z}_i\underline{\theta}^T + s_i\beta), & a_i = 0, \text{ or } a_i = 1, y_i = 1 \\ 0, & a_i = 1, y_i = 0. \end{cases} \end{aligned} \tag{3.10}$$

$$p\left(x_i = 0 \mid \underline{y}, \underline{a}, \underline{x}_{-i}, \beta, \underline{\theta}\right) = p(y_i, a_1 \mid x_i = 0)\, \pi\left(x_i = 0 \mid \underline{x}_{\delta_i}, \beta, \underline{\theta}\right)$$

$$\propto \begin{cases} 1, & a_i = 0, \text{ or } a_i = 1, y_i = 0 \\ 0, & a_i = 1, y_i = 1. \end{cases} \tag{3.11}$$

Combining these proportional distributions, we obtain the full conditional distribution for $x_i = 1$.

$$p\left(x_i = 1 \mid \underline{y}, \underline{a}, \underline{x}_{-i}, \beta, \underline{\theta}\right) = \begin{cases} \frac{\exp(\underline{z}_i \underline{\theta}^T + s_i \beta)}{1+\exp(\underline{z}_i \underline{\theta}^T + s_i \beta)}, & a_i = 0, \\ \frac{\exp(\underline{z}_i \underline{\theta}^T + s_i \beta)}{0+\exp(\underline{z}_i \underline{\theta}^T + s_i \beta)} = 1, & a_i = 1, y_i = 1, \\ \frac{0}{1+0} = 0, & a_i = 1, y_i = 0. \end{cases} \tag{3.12}$$

For our sample data setup, only sites that were not sampled, $a_i = 0$, are updated. The $N - n$ full conditional distributions for x_i are fully specified by Equation 3.12 and can be readily sampled from in the Gibbs sampler.

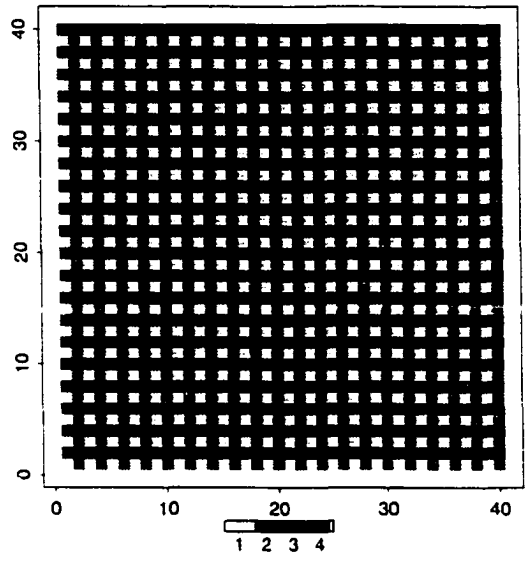
Updating $\underline{x}$ in the Gibbs Sampler

We use the full conditional distribution for true presence at site i (3.12) to update the predicted presence, $\underline{x}$, in three steps. The first step is to calculate $\underline{\hat{p}}$, the predicted probability of presence, for unsampled sites, $a_i = 0$. The second step is to calculate the spatial covariate, $\underline{s}$. The third step is to update the prediction of presence/absence, $\underline{x}$. Below we describe this three step process for obtaining sample values at iteration t. The initialization step, $t = 0$, is described in Section 3.3.4.

For iteration t, the first step for sites i such that $a_i = 0$, $\hat{p}_i = Pr(x_i = 1)$ is calculated as

$$\begin{aligned} \hat{p}_i^{(t)} &= P(x_i^{(t)} = 1) \\ &= \begin{cases} \frac{\exp\left(\underline{z}_i \underline{\theta}^{T(t-1)} + s_i^{(t-1)} \beta^{(t-1)}\right)}{1+\exp\left(\underline{z}_i^T \underline{\theta}^{T(t-1)} + s_i^{(t-1)} \beta^{(t-1)}\right)}, & a_i = 0, \\ y_i, & a_i = 1, \end{cases} \end{aligned} \tag{3.13}$$

Figure 3.1: Groups of Independent Sites



where $\underline{s}^{(t-1)}, \beta^{(t-1)}, \underline{\theta}^{(t-1)}$ are parameters from iteration $(t-1)$.

The second step is to calculate the spatial covariate $\underline{s}$. By definition, the LDMRF utilizes the neighboring presence/absence information, x_{δ_i}, not estimates of the probability of presence. However the estimation algorithm does not converge when the spatial covariate is a function of estimated presence/absence, so we use the spatial covariate defined below to approximate a LDMRF. We use estimates of the probability of presence from iteration t, $\hat{\underline{p}}^{(t)}$, from the first step. We calculate the spatial covariate defined in Section 3.1 as,

$$s_i^{(t)} = \sum_{j \in \delta_i} \hat{p}_j^{(t)}.$$

The parameters $\hat{\underline{p}}^{(t)}$ and $\underline{s}^{(t)}$ are not updated simultaneously for all sites. We calculate $\hat{p}_i^{(t)}$ simultaneously within a group of independent sites and then calculate $s_i^{(t)}$ simultaneously within that same group of independent sites. This is done for each group of independent sites in turn. We use four groups of independent sites, consisting of groups of sites from every other row and every other column. This

independent group structure is seen for the whole lattice in Figure 3.1. Group 1 starts with the bottom left corner and is represented by white sites. Group 2 contains the next site to the right and all other light grey sites. The algorithm proceeds analogously for groups 3 and 4. The importance of sampling simultaneously only within independent groups will be discussed in Section 3.3.5.

The third step is to update the presence/absence parameters, $\underline{x}$. To assign updated presence/absence at site i, at iteration t, we sample from a Bernoulli distribution with probability $\hat{p}_i^{(t)}$ (Eqn 3.13). Thus, the probability that $x_i^{(t)}$ equals 1 is $\hat{p}_i^{(t)}$.

3.3.3 Full Conditional Distributions for the Model Parameters

The full conditional distributions for β and $\underline{\theta}$ are formulated using the full conditional distribution of $\underline{x}$ and the assumption of *a priori* mutual independence of $\underline{y}$, β, and $\underline{\theta}$ as,

$$\begin{aligned} p\left(\beta \mid \underline{y}, \underline{x}, \underline{\theta}\right) &\propto p\left(\underline{x} \mid \underline{y}, \beta, \underline{\theta}\right) p\left(\beta \mid \underline{y}, \underline{\theta}\right) \\ &= p\left(\underline{x} \mid \underline{y}, \beta, \underline{\theta}\right) \pi(\beta), \end{aligned} \tag{3.14}$$

and

$$\begin{aligned} p\left(\theta_k \mid \underline{y}, \underline{x}, \beta, \underline{\theta}_{-k}\right) &\propto p\left(\underline{x} \mid \underline{y}, \beta, \underline{\theta}\right) p\left(\theta_k \mid \underline{y}, \beta, \underline{\theta}_{-k}\right) \\ &= p\left(\underline{x} \mid \underline{y}, \beta, \underline{\theta}\right) \pi(\theta_k), \end{aligned} \tag{3.15}$$

for $k = 0, \ldots, p$. All of the full conditional distributions for the coefficient parameters are analytically intractable since they involve the intractable full conditional distribution for $\underline{x}$ and thus we cannot sample directly from the full conditional distributions. To estimate the parameters we obtain sample values from these full conditional distributions using a Hastings-Metropolis step within the Gibbs sampler for each β and θ_k, $k = 0, \ldots, p$.

3.3.3.1 Hastings-Metropolis Steps

The Hastings-Metropolis algorithm (Hastings, 1970) is an MCMC sampling method for distributions that are intractable or distributions from which it is difficult to sample. The difficult distribution of interest is called the target distribution, denoted by $p(\cdot)$. In general, although p is intractable, the ratio $p(x)/p(x')$, for $x \neq x'$, usually is not intractable. The method of Hastings-Metropolis utilizes a user-specified, easy to sample from, proposal distribution, denoted by $g(\cdot)$. The domain of the proposal distribution must include the domain of the target distribution.

The Hastings-Metropolis algorithm proceeds as follows. At iteration t, a value is sampled from the proposal distribution $g(\cdot)$. This value is either accepted as a valid value from the target distribution or rejected. The probability that a value is accepted is calculated as the ratio of target and proposal distributions at the previous value and the new proposed value. In general the acceptance probability, α, has the form,

$$\alpha = \frac{p(\text{new})\, g(\text{previous})}{p(\text{previous})\, g(\text{new})}. \tag{3.16}$$

If the value is rejected, then the 'updated' value takes on the previously sampled value from iteration $(t-1)$.

Usually the ratio of the target distributions, $p(\text{new})/p(\text{previous})$, is not intractable because the difficult normalizing constant in p usually cancels. The autologistic model with covariates for sample data involves a sum which is a function of the parameters of interest. The normalizing constants in $p(\text{previous})$ and $p(\text{new})$ contain the previous value or the new value, respectively, so the normalizing constants do not cancel. Thus, the ratio of the target distributions in α is intractable, and α is not calculable. We use alternative target distributions instead of the full

conditional distributions in order to calculate α. As in HH, we use the full conditional *pseudolikelihood* distributions. Next we define the target distributions, which are the full conditional pseudolikelihood distributions, and specify proposal distributions.

3.3.3.2 Pseudolikelihood

We define the full conditional pseudolikelihood distributions for the parameters of interest, β and $\underline{\theta}$, as the target distributions for the Hastings-Metropolis steps. The full conditional distributions for β and $\underline{\theta}$ are functions of the full conditional distribution for $\underline{x}$. Thus, the full conditional *pseudolikelihood* distributions for β and $\underline{\theta}$ are functions of the full conditional *pseudolikelihood* for $\underline{x}$.

The full conditional pseudolikelihood distribution for $\underline{x}$ is the distribution resulting from the assumption that true presence at each site is conditionally independent given the other site values. Since the border sites, those sites with fewer than eight neighbors, are estimated from a reduced pool of information, we follow the suggestion of HH and use only interior sites to estimate the model parameters. For site i, we define $b_i = 1$ if the site is on the interior of the area of interest and define $b_i = 0$ otherwise. The pseudolikelihood distribution for $\underline{x}$ is the product over only sites for which $b_i = 1$. The pseudolikelihood distribution will be represented by $pl(\cdot)$. Assuming this conditional independence, the full conditional pseudolikelihood distribution for $\underline{x}$ is

$$\begin{aligned} pl\left(\underline{x} \mid \underline{y}, \beta, \underline{\theta}\right) &= \prod_{i=1}^{N} p\left(x_i \mid \underline{y}, \underline{x}_{-i}, \beta, \underline{\theta}\right) \\ &= \prod_{\forall i \ni : b_i = 1} \frac{\exp\left(x_i\left(\underline{z}_i \underline{\theta}^T + s_i \beta\right)\right)}{1 + \exp\left(\underline{z}_i \underline{\theta}^T + s_i \beta\right)}. \end{aligned} \tag{3.17}$$

Thus the full conditional pseudolikelihood distribution for β is defined

$$\begin{aligned} pl\left(\beta \mid \underline{y}, \underline{x}, \underline{\theta}\right) &\propto pl\left(\underline{x} \mid \underline{y}, \beta, \underline{\theta}\right) \pi(\beta) \\ &= \left(\prod_{\forall i \ni : b_i = 1} \frac{\exp\left(x_i\left(\underline{z}_i \underline{\theta}^T + s_i \beta\right)\right)}{\left[1 + \exp\left(\underline{z}_i \underline{\theta}^T + s_i \beta\right)\right]}\right) \frac{\beta^{\psi-1} \exp\left(-(\beta/\alpha)\right)}{\alpha^{\psi} \Gamma(\psi)} \end{aligned} \quad (3.18)$$

and the full conditional pseudolikelihood distribution for θ_k is defined

$$\begin{aligned} pl\left(\theta_k \mid \underline{y}, \underline{x}, \beta, \underline{\theta}_{-k}\right) &\propto pl\left(\underline{x} \mid \underline{y}, \beta, \underline{\theta}\right) \pi(\theta_k) \\ &= \left(\prod_{\forall i \ni : b_i = 1} \frac{\exp\left(x_i\left(\underline{z}_i \underline{\theta}^T + s_i \beta\right)\right)}{1 + \exp\left(\underline{z}_i \underline{\theta}^T + s_i \beta\right)}\right) \\ &\quad \left(2\pi\tau_k^2\right)^{-1/2} \exp\left(-\frac{\theta_k^2}{2\tau_k^2}\right). \end{aligned} \quad (3.19)$$

for $k = 0, 1, \ldots, p$. These target distributions, the full conditional pseudolikelihood distributions, are fully specified, so the acceptance probabilities, α in Equation 3.16, can be calculated. We present the proposal distributions and acceptance probabilities which complete the sampling steps for β and $\underline{\theta}$ at iteration t.

3.3.3.3 Maximum Pseudolikelihood Estimation

We specify the proposal distributions to be used in the Hastings-Metropolis steps using the recommendation of HH. Under regularity conditions in Cox and Hinkley (1974), the log of the full conditional pseudolikelihood distribution of $\underline{x}$ is asymptotically normal with mean and variance equal to the maximum likelihood estimates from the pseudolikelihood. Since each pseudolikelihood distribution is a nonlinear function, we use a quasi-Newton optimizer to find the maximum pseudolikelihood estimate for the mean. The variance for each coefficient parameter is obtained using the observed Fisher information from its pseudolikelihood distribution. The specification of the proposal distribution is done at every iteration.

At iteration t, for each coefficient parameter, β and $\underline{\theta}$, we calculate the maximum pseudolikelihood estimate of the mean and the Fisher information from the full conditional pseudolikelihood distribution. The mean and variance obtained for each coefficient parameter define the parameters of the normal distribution used as the proposal distribution for each coefficient parameter. We use a nonlinear minimization procedure to obtain maximum pseudolikelihood estimates. We minimize the negative of the log of the full conditional pseudolikelihood distribution. For β, the negative log full conditional pseudolikelihood distribution is

$$\begin{aligned} -\log pl\left(\beta \mid \underline{y}, \underline{x}, \underline{\theta}\right) &= \sum_{\forall i \ni : b_i = 1} \log\left(1 + \exp\left(\underline{z}_i \underline{\theta}^T + s_i \beta\right)\right) \\ &\quad - \left(\sum_{\forall i \ni : b_i = 1} x_i \left(\underline{z}_i \underline{\theta}^T + s_i \beta\right)\right) \\ &\quad -(\omega - 1)\log(\beta) + \frac{\beta}{\alpha} + \omega \log(\alpha) + \log \Gamma(\omega), \end{aligned}$$

and for $\theta_k, k = 0, \ldots, p$, the negative log full conditional pseudolikelihood distribution is

$$\begin{aligned} -\log pl\left(\theta_k \mid \underline{y}, \underline{x}, \beta, \underline{\theta}_{-k}\right) &= \sum_{\forall i \ni : b_i = 1} \log\left(1 + \exp\left(\underline{z}_i \underline{\theta}^T + s_i \beta\right)\right) \\ &\quad - \left(\sum_{\forall i \ni : b_i = 1} x_i \left(\underline{z}_i \underline{\theta}^T + s_i \beta\right)\right) \\ &\quad -\frac{1}{2}\log\left(2\pi\tau_k^2\right) - \frac{\theta_k^2}{2\tau_k^2}. \end{aligned}$$

We use the `nlmin` function in Splus to obtain the maximum pseudolikelihood estimates, $\hat{\beta}_{\text{MPLE}}$ and $\underline{\hat{\theta}}_{\text{MPLE}}$. These values are then used as the means in the proposal distributions.

To obtain the variance for each proposal distribution we use the inverse of the observed Fisher information from each full conditional pseudolikelihood distribution.

The observed Fisher information is calculated as the negative of the second derivative of the log of the full conditional pseudolikelihood distribution, evaluated at the maximum pseudolikelihood estimate. The variance for the proposal distribution for β is derived as follows.

$$\begin{aligned}
\sigma^2_{\beta_{\text{MPLE}}} &= \left[\left[-\frac{\partial^2 \log pl(\beta)}{\partial \beta^2}\right]^{-1}\right]_{\beta=\hat{\beta}_{\text{MPLE}}} \\
&= \left[\left[-\frac{\partial^2}{\partial \beta^2}\left[-\sum_{\forall i \ni : b_i=1} \log\left(1+\exp\left(\underline{z}_i\,\underline{\theta}^T + s_i\beta\right)\right) + \sum_{\forall i \ni : b_i=1} x_i\left(\underline{z}_i\,\underline{\theta}^T + s_i\beta\right)\right.\right.\right. \\
&\qquad \left.\left.\left. +(\psi-1)\log(\beta) - \frac{\beta}{\alpha} - \psi\log(\alpha) - \log\Gamma(\psi)\right]\right]^{-1}\right]_{\beta=\hat{\beta}_{\text{MPLE}}} \\
&= \left[\left[-\frac{\partial}{\partial \beta}\left[-\sum_{\forall i \ni : b_i=1} \frac{s_i \exp\left(\underline{z}_i\,\underline{\theta}^T + s_i\beta\right)}{1+\exp\left(\underline{z}_i\,\underline{\theta}^T + s_i\beta\right)} + \sum_{\forall i \ni : b_i=1} x_i s_i + \frac{\psi-1}{\beta} - \frac{1}{\alpha}\right]\right]^{-1}\right]_{\beta=\hat{\beta}_{\text{MPLE}}} \\
&= \left[\left[\sum_{\forall i \ni : b_i=1} \frac{s_i^2 \exp\left(\underline{z}_i\,\underline{\theta}^T + s_i\beta\right)}{\left(1+\exp\left(\underline{z}_i\,\underline{\theta}^T + s_i\beta\right)\right)^2} + \frac{\psi-1}{\beta^2}\right]^{-1}\right]_{\beta=\hat{\beta}_{\text{MPLE}}} \qquad (3.20)
\end{aligned}$$

The hyperparameter ψ, is the specified scale from the hyperprior distribution on β given in Equation 3.6. The variance for the proposal distributions for θ_k, $k = 0, \ldots, p$, is derived as follows.

$$\begin{aligned}
\sigma^2_{\hat{\theta}_{k,\text{MPLE}}} &= \left[\left[-\frac{\partial^2 \log pl(\theta_k)}{\partial \theta_k^2}\right]^{-1}\right]_{\theta_k=\hat{\theta}_{k,\text{MPLE}}} \\
&= \left[\left[-\frac{\partial^2}{\partial \theta_k^2}\left[-\sum_{\forall i \ni : b_i=1} \log\left(1+\exp\left(\underline{z}_i\,\underline{\theta}^T + s_i\beta\right)\right) + \sum_{\forall i \ni : b_i=1} x_i\left(\underline{z}_i\,\underline{\theta}^T + s_i\beta\right)\right.\right.\right. \\
&\qquad \left.\left.\left. +\log\left(2\pi\tau_k^2\right) + \frac{\theta_k^2}{2\tau_k^2}\right]\right]^{-1}\right]_{\theta_k=\hat{\theta}_{k,\text{MPLE}}} \\
&= \left[\left[-\frac{\partial}{\partial \theta_k}\left[-\sum_{\forall i \ni : b_i=1} \frac{z_{ki} \exp\left(\underline{z}_i\,\underline{\theta}^T + s_i\beta\right)}{1+\exp\left(\underline{z}_i\,\underline{\theta}^T + s_i\beta\right)} + \sum_{\forall i \ni : b_i=1} x_i z_{ki} + \frac{\theta_k}{\tau_k^2}\right]\right]^{-1}\right]_{\theta_k=\hat{\theta}_{k,\text{MPLE}}} \\
&= \left[\left[\sum_{\forall i \ni : b_i=1}^{N} \frac{z_{ki}^2 \exp\left(\underline{z}_i\,\underline{\theta}^T + s_i\beta\right)}{\left(1+\exp\left(\underline{z}_i\,\underline{\theta}^T + s_i\beta\right)\right)^2} - \frac{1}{\tau_k^2}\right]^{-1}\right]_{\theta_k=\hat{\theta}_{k,\text{MPLE}}} \qquad (3.21)
\end{aligned}$$

The hyperparameter τ_k^2 is the specified variance from the normal hyperprior distribution on θ_k.

To increase efficiency in the sampling algorithm we multiply the variances in the proposal distributions by 2.38, as recommended by Gelman *et al.* (1995). Using an inflated variance in the proposal distributions allows for larger 'steps' in the Gibbs sampler. A sampled value which is far from the previous value is considered to be a large step. For the first iterations, these large steps allow for greater speed in coverage over the range of values. For the later steps, the difference that the increased variance makes is negligible.

The proposal distributions using the mean and variance from the full conditional pseudolikelihood distributions are,

$$g(\beta) = N\left(\hat{\beta}_{\text{MPLE}}, \left[\left[\sum_{\forall i \ni b_i=1} \frac{s_i^2 \exp\left(z_i \underline{\theta}^T + s_i\beta\right)}{\left(1+\exp\left(z_i \underline{\theta}^T + s_i\beta\right)\right)^2} + \frac{v-1}{\beta^2}\right]^{-1}\right]_{\beta=\hat{\beta}_{\text{MPLE}}}\right),$$

$$g(\theta_k) = N\left(\hat{\theta}_{k,\text{MPLE}}, \left[\left[\sum_{\forall i \ni b_i=1} \frac{z_{ki}^2 \exp\left(z_i \underline{\theta}^T + s_i\beta\right)}{\left(1+\exp\left(z_i \underline{\theta}^T + s_i\beta\right)\right)^2} + \frac{2}{\tau_k^2}\right]^{-1}\right]_{\theta_k=\hat{\theta}_{k,\text{MPLE}}}\right),$$

for $k = 0, \ldots, p$. To obtain sample values for β and $\underline{\theta}$ at iteration t, we sample a value for each coefficient parameter from these proposal distributions. The coefficients, β and $\underline{\theta}$, are sampled in random order. This eliminates any order bias in sampling. If β is always sampled first, for example, this new value is used in the target and proposal distributions for each θ_k. This constant order may prevent full exploration of the parameter space.

Each sampled value is then either accepted or rejected. The general acceptance probability was given in Equation 3.16. For $\beta^{(t)}$, this acceptance probability is,

$$\alpha_{\beta^{(t)}} = \frac{p\left(\beta^{(t)}\right) g\left(\beta^{(t-1)}\right)}{p\left(\beta^{(t-1)}\right) g\left(\beta^{(t)}\right)}$$

$$= \frac{\left(\prod_{i=1}^{N} pl\left(x_i^{(t)} \mid \underline{y}, \underline{x}_{-i}^{(t)}, \beta^{(t)}, \underline{\theta}\right)\right) \pi\left(\beta^{(t)}\right) \phi\left(\frac{\beta^{(t-1)} - \hat{\beta}_{\text{MPLE}}}{\sigma_{\hat{\beta}_{\text{MPLE}}}}\right)}{\left(\prod_{i=1}^{N} pl\left(x_i^{(t-1)} \mid \underline{y}, \underline{x}_{-i}^{(t-1)}, \beta^{(t-1)}, \underline{\theta}\right)\right) \pi\left(\beta^{(t-1)}\right) \phi\left(\frac{\beta^{(t)} - \hat{\beta}_{\text{MPLE}}}{\sigma_{\hat{\beta}_{\text{MPLE}}}}\right)}. \quad (3.22)$$

where $\phi(\cdot)$ is the density function for a standard normal random variable.

Similarly for θ_k, $k = 0, \ldots, p$, the value $\theta_k^{(t)}$ is either accepted or $\theta_k^{(t-1)}$ is used again. For $\theta_k^{(t)}$, $k = 0, \ldots, p$, the acceptance probabilities are

$$\alpha_{\theta_k^{(t)}} = \frac{p\left(\theta_k^{(t)}\right) g\left(\theta_k^{(t-1)}\right)}{p\left(\theta_k^{(t-1)}\right) g\left(\theta_k^{(t)}\right)}$$

$$= \frac{\left(\prod_{i=1}^{N} pl\left(x_i \mid \underline{y}, \underline{x}_{-i}, \beta, \underline{\theta}_{-k}, \theta_k^{(t)}\right)\right) \pi\left(\theta_k^{(t)}\right) \phi\left(\frac{\theta_k^{(t-1)} - \hat{\theta}_{k,\text{MPLE}}}{\sigma_{\hat{\theta}_{k,\text{MPLE}}}}\right)}{\left(\prod_{i=1}^{N} pl\left(x_i \mid \underline{y}, \underline{x}_{-i}, \beta, \underline{\theta}_{-k}, \theta_k^{(t-1)}\right)\right) \pi\left(\theta_k^{(t-1)}\right) \phi\left(\frac{\theta_k^{(t)} - \hat{\theta}_{k,\text{MPLE}}}{\sigma_{\hat{\theta}_{k,\text{MPLE}}}}\right)} \quad (3.23)$$

In this way, sampled values at iteration t are obtained.

3.3.4 Gibbs-Hastings Sampling

We have now described the Gibbs sampler and the Hastings-Metrolpolis steps in detail which used together construct our Gibbs-Hastings sampler. The Gibbs-Hastings sampler is used to obtain estimates for the $N - n + p + 2$ parameters, $\underline{x}$, β, and $\underline{\theta}$. We summarize the Gibbs-Hastings sampler steps below.

First the parameters $\underline{s}$, β, and $\underline{\theta}$ are initialized. There are two parts to the initialization step. First, we calculate the initial values of the spatial covariate, $\underline{s}$. Recall that we defined

$$s_i = \sum_{j \in \delta_i} P(x_j = 1). \quad (3.24)$$

The data, $\underline{y}$, consist of observed presence/absence for each site, not probability of presence, so we don't have values for $P(x_i = 1)$ before the algorithm is initialized. Thus for the initial estimate for the spatial covariate we use,

$$s_i^{(0)} = \sum_{j \in \delta_i} y_i. \quad (3.25)$$

This spatial covariate is similar to spatial covariates typically used in image analysis.

Second we initialize the coefficient parameters β and $\underline{\theta}$. We use logistic regression to estimate these parameters based on the initial spatial covariate, $\underline{s}^{(0)}$, the observed presence/absence, $\underline{y}$, and the covariates, Z, for sampled sites. The parameter estimates obtained from the logistic regression are used as the initial values, $\beta^{(0)}$, and $\underline{\theta}^{(0)}$. This assumes that response, $\underline{y}$, is independent between sites, and uses only information from sampled sites, but starts the estimation routine closer to the truth than arbitrary initial values would.

At iteration t, one cycle of the Gibbs sampler is described as follows:

1. Calculate the predicted probability of species presence $\underline{\hat{p}}^{(t)}$ given in Equation 3.13 and the spatial covariate $\underline{s}^{(t)}$ given in Equation 3.24.

2. Sample $\underline{x}^{(t)}$, the proposed estimate of the truth, from Bernoulli ($\underline{\hat{p}}^{(t)}$).

3. Sample each $\theta_k^{(t)}$, for $k = 0, \ldots, p$, and $\beta^{(t)}$ using a Hastings-Metropolis step as described in Section 3.3.3.1.

3.3.5 Convergence

Under only weak conditions, Geman and Geman (1984) show that the Gibbs sampler for Bayesian image analysis using a LDMRF assumption converges to the proper stationary distribution. The condition they specify is on the manner in which site parameters are updated. No two sites which are neighbors should be updated together, to ensure convergence of the Gibbs sampler.

The Gibbs sampling algorithm we suggest utilizes the full conditional distributions to obtain estimates from the posterior distribution. The parameter values,

$\underline{x}^{(t)}$, $\beta^{(t)}$, and $\underline{\theta}^{(t)}$, are sampled from their full conditional distributions and approximations thereto. Since the update schedule we employ abides by the assumption specified by Geman and Geman (see Section 3.3.1), these sampled values are considered to be values from the posterior distribution. This simply states that it is reasonable to assume that the algorithm converges after some moderate number of iterations. We use a variety of convergence diagnostics in our examples to ensure that the sampler indeed converges and to monitor *when* the sampler converges.

Chapter 4

EXAMPLES

The USFS project which motivated our work is still in progress. The data are not available so we are unable to use that real data. We test the model on simulated examples.

We use simulated data to evaluate the performance of our methodology. This simulated data is treated as the truth which allows us to determine whether our model produces accurate results. The model is evaluated on how well it reproduces the truth using various sets of input data. The input data consist of observed presence/absence information at a sample of sites and covariate information at all sites in the area of interest. The information contained in the observed presence/absence is varied by using different sampling plans and sample sizes. The information contained in the covariate is varied to determine how well the model discriminates among sites with good habitat but no true presence, good habitat and true presence, poor habitat but species presence, and poor habitat with species absence.

We compare the predicted probability of presence results from the autologistic model with covariates for sample data to results obtained from two simpler models. The logistic regression model assumes independence of species presence/absence among all sites. The basic autologistic model uses spatial correlation but does not use covariates related to species presence. Thus we determine whether our model improves upon existing models. We introduce the setup of the simulations,

present the example results from the three models, discuss and present convergence diagnostics, and finally present the results of a sensitivity analysis.

4.1 Simulation Setup

To setup the simulations we produce a true scene, determine covariate values for the sites, and then specify a sample plan which is used to obtain the observed data. We use a 40×40 lattice for a total of $N = 1600$ sites in the area of interest for the simulations. The true scene was generated by placing patches of presence over our area of interest. We did this to attain a picture which had separated clusters and clusters that seemed to be joined, and also covered the area of interest.

We limit our simulations setup to utilize one covariate which has high covariate values corresponding to 'good' habitat for the species. The covariate is subjectively formulated in three information patterns: one which has good habitat only in areas of species presence, and two confusing patterns which have either good habitat in areas of species absence or poor habitat in areas of species presence. For each information pattern the covariate is generated from two normal distributions each with mean 0. The normal distribution which corresponds to good habitat sites has variance 1 and the normal distribution corresponding to poor habitat sites has variance 5.

We use two sampling plans at two levels of coverage and each at two sampling intensities to obtain a variety of observed presence/absence images. The sampling plans we use are simple random sampling and systematic grid sampling of clusters. The systematic grid sample of clusters starts with a one random start systematic sample and at each selected site we also sample the neighborhood which is the eight surrounding sites ("Observations = Y" in Figure 4.1). The two levels of sample coverage are either over the whole area of interest or over only a portion

("Observations = Y" in Figures 4.1 and 4.4). We use two sample sizes corresponding to either a large or small sample for each coverage and sample plan combination. The sample size, which varies for the different sampling plans, will be presented for each example.

4.2 Gibbs Sampler Set-up

The estimation procedure which was described in Chapter 3 requires hyperprior distribution specification. Hyperparameters are chosen using pilot runs of the Gibbs-Hastings sampler. For the spatial parameter, β, we specify a broad gamma hyperprior distribution, $\Gamma(\omega, \alpha)$, with $(\omega, \alpha) = (2, 1.5)$. This distribution has most of its mass between 0 and 3. For each covariate parameter, θ_j, we specify a vague normal hyperprior distribution, $N(0, 10^2)$. These hyperprior distributions are shown in the middle row of Figure 4.14. Sensitivity analyses are described in Section 4.6.

We assume a second order neighborhood (Figure 2.1) for the locally dependent Markov random field (LDMRF) spatial structure. We also tried using a first order neighborhood but found the results to be less consistent, especially for small sample sizes.

For all simulations we run the Gibbs-Hastings sampler for 10,000 iterations, discarding the first 2000 iterations as burn-in. To obtain an independent sample for estimating parameters we use every other iteration value. These values are set using recommendations from Gibbsit (Raftery and Lewis, 1992), a program for identifying proper parameters for Gibbs sampling convergence, which will be discussed in Section 4.5.

4.3 Performance Evaluation

Three simulation examples follow. Nine scenes are given for each simulation: the true scene, the observed scene, the covariate, the prediction scenes from the logistic regression model, the autologistic model, and the autologistic model with covariates for sample data, and difference scenes described below. In the true scene black sites indicate presence. In the observed scenes unsampled sites are light grey, sampled sites where species are not present are medium grey, and sampled sites where species are present and hence observed are black. We assume perfect detection in obtaining observed presence/absence data for these simulations. In the covariate scenes higher covariate values are darker in color which corresponds to better habitat for the species. For prediction scenes darker colors indicate that the probability or posterior probability of presence is closer to one, or the species is more likely to be present.

We use several methods to evaluate performance. The difference of scenes are the predicted probabilities ($\hat{p}$) minus the true presence/absence (x) for each site. These plots highlight where the models are having difficulty. Light colors indicate underestimation of predicted probability of presence and dark colors indicate overestimation of predicted probability of presence. Shades corresponding to zero in the legend indicate correct prediction.

The second measure of performance is sensitivity and specificity. Sensitivity is the proportion of occupied sites that are predicted to contain the species. Specificity is the proportion of unoccupied sites which are predicted to be absent of the species. If both sensitivity and specificity equal 1, the model produces perfect predictions. Mathematically, these are formulated as

$$\begin{aligned} \text{Sensitivity} &= \frac{\sum \hat{x}_i I_{[x_i=1]}}{\sum_{i=1}^{N} x_i}. \\ \text{Specificity} &= \frac{\sum (1-\hat{x}_i) I_{[x_i=0]}}{\sum_{i=1}^{N} (1-x_i)}. \end{aligned} \tag{4.1}$$

We predict presence, $\hat{x}_i = 1$, if $\hat{p}_i > 0.5$. This cutoff value of 0.5 is not arbitrary; since we know the truth we investigate cutoff values which maximize the number of correctly classified sites. For our examples, this maximum is obtained for a cutoff in the range from 0.4 to 0.6. The sensitivity and specificity results are given in Table 4.1.

We also compare performance of the models by considering the predicted probability of presence for the two groups, the true presence sites and the true absence sites. A histogram of predicted probability of presence for the absence sites should be concentrated around zero with few high predicted probabilities of presence. A histogram of predicted probability of presence for presence sites should be concentrated around one with few small predicted probabilities of presence. We present these two histograms for each model for each example. Since there are many more absence sites we use two different scales for the histograms of presence sites and absence sites. The histograms are not directly comparable across the groups of presence/absence sites, but are comparable across models within groups of presence/absence sites.

We present histograms and smoothed distributions of the posterior distributions for the model coefficients in Figures 4.2, 4.5, and 4.8. We summarize estimated values and give 95% confidence intervals for the logistic regression parameters, and 95% highest posterior density intervals for the autologistic model and autologistic model with covariates for sample data in Tables 4.2, 4.3, and 4.4.

Table 4.1: Sensitivity and Specificity of Models for All Examples

	Example 1		Example 2		Example 3	
	Sens.	Spec.	Sens.	Spec.	Sens.	Spec.
Logistic Regression	.76	.92	.13	.97	.11	.98
Autologistic	.43	.91	.50	.78	.66	.98
Autologistic with Covariate	.99	.96	.79	.97	.66	.98

Table 4.2: Coefficient Estimates and 95% Intervals for Example 1 from Logistic Regression(LR), Basic Autologistic Model(AL), and Autologistic Model with Covariates for Sample Data(AMCS).

Model	$\hat{\beta}$	95% Interval	$\hat{\theta}_0$	95% Interval	$\hat{\theta}_1$	95% Interval
LR	-	-	-11.20	(-13.50, -8.90)	2.52	(0.22, 4.82)
AL	1.90	(0.61, 3.20)	-	-	-	-
AMCS	2.68	(0.56, 6.52)	-16.75	(-40.89, -3.10)	1.86	(-0.40, 5.83)

4.4 Simulation Results

4.4.1 Example 1

The first example uses observations from a small systematic grid sample of clusters over the top portion of the area of interest (Figure 4.1). The sample size is 54, a 3.4% sample, with 12 observed presence sites. The covariate has reliable information: good habitat occurs where the species is actually present.

The predictions from the three models are given in Figure 4.1. The logistic regression model mirrors the covariate map as expected when only one explanatory variable is included. The basic autologistic model has trouble with this small amount of observed presence/absence information. The autologistic model with covariates for sample data reproduces the true image well.

The bottom scenes, which are the differences between predicted probability ($\hat{p}$) and true presence/absence (x) show that the logistic regression has a tendency to

underestimate presence and the autologistic model with covariates for sample data has a tendency to overestimate presence around the perimeter of clusters.

The sensitivity and specificity (Table 4.1) for the autologistic model with covariates for sample data are 0.99 and 0.96, respectively indicating almost perfect prediction. The logistic regression model also has high specificity, 0.92, but considerably lower sensitivity, 0.76, due to the tendency to predict moderate probabilities of presence. The autologistic model with covariates for sample data is more discriminating than the logistic regression since it forms groups of sites with high probability of presence. The performance measures for the autologistic model (sensitivity = 0.43 and specificity = 0.91) reflect the poor prediction we see in Figure 4.1.

The coefficient estimates from all three models (Table 4.2) all have overlapping confidence/probability intervals, where applicable. Thus, although the models obtained from the methods do not have different coefficient estimates, the presence of all coefficients in the model certainly produces a better prediction of presence/absence over the area, as we discussed above. The posterior distributions of the coefficient estimates are shown in Figure 4.2.

The histograms of predicted probability of presence for groups of absence and presence sites are shown in Figure 4.3. The logistic regression model tends to predict absent sites as having lower predicted probability of presence and present sites as having higher predicted probability of presence, but there is no distinct break in $\hat{p}$ between the present and absent sites. The histograms for the autologistic model with covariates for sample data show two distinct groups based on predicted probability of presence, where true presence sites all had $\hat{p} > 0.4$ and over 95% of true absence sites had $\hat{p} < 0.4$.

Table 4.3: Coefficient Estimates and 95% Intervals for Example 2 from Logistic Regression(LR), Basic Autologistic Model(AL), and Autologistic Model with Covariates for Sample Data(AMCS).

Model	$\hat{\beta}$	95% Interval	$\hat{\theta}_0$	95% Interval	$\hat{\theta}_1$	95% Interval
LR	-	-	-7.20	(-10.30, -4.10)	1.20	(-1.90, 4.30)
AL	2.24	(0.70, 3.70)	-	-	-	-
AMCS	2.10	(1.19, 3.51)	-20.90	(-30.80, -9.71)	2.63	(0.74, 4.28)

4.4.2 Example 2

In the second example the systematic grid sample of clusters covers the entire area of interest (Figure 4.4). The sample size is 81, a 5% sample, with observed presence at 13 sites. The covariate gives confusing information as the range of good habitat is larger than the actual range of the species.

The prediction scenes for Example 2 are given in Figure 4.4. The logistic regression model mirrors the covariate, which in this case gives poor predictions in areas of good habitat even with species absence. The autologistic model requires more observed presence/absence data to produce reliable results. The predictions from the autologistic model with covariates for sample data are similar to those from the logistic regression except in the area of good habitat and species absence. The autologistic model with covariates for sample data is more discriminating about sites with good habitat; the model forms clusters within areas of good habitat.

The difference images show the tendency for the logistic regression model to produce moderate values for predicted probability of presence and hence underestimates true presence. The autologistic model with covariates for sample data overestimates presence on the perimeter of clusters of species presence and underestimates presence in the lower right cluster of species presence.

Table 4.4: Coefficient Estimates and 95% Intervals for Example 3 from Logistic Regression(LR), Basic Autologistic Model(AL), and Autologistic Model with Covariates for Sample Data(AMCS).

Model	$\hat{\beta}$	95% Interval	$\hat{\theta}_0$	95% Interval	$\hat{\theta}_1$	95% Interval
LR	-	-	-5.70	(-7.40, -4.00)	0.95	(-0.75, 2.65)
AL	0.73	(0.48, 0.98)	-	-	-	-
AMCS	1.60	(0.82, 2.60)	-5.20	(-8.90, -1.70)	-0.20	(-0.90, 0.50)

The autologistic model with covariates for sample data has sensitivity of 0.79, and specificity of 0.97 which indicates good correspondence between predictions and the truth (Table 4.1). For the logistic regression model sensitivity is 0.13 and specificity is 0.97. High specificity for all models is expected due to the large number of absence sites. The measures for the autologistic model (sensitivity = 0.5 and specificity = 0.78) again correspond to the poor image reproduction.

The coefficient estimates display the same properties as in example 1, the 95% intervals overlap (Table 4.3, and the distributions from the autologistic model with covariates for sample data have reasonable spread (Figure 4.5).

The histograms of predicted probability of presence are displayed in Figure 4.6. The logistic regression results are poor due to the moderate predicted probability of presence. The autologistic model with covariates for sample data shows some overlap between present and absent sites as compared to example 1, but the predictions generally fall into two correctly classified groups.

4.4.3 Example 3

Our third example uses observations from a large simple random sample with coverage over the top portion of the area of interest. The sample size is 180, an 11.25% sample, with 31 sites of observed presence. The covariate provides confusing information; the range of good habitat is larger than the actual range of the species.

The results from this example are in Figure 4.7. The logistic regression mirrors the covariate values which do not match up with true presence. The autologistic model performs well with this large quantity of information except in the area where no sample data was obtained. The autologistic model with covariates for sample data does a good job for the sampled area, similar to the autologistic model, but shows a slight increase in probability of presence for the area of good habitat in the non-sampled sites in the lower half. This is a clear example of the autologistic model with covariates for sample data utilizing the best of both the autologistic model and the logistic regression models.

The difference images again show the moderate predictions from the logistic regression model. The difference images for the autologistic model and autologistic model with covariates for sample data appear to be quite similar since the increase in predicted probability of presence from the autologistic model with covariates for sample data in the bottom half is very slight.

The sensitivity and specificity from the logistic regression model are similar to the results from the previous example since the covariate is the same (Table 4.1). The sensitivity and specificity performance of the autologistic and autologistic model with covariates for sample data are identical. This equality, even in the face of a hint of presence in the lower right corner, is due to the classification of predicted presence using $\hat{p}_i > .5$. A more discriminating rule for predicted probability scenes with this hint of presence might be used to include the lower right cluster in the predicted presence classification.

The coefficient estimates again overlap, but show more dissimilarity than the other examples (Table 4.4). The distributions of the estimates from the autologistic model with covariates for sample data are fairly Gaussian with reasonable spread.

The histograms of predicted probability of presence are displayed in Figure 4.9. The logistic regression results are again fairly poor since they form no definite groups. The autologistic model histograms display the 'missed' cluster as the low predicted probability of presence for the true presence sites ($x = 1$). Otherwise the predictions for the true presence sites are relatively uniform. The autologistic model with covariates for sample data results show fewer present sites with low predicted probability of presence than the autologistic model but still display the missed cluster.

4.5 Convergence

The Gibbs-Hastings sampler is an iterative estimation procedure. The results converge to the appropriate estimates. Besides setting up the algorithm, we must monitor and judge convergence. There are three issues to the number of iterations that are necessary for convergence. The first is how many iterations are necessary until the procedure is sampling from the appropriate distributions. The second is to determine if all samples are independent and if not, how many in-between samples should be dropped to ensure independence of the remaining. The third number is how many saved samples are necessary to produce valid estimates. We investigate these in the following sections.

4.5.1 Burn-in and Number of Iterations

We use the program Gibbsit (Raftery and Lewis, 1992) to determine run lengths for the Gibbs-Hastings sampler. We introduce the sampler parameters, describe the Gibbsit program, and present the results.

The number of iterations used in the Gibbs-Hastings sampler is denoted by M. The burn-in period is the number of initial iterations which are not used in estimation. The Gibbs-Hastings sampler provides values from the proper distribution

asymptotically so we disregard the first iterations until we can assume the Markov chain has converged to the appropriate stationary distribution. We define K to be the step number: after burn-in we use only every K^{th} iteration value in order to obtain an independent sample for estimating parameters.

The program Gibbsit uses output from an initial short run of a Gibbs sampler to determine the necessary sampler parameters to achieve specified precision on the estimated parameters. We specify the recommended precision bound of .01 on a 95% confidence interval for the parameters $\underline{x}$, β, and $\underline{\theta}$.

The results we obtained from Gibbsit are as follows. For the simulation examples, slightly conservative values for the sampler parameters are $M = 10,000$, burn-in= 2000, and $K = 2$.

4.5.2 Convergence Diagnostics

The sampler parameters are not guaranteed to provide convergence so we also use several convergence diagnostics to monitor convergence of the Gibbs-Hastings sampler. One convergence diagnostic method is to monitor sample paths of the parameter values. The sample values should settle into a random variation around some central value. This procedure is straightforward for β, θ_0, and θ_1, but is not feasible for site values, either $\underline{x}$, or $\underline{\hat{p}}$, since there are 1600 values of each parameter for every iteration.

Since monitoring each site is unrealistic we consider a function of the site values which we call the C-index. The C-index measures the number of sites whose difference between predicted probability of presence at iteration t and iteration $(t-1)$ differs by more than 0.05. Thus the C-index measures the number of sites whose predicted probability of presence is changing even moderately. The C-index at the

t^{th} iteration, C-index$_t$, is defined

$$\text{C-index}_t = \sum_{i=1}^{N} \mathrm{I}\left[\left|\, \hat{p}_{i,t-1} - \hat{p}_{i,t} \,\right| > 0.05\right].$$

We diagnose convergence by monitoring the sample path of the C-index$_t$ as well as β, θ_0, and θ_1.

Figure 4.10, Figure 4.11, and Figure 4.12 show the sample path convergence diagnostics for Examples 1 - 3. The top plots are sample paths for β, θ_0, and θ_1, respectively. There are 4000 values shown because M=10,000, burn-in =2000, and only every other value was retained for estimation purposes. The bottom plot is the sample path for the index on $\underline{\hat{p}}$, C-index$_t$. The C-index is shown for all 10,000 iterations. These plots show, as expected, that the parameters are correlated. There is some question as to whether the sample paths for the parameters have converged and longer runs may be needed for improved parameter estimates. However, the main interest is in the maps of mean predicted probability of presence.

To investigate convergence of the scene, we also monitor the mean predicted probability of presence map for sequences of 400 iterations. We use this method for all examples but display the resulting sample path of images for Example 2 in Figure 4.13. These plots indicate convergence of the predicted probability of presence maps.

4.6 Sensitivity Analysis

The estimation procedure for the autologistic model with covariates for sample data may be sensitive to hyperparameter values chosen for β, θ_0, and θ_1. The hyperparameter values for β are the gamma distribution parameters shape = ψ and scale = α. The hyperparameter value for each θ_k is τ_k^2 the variance for the normal

distribution. We performed a sensitivity analysis to determine if the estimation procedure is robust to these specifications using the setup of Example 2 in Figure 4.4. The regular hyperprior distributions we used in the examples presented above and also the hyperprior distributions we consider for the sensitivity analysis are shown in Figure 4.14. The first row of densities represent narrow hyperprior distributions.

$$\begin{aligned} \pi(\beta) &= \text{Gamma}(1,2). \\ \pi(\theta_k) &= \text{N}\left(0,3^2\right). \end{aligned} \tag{4.2}$$

for $k = 0,1$. The second row of densities are the hyperprior distributions used in the examples.

$$\begin{aligned} \pi(\beta) &= \text{Gamma}(2,1.5). \\ \pi(\theta_k) &= \text{N}\left(0,10^2\right). \end{aligned} \tag{4.3}$$

for $k = 0,1$. The bottom row of the figure displays hyperprior distributions which are more broad.

$$\begin{aligned} \pi(\beta) &= \text{Gamma}(3,2). \\ \pi(\theta_k) &= \text{N}\left(0,20^2\right). \end{aligned} \tag{4.4}$$

for $k = 0,1$.

We do not consider a hyperprior distribution for β which allows values less than zero as clustering of species was an assumption of the analysis. We also do not consider different location hyperparameters for the θ_k's since we have no reason to choose a hyperprior distribution which is not vague in its location. We use the same starting values that we use in the examples since the chain is robust to these inputs. We run the sensitivity analyses using the same sampler parameters of burn-in =

2000 and N = 10,000. We use every second sampled value (K=2) for estimation. Based on monitoring of scenes of predicted probability of presence, we concluded that these sensitivity analyses had converged.

The initial comparison for sensitivity to hyperparameter specification is on the prediction scenes shown in Figure 4.15. The top image was produced using the narrow hyperprior distributions in Equation 4.2. The narrow hyperprior distributions constrained the sample values too much to allow proper posterior coverage. The posterior distributions are displayed in Figure 4.16. The posterior distributions from the narrow hyperprior distributions are narrow themselves.

The regular and broad hyperprior distributions result in very similar images and posterior distributions. This indicates that the procedure is robust to hyperparameter choice when the spread in the hyperprior distribution is broad. Thus hyperparameters should be chosen conservatively.

We further investigate the effect of hyperparameter specification, by comparing the estimates of the parameters and their posterior distributions. Table 4.5 displays the estimates and standard deviations for the coefficient parameters from three hyperparameter value specifications. Again, these results indicate that the narrow hyperprior distributions can impact the results. However, note that inferences based on all three set-ups will be similar. All methods show a positive spatial correlation between sites and a negative intercept. All methods also indicate a positive relationship between the covariate and probability of presence, however the parameter is not significantly different from zero for the narrow hyperprior distributions.

We include two measures used to compare the site parameters in the sensitivity analysis (Table 4.5). We consider the number of misclassified sites which is the sum of the absolute value of the difference between predicted presence/absence at

a site and the truth. Presence is predicted for sites with $\hat{p} > 0.5$. The number of misclassified sites for the regular and broad hyperprior distributions are similar at 219 and 211, while the narrow hyperprior distributions result in more misclassified sites, 336. Since comparison of all individual sites is not efficient we compare the predicted probability of presence for only the three sites pictured in Figure 4.17. Site # 1010 is a definite presence site, site #1298 is a site on the border of a cluster, so estimation is more difficult, and site #205 is a definite absence site. Note the similarity between the regular and broad hyperprior distributions and the comparatively poor predictions from the narrow hyperprior distribution.

4.7 Conclusions

The autologistic model with covariates for sample data provides improved prediction over the logistic regression and the basic autologistic model. The examples we present demonstrate the ability of the model to utilize both spatial information and covariate information to form reliable predictions based on limited observed presence/absence information. The estimation procedure we suggest provides reasonable posterior estimates using broad hyperprior distributional information.

Figure 4.1: Setup and Output Maps for Example 1

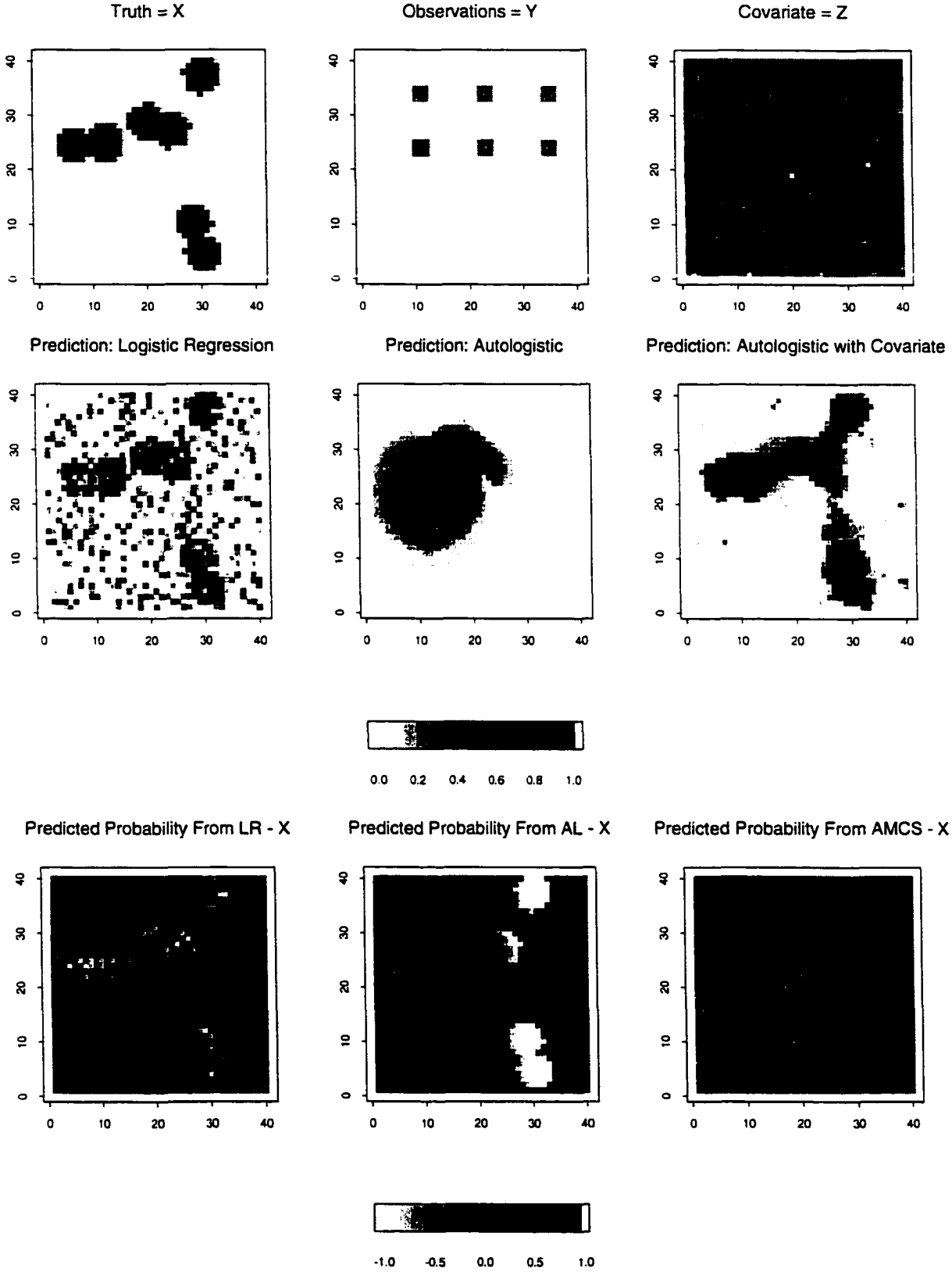


Figure 4.2: Posterior Distributions of Coefficients for Example 1 with Kernel Estimates

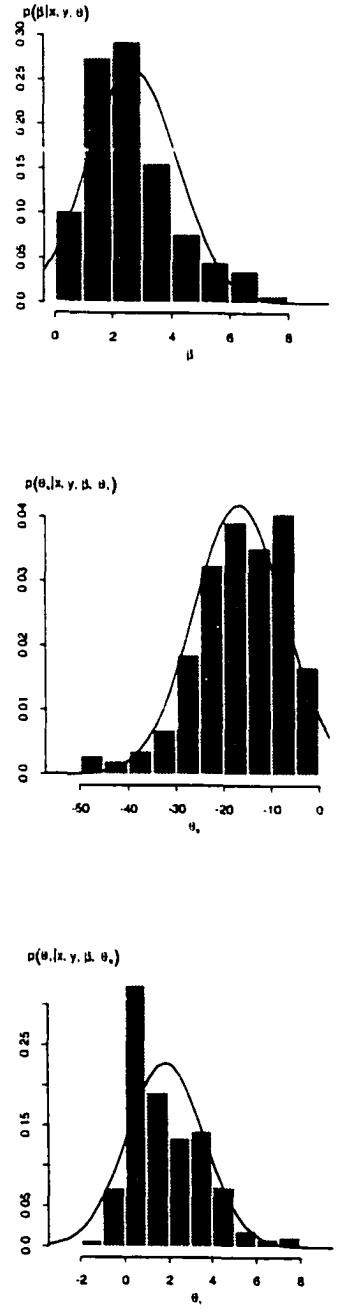


Figure 4.3: Histogram of Predicted Probability of Presence for Presence and Absence Sites Separately for Example 1

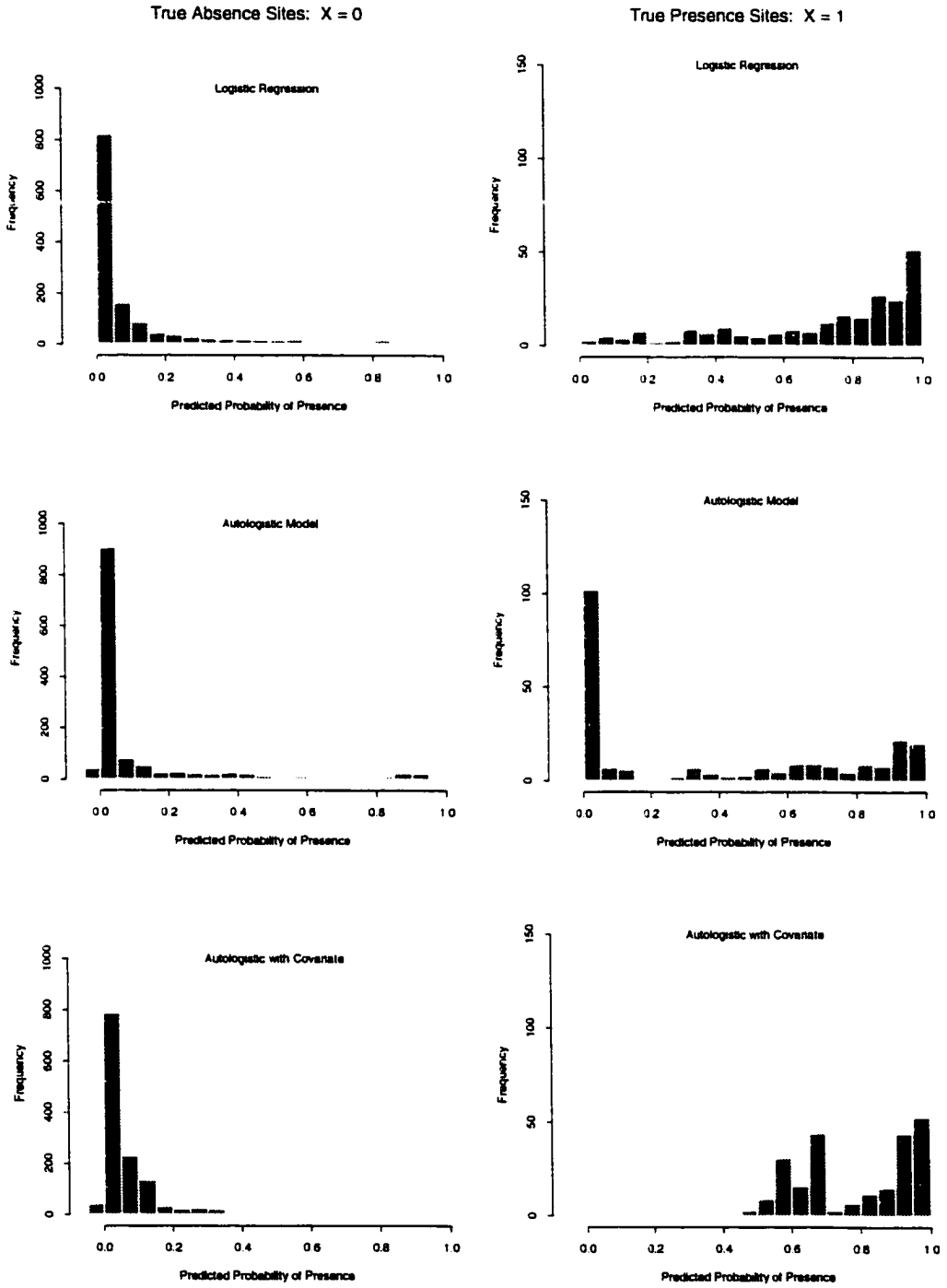


Figure 4.4: Setup and Output Maps for Example 2

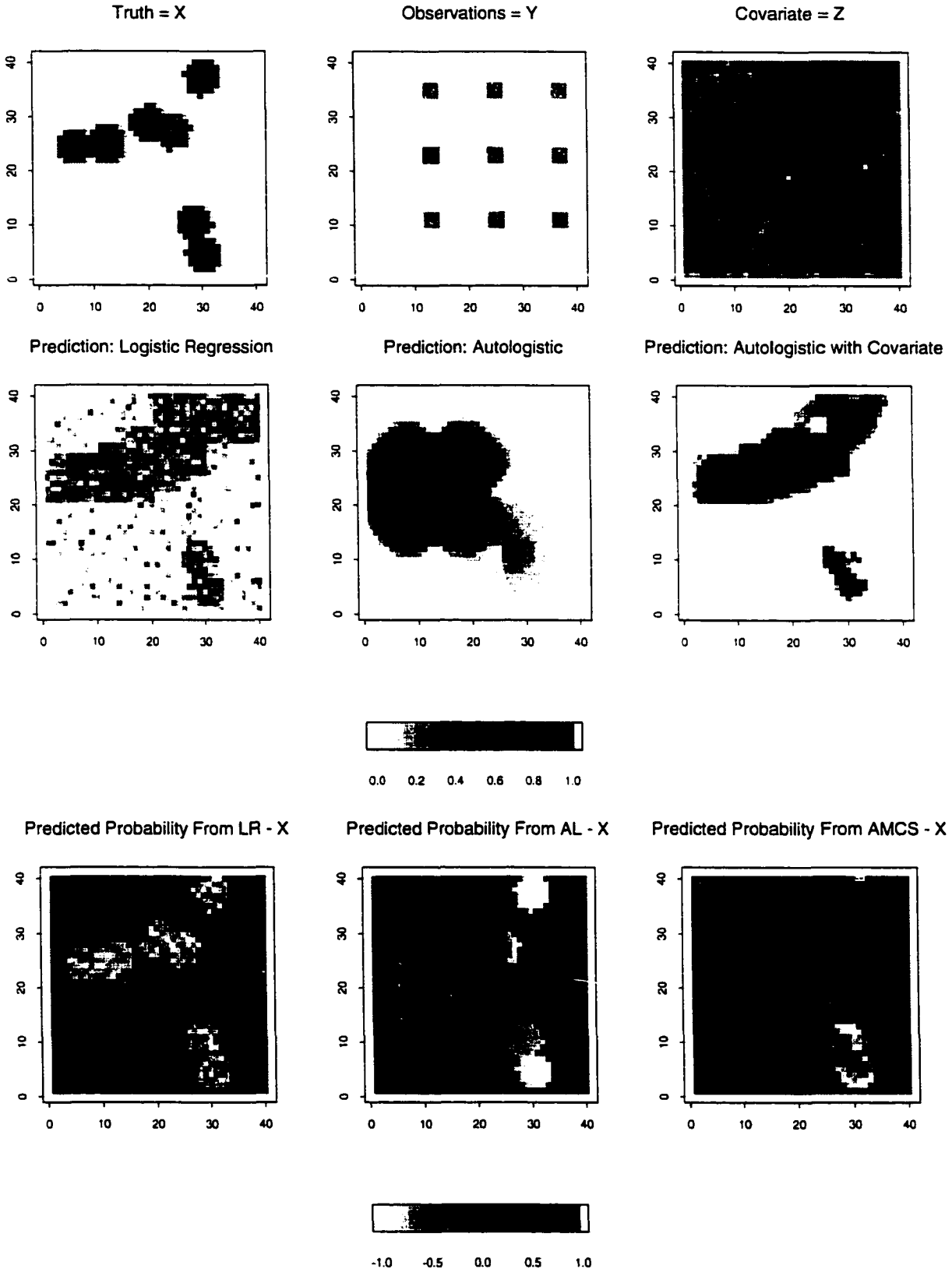


Figure 4.5: Posterior Distributions of Coefficients for Example 2 with Kernel Estimate

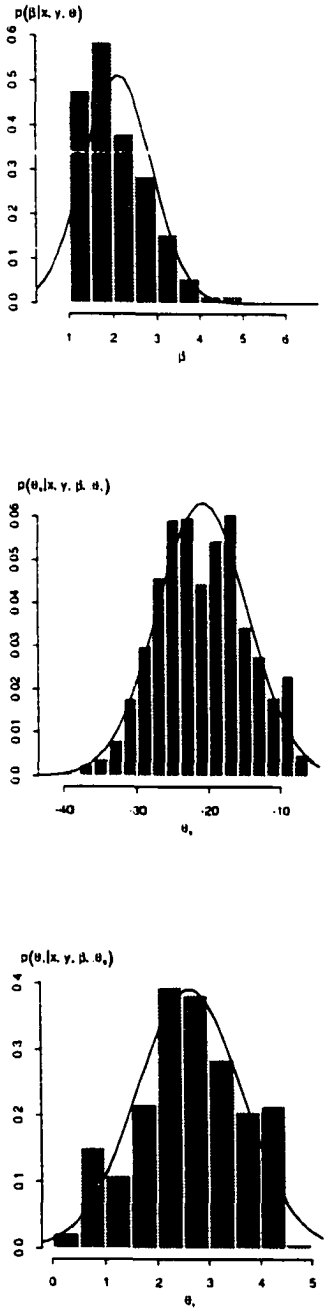


Figure 4.6: Histogram of Predicted Probability of Presence for Presence and Absence Sites Separately for Example 2

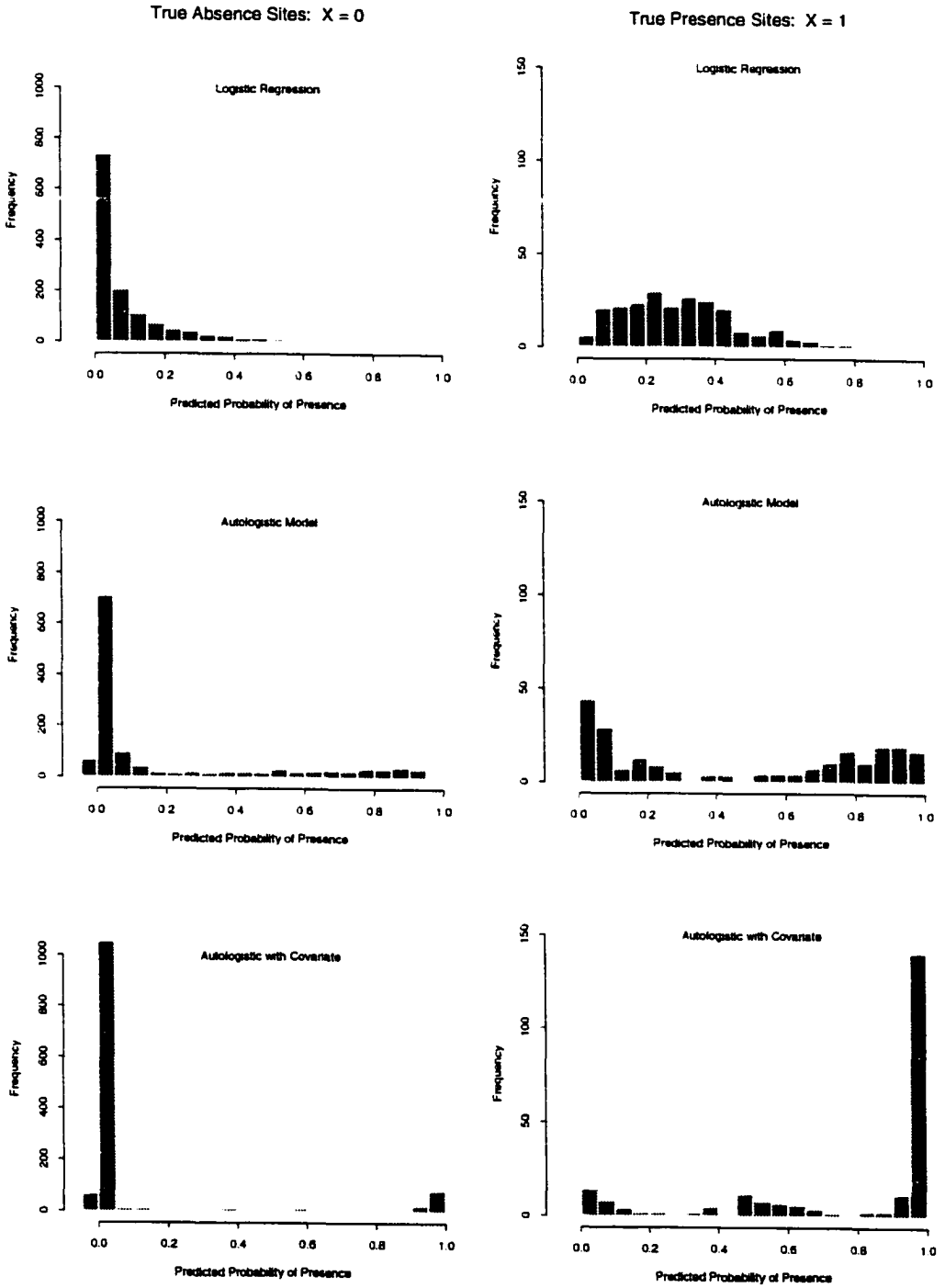


Figure 4.7: Setup and Output Maps for Example 3

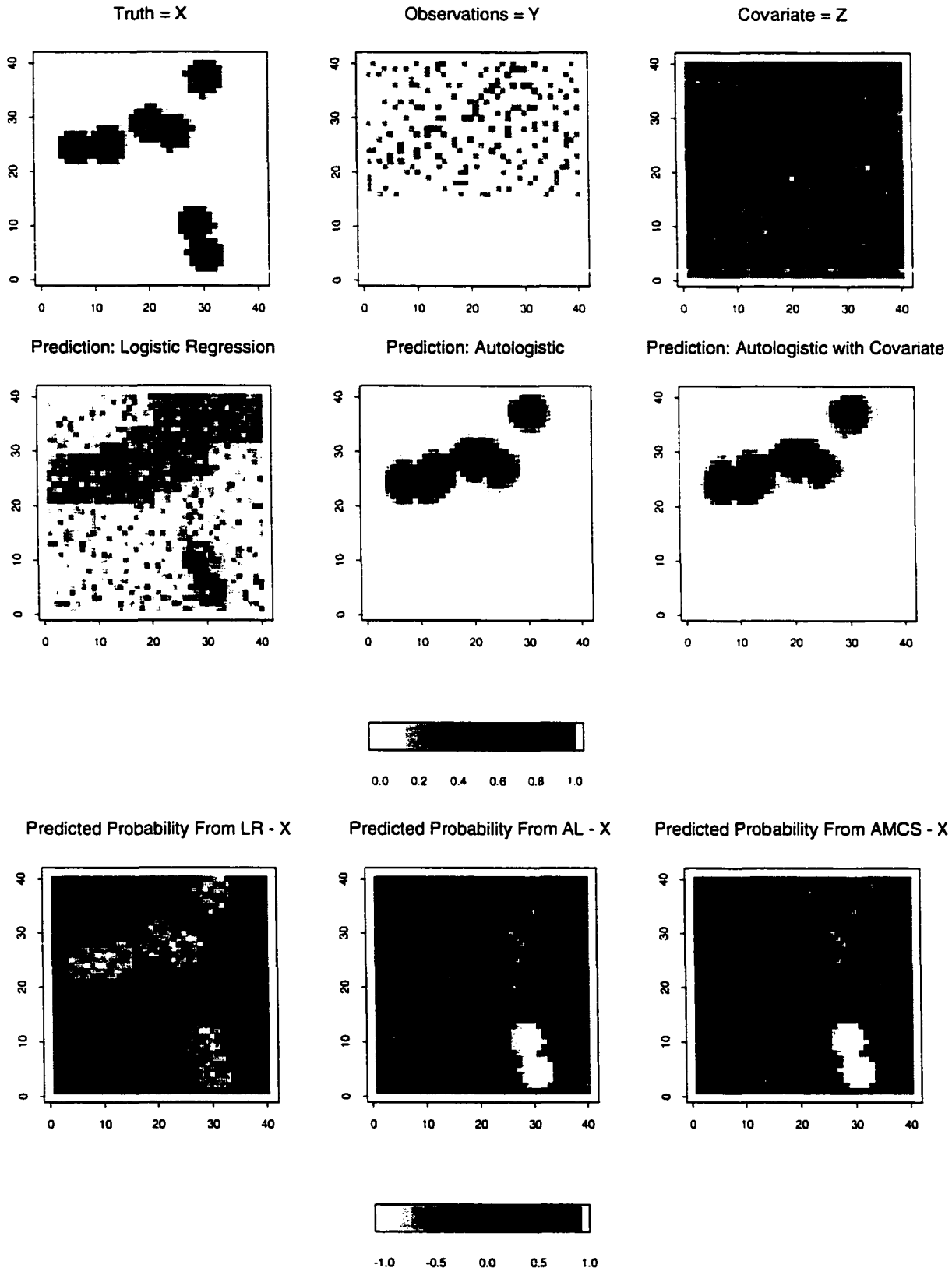


Figure 4.8: Posterior Distributions of Coefficients for Example 3 with Kernel Estimate

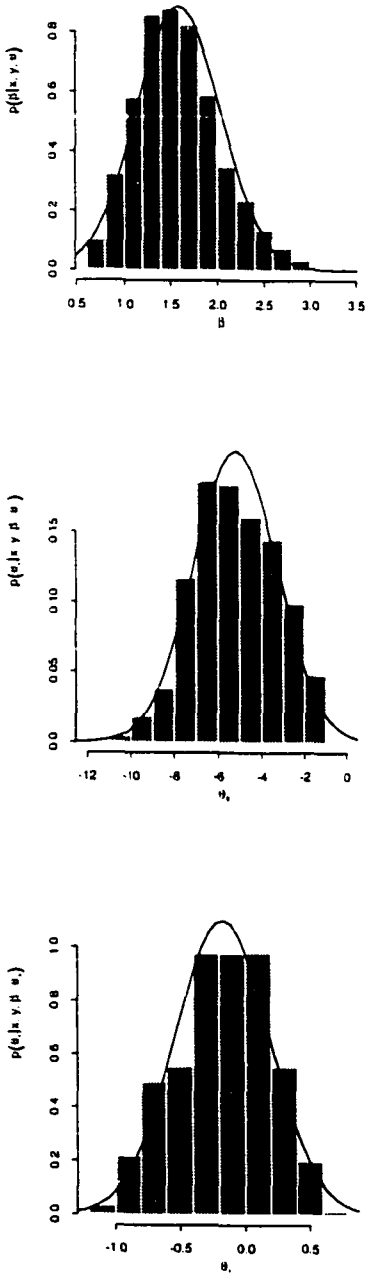


Figure 4.9: Histogram of Predicted Probability of Presence for Presence and Absence Sites Separately for Example 3

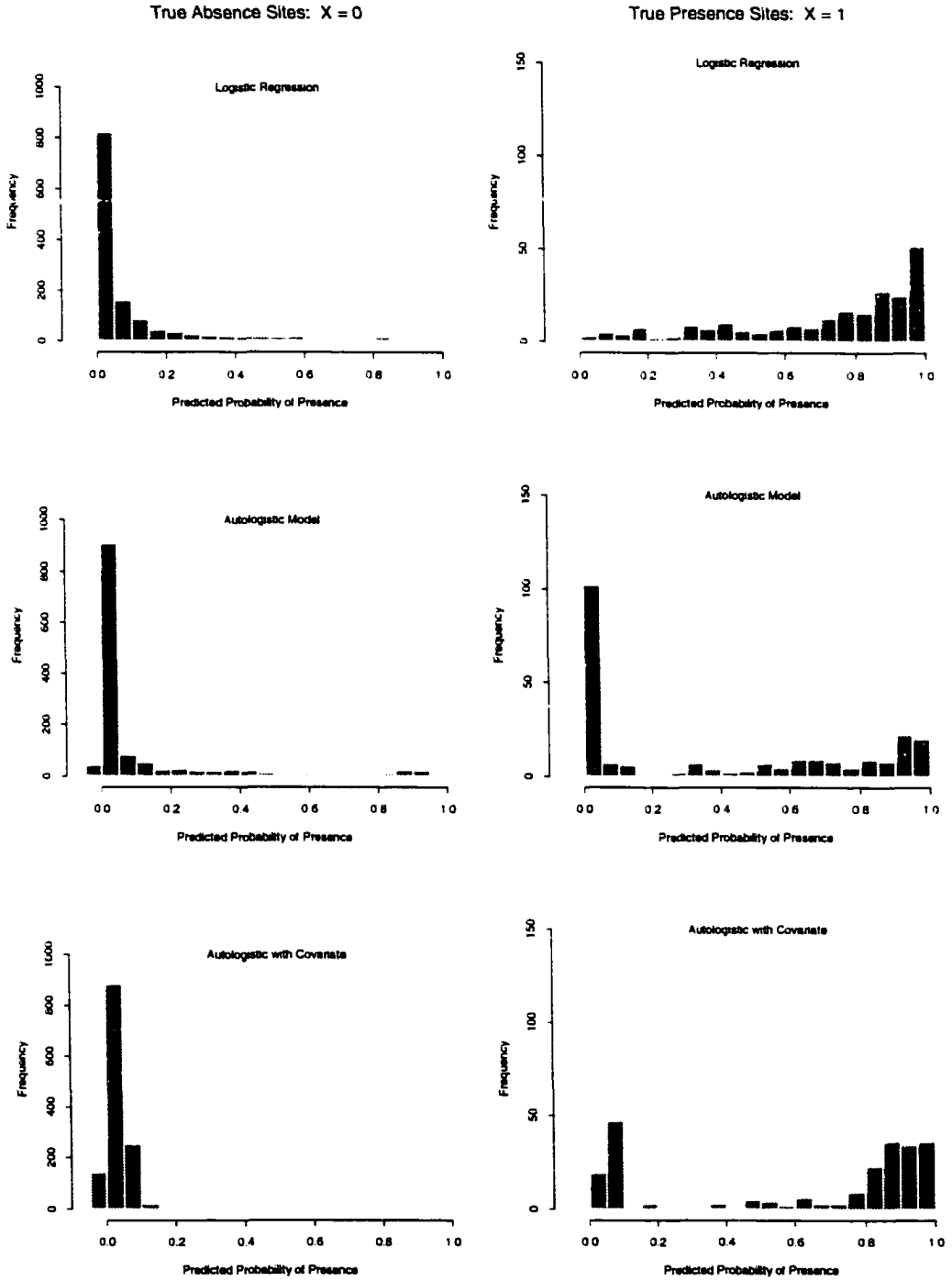


Figure 4.10: Convergence Diagnostics: Sample Paths for Example 1

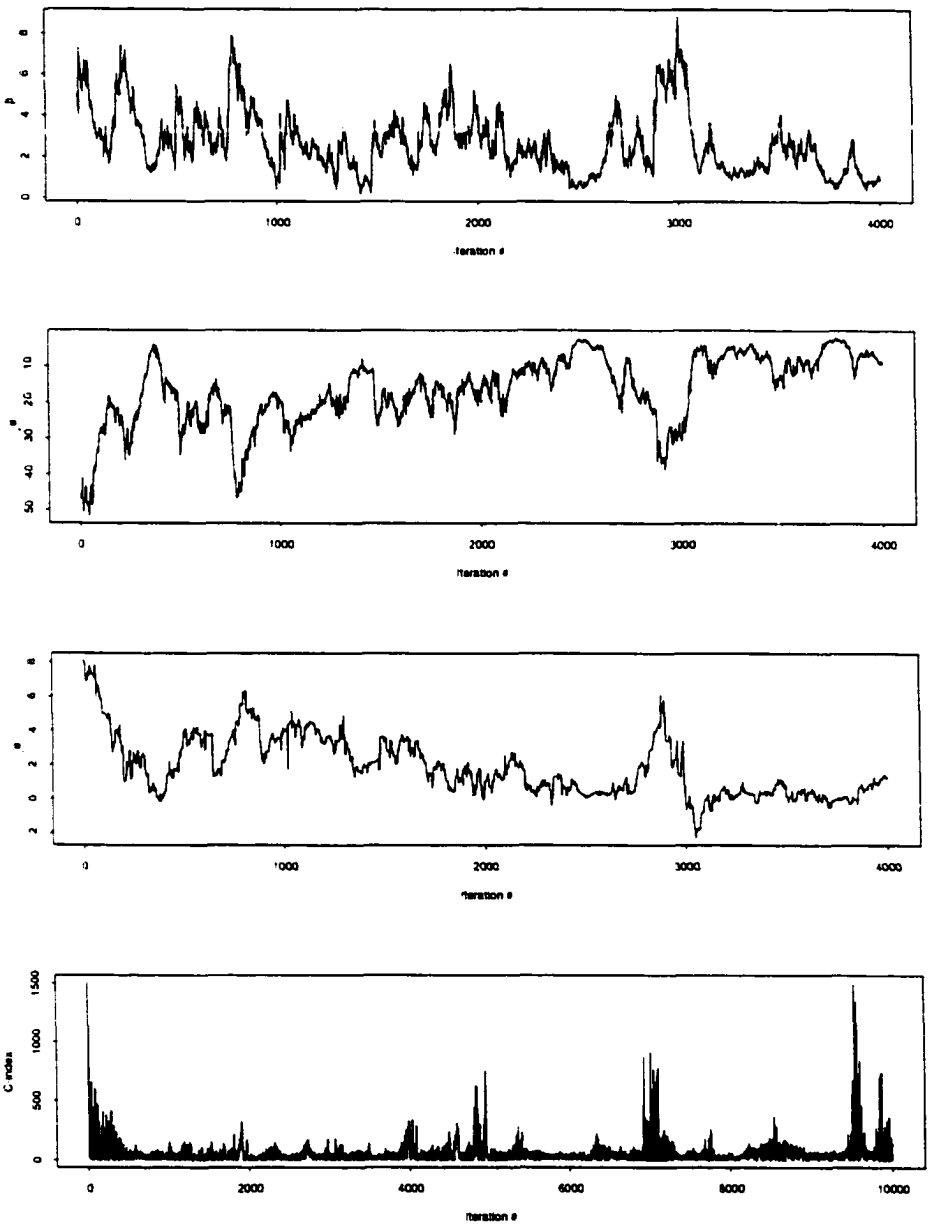


Figure 4.11: Convergence Diagnostics: Sample Paths for Example 2

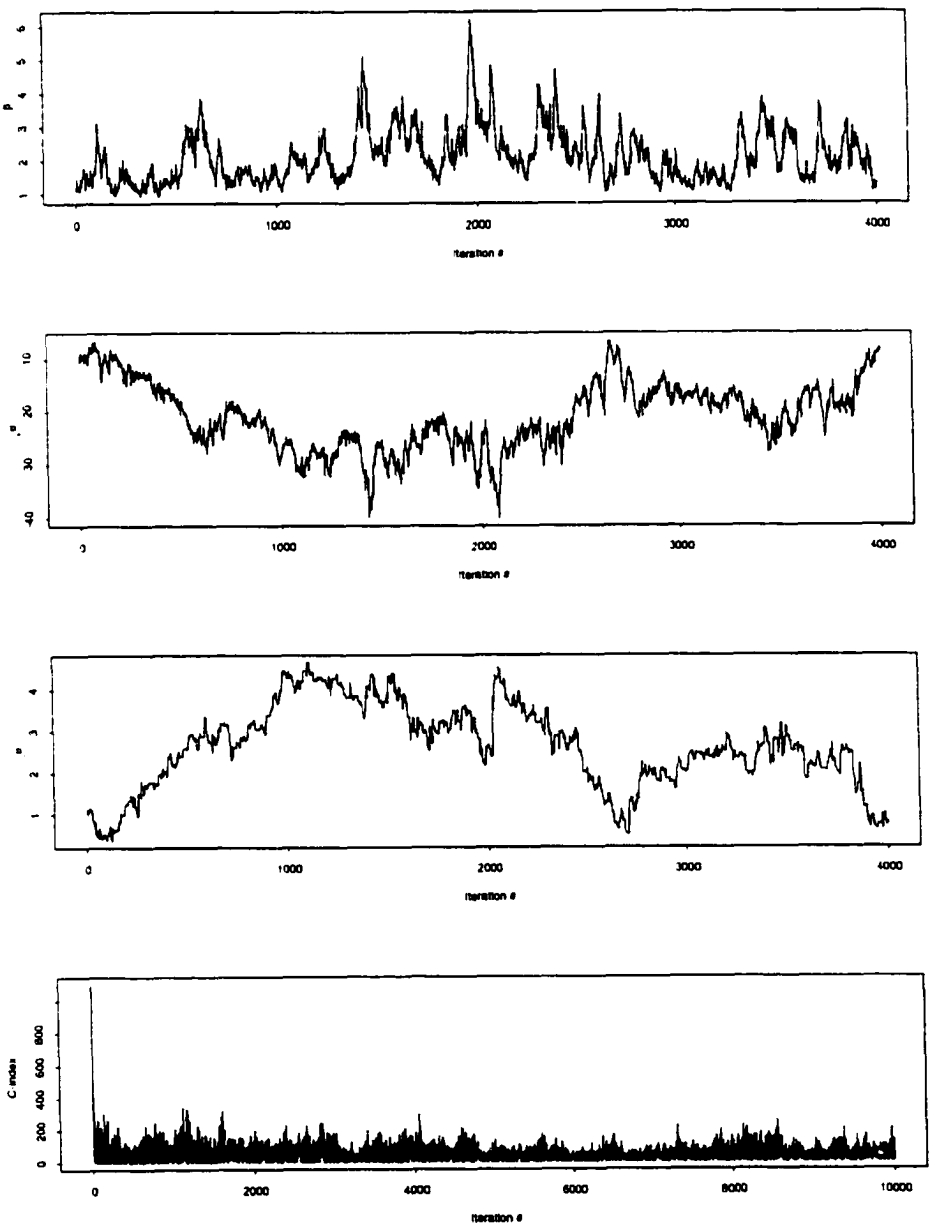


Figure 4.12: Convergence Diagnostics: Sample Paths for Example 3

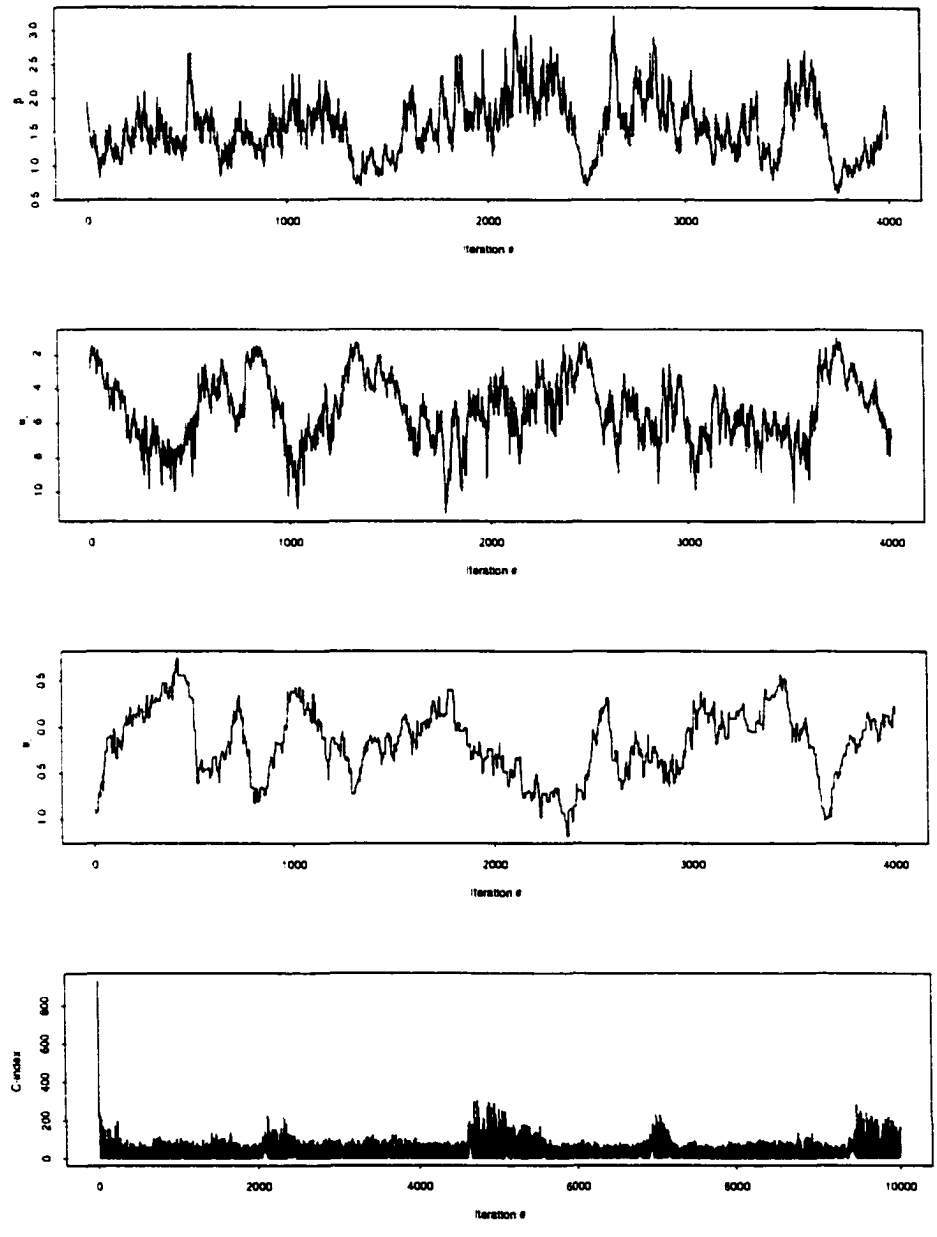
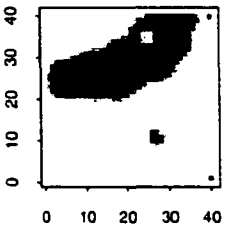
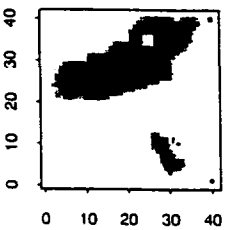


Figure 4.13: Convergence Diagnostics: Sequence of Images for Example 2

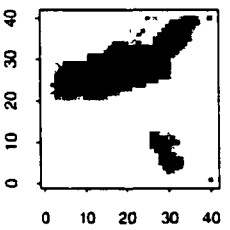
Iterations 2000 - 2800



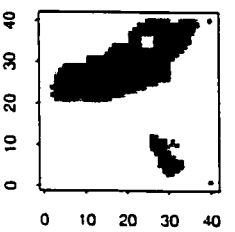
Iterations 2800 - 3600



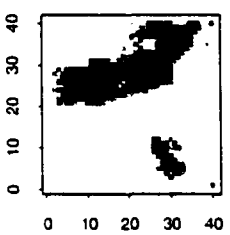
Iterations 3600 - 4400



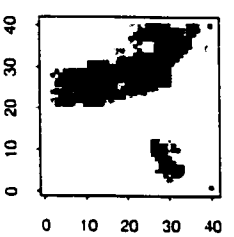
Iterations 4400 - 5200



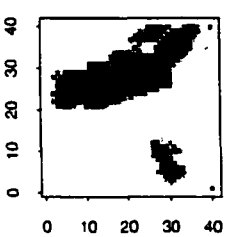
Iterations 5200 - 6000



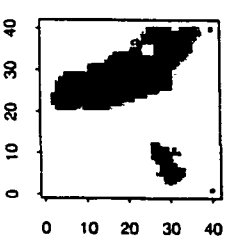
Iterations 6000 - 6800



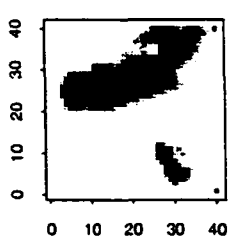
Iterations 6800 - 7600



Iterations 7600 - 8400



Iterations 8400 - 9200



Iterations 9200 - 10000

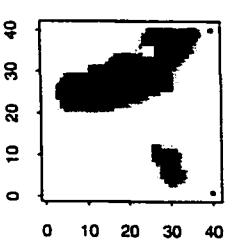


Figure 4.14: Hyperprior Distributions for Sensitivity Analysis

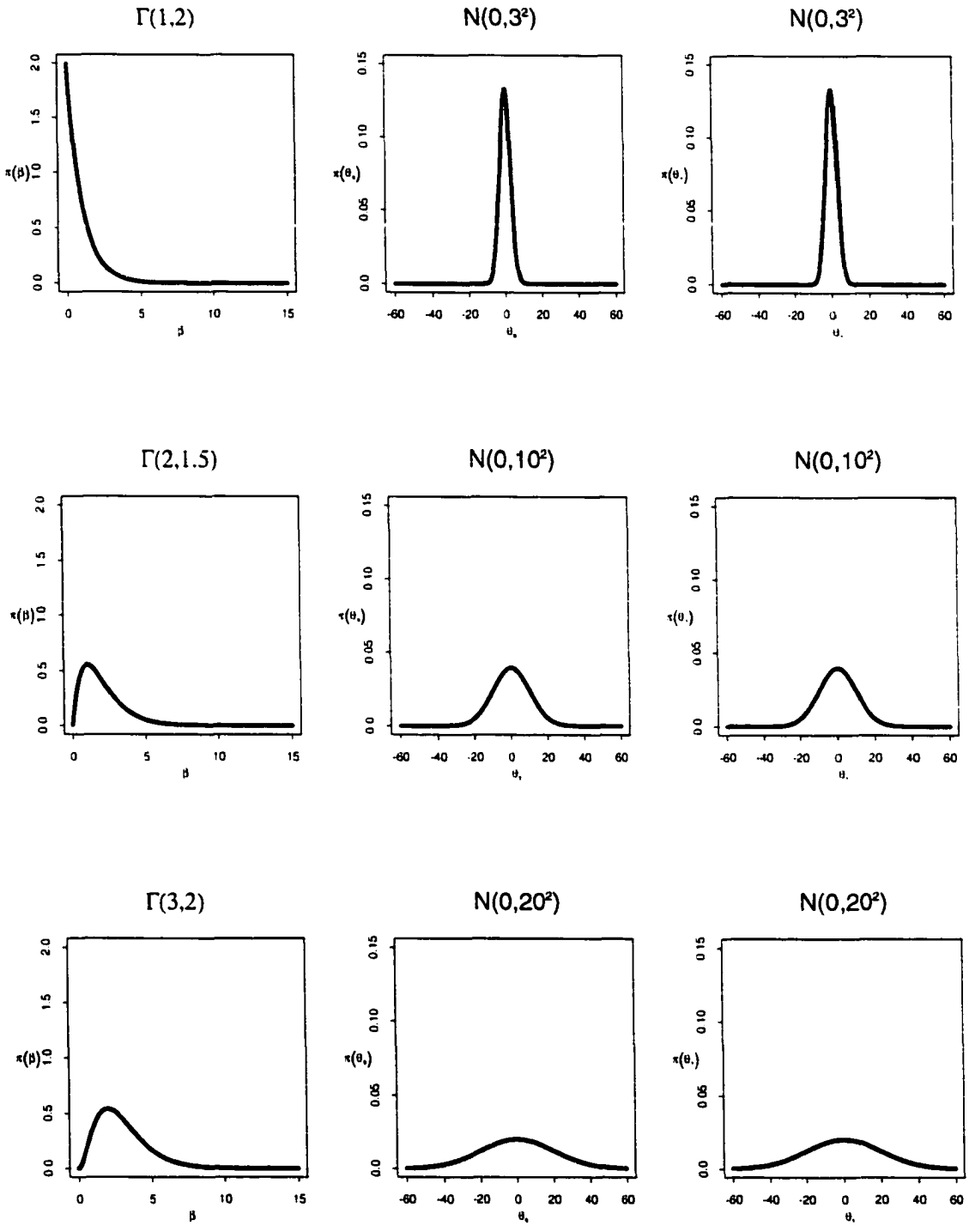


Figure 4.15: Output Maps from Sensitivity Analysis

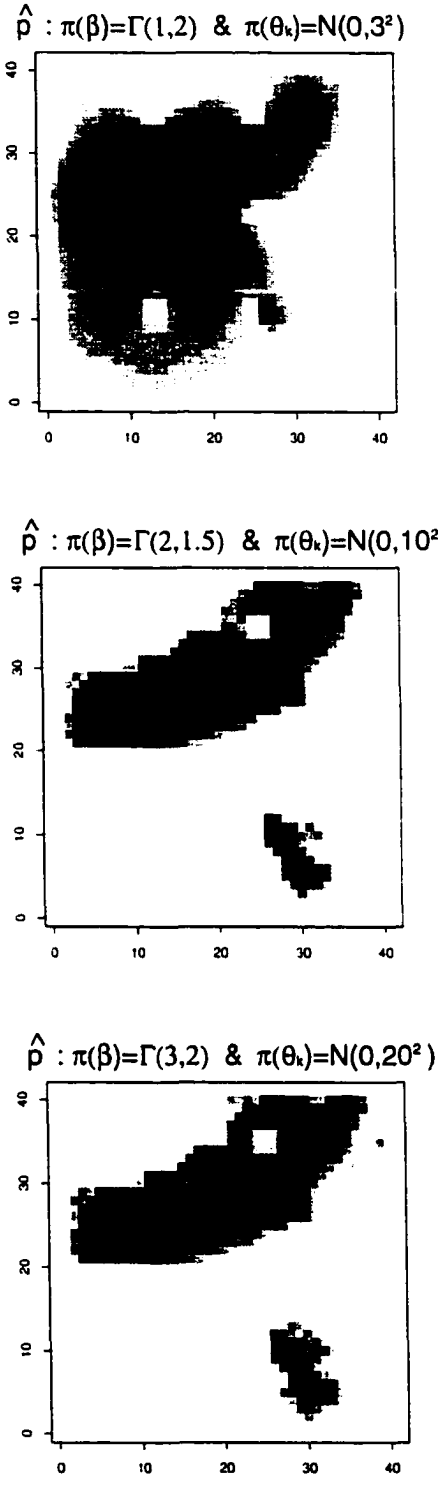


Figure 4.16: Posterior Distributions of Coefficients from Sensitivity Analysis with kernel estimate

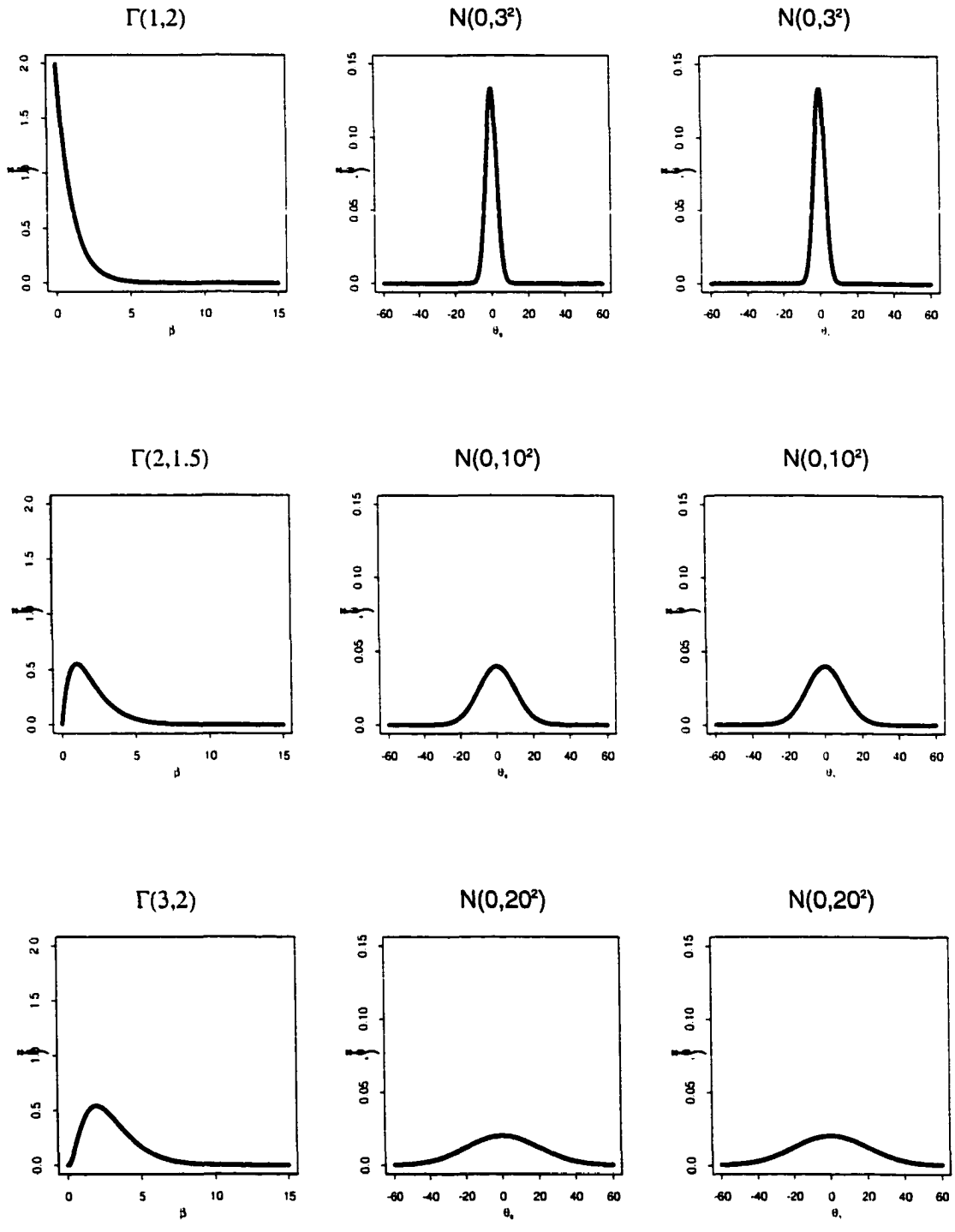
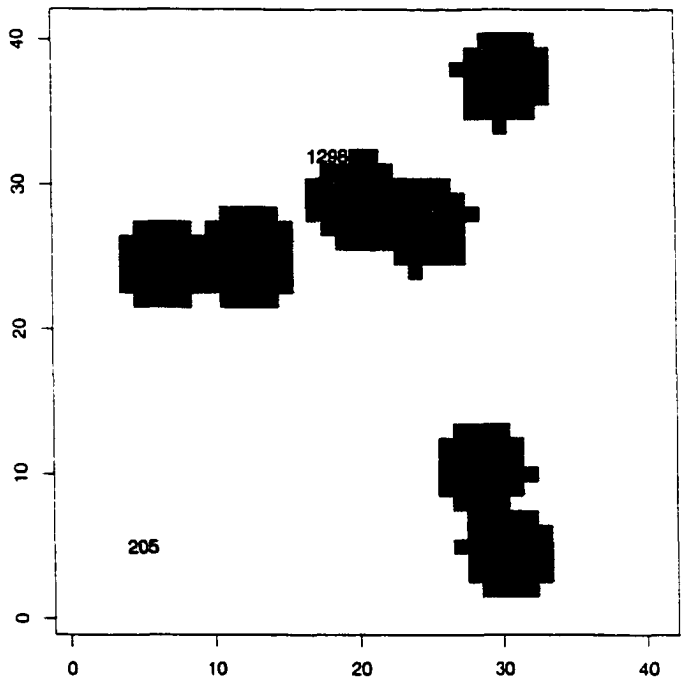


Table 4.5: Parameter Estimates from Sensitivity Analysis

$\pi(\beta)$	$\pi(\theta_k)$ $k = 0, 1$	$\hat{\beta}$ (s.d. $\hat{\beta}$)	$\hat{\theta}_0$ (s.d. $\hat{\theta}_0$)	$\hat{\theta}_1$ (s.d. $\hat{\theta}_1$)	$\sum_{i=1}^{N} \mid x_i - \hat{x}_i \mid$	$\hat{p}_{1010}$ $(x = 1)$	$\hat{p}_{1298}$ $(x = 0)$	$\hat{p}_{205}$ $(x = 0)$
$\Gamma(1,2)$	$N(0,3^2)$	1.20 (0.30)	-5.00 (1.70)	0.13 (0.29)	336	0.76	0.62	0.1400
$\Gamma(2,1.5)$	$N(0,10^2)$	2.10 (0.78)	-20.90 (6.30)	2.63 (1.02)	219	0.99	0.53	0.0003
$\Gamma(3,2)$	$N(0,20^2)$	1.94 (0.81)	-21.61 (6.46)	2.91 (1.16)	211	0.94	0.56	0.0003

Figure 4.17: Comparison Sites for Sensitivity Analysis



Chapter 5

ESTIMATION OF DETECTABILITY

We have assumed, thus far, that the observed presence/absence data are obtained without error. However, for many species the probability of detection, given presence, is less than one especially under various habitat, weather, and sampling conditions. In Chapter 3 we extended the autologistic model to account for sample data and included covariates related to species presence. We make a further extension to the autologistic model to estimate detectability for a species at each site in the area of interest. Each site has a possibly different probability of detection which we assume can be estimated using covariates related to detection. Possible covariates related to detection might be aspect, percent ground cover, weather conditions, or the amount of time spent searching a given site. We use a function of these covariates to estimate probability of detection for a species at each site via the likelihood function of the data given the parameters.

We derive this extension to the autologistic model with covariates for sample data for estimation of detectability, describe an estimation algorithm, and demonstrate its utility through two examples.

5.1 Notation

We denote the q covariates that we suspect are related to detection by an $N \times (q+1)$ matrix W. We include an intercept term in the detection function so

that the first column of W is a column of ones. We introduce new parameters $\underline{\lambda} = (\lambda_0, \ldots, \lambda_q)$ which are the coefficients for the detection covariates in the likelihood function. All other notation is the same as the notation summarized in Table 3.1.

5.2 Likelihood Function

We propose a likelihood function for the observed presence/absence given the truth wherein the probability of nondetection using a logit function of the detection covariates. The probability of nondetection is the probability that we observe absence at a sampled, presence site. The probability of nondetection is calculable for sites which were searched, $a_i = 1$, and absence was recorded, $y_i = 0$. We assume that species presence is observed without error so that if $y_i = 1$ the probability of nondetection is zero. For sites which were not searched, $a_i = 0$, the probability of nondetection is one since we cannot observe the species if we don't sample. The likelihood function for the data given the truth that distinguishes these scenarios is as follows

$$f\left(\underline{y} \mid \underline{a}, \underline{x}, \underline{\lambda}\right) = \prod_{i=1}^{N} p(y_i \mid a_i, x_i, \underline{\lambda}).$$

where

$$p(y_i \mid x_i, a_i, \underline{\lambda}) = \begin{cases} 1 - y_i, & a_i = 0, \text{ or } x_i = 0, \\ \frac{\exp(y_i(\underline{w}_i \underline{\lambda}^T))}{1+\exp(\underline{w}_i \underline{\lambda}^T)}, & a_i = 1 \text{ and } x_i = 1, \end{cases} \tag{5.1}$$

for $i = 1, \ldots, N$ and w_i is the ith row of the detection covariates matrix, W. Thus the likelihood function simply states that the probability of nondetection is a logit function of the detection covariates, W, and coefficients, $\underline{\lambda}$, for sites with possible nondetection. To illustrate, consider Figure 5.1 which portrays a hypothetical relationship between percent ground cover and probability of nondetection for sites

with true species presence given by

$$Pr(y_i = 0 \mid x_i, a_i, \underline{\lambda}) = \frac{1}{1 + \exp(2 - 2 * w_i)} \tag{5.2}$$

Probability of nondetection for individual sites will fall somewhere along the line shown here.

This likelihood function extends the model of Heikkinen and Högmander (1994, HH hereafter) which involves only three levels of search intensity. HH use a categorical variable multiplier to include detectability which in their case represented search intensity. One category corresponded to no search which was their approach to utilizing sample data. Our approach will be much more widely applicable.

5.3 Prior and Posterior Distributions

In a hierarchical Bayesian setup we define prior distributions, hyperprior distributions, and the posterior distribution of the parameters given the data. Based on the likelihood function in Equation 5.1 we define two prior distributions, $\pi(\underline{x})$ and $\pi(\underline{\lambda})$. The prior distribution for $\underline{x}$, by the locally dependent Markov random field (LDMRF) assumption, is given in Equation 3.5 as

$$\pi(\underline{x} \mid \beta, \underline{\theta}) \propto \exp\left\{\left(\sum_{i=1}^{N} z_i x_i\right)\underline{\theta}^T + \beta \sum_{i<j} x_i x_j I_{[i\sim j]}\right\}. \tag{5.3}$$

This is numerically intractable due to the sum in the denominator over all possible configurations of presence/absence on the lattice.

We specify the prior distribution for the nondetection coefficients, $\underline{\lambda}$, to be a vague multivariate normal distribution, $\underline{\lambda} \sim N(\underline{0}, \Sigma_{\underline{\lambda}})$, where $\Sigma_{\underline{\lambda}}$ is a diagonal covariance matrix with $\left(\tau^2_{\lambda_0}, \tau^2_{\lambda_1}, \ldots, \tau^2_{\lambda_q}\right)$ on the diagonal. The values $\left(\tau^2_{\lambda_0}, \tau^2_{\lambda_1}, \ldots, \tau^2_{\lambda_q}\right)$ $\tau^2_{\lambda_q}\Big)$ are hyperparameters to be chosen. We assume *a priori* mutual independence of $\lambda_0, \ldots, \lambda_q$.

We use the hyperprior distributions for β and $\underline{\theta}$ that we specified in Section 3.2.2, so $\beta \sim \Gamma(\psi, \alpha)$ and $\underline{\theta} \sim N(\underline{0}, \Sigma)$.

The posterior distribution of the parameters given the data is

$$p\left(\underline{x}, \underline{\lambda}, \beta, \underline{\theta} \mid \underline{y}, \underline{a}\right) \propto f\left(\underline{y}, \underline{a} \mid \underline{x}, \underline{\lambda}\right) \pi\left(\underline{\lambda}\right) \pi\left(\underline{x} \mid \beta, \underline{\theta}\right) \pi\left(\beta\right) \pi\left(\underline{\theta}\right). \qquad (5.4)$$

5.4 Estimation

The posterior distribution is intractable due to the sum in the denominator so we present an extended version of the Gibbs-Hastings sampling algorithm described in Section 3.3 for estimation of the parameters $\underline{x}$, $\underline{\lambda}$, β, and $\underline{\theta}$. The Gibbs-Hastings sampler utilizes the full conditional distributions for all parameters to obtain posterior estimates. The distributions and quantities which are necessary for the estimation procedure are the full conditional distributions, the pseudolikelihood distributions for β and $\underline{\theta}$, the negative log distributions used to estimate means for the proposal distributions in the Hastings-Metropolis steps, and the Fisher information used to estimate variances for the proposal distributions for the Hastings-Metropolis steps.

We derive these new distributions and quantities for sampling $\underline{x}$ and $\underline{\lambda}$ values. We merely present the updated distributions and quantities for β and $\underline{\theta}$ since they are analogous to those derived in Section 3.3. We then review the steps of the Gibbs-Hastings estimation procedure for this extension.

5.4.1 Full Conditional Distributions for Presence/Absence (X)

The full conditional distribution for $\underline{x}$ is derived using the posterior distribution (Eqn 5.4), the prior distribution for $\underline{x}$ (Eqn 5.3), the likelihood function (Eqn 5.1),

and the *a priori* mutual independence of $\underline{y}$, β, and $\underline{\theta}$. The full conditional distribution for $\underline{x}$ is defined

$$
\begin{aligned}
p\left(\underline{x} \mid \underline{a}, \underline{y}, \underline{\lambda}, \beta, \underline{\theta}\right) &= \frac{p\left(\underline{x}, \underline{\lambda}, \beta, \underline{\theta} \mid \underline{y}, \underline{a}\right)}{p\left(\underline{\lambda}, \beta, \underline{\theta} \mid \underline{y}, \underline{a}\right)} \\
&\propto \frac{f\left(\underline{y}, \underline{a} \mid \underline{x}, \underline{\lambda}\right) \pi\left(\underline{\lambda}\right) \pi\left(\underline{x} \mid \right.}{\pi\left(\underline{\lambda}\right) \pi(\beta) \pi(} \\
&= f\left(\underline{y}, \underline{a} \mid \underline{x}, \underline{\lambda}\right) \pi\left(\underline{x} \mid \beta, \underline{\theta}\right) \\
&\propto \left(\prod_{i=1}^{N} p\left(y_i, a_i \mid x_i, \underline{\lambda}\right)\right) \exp\left\{\left(\sum_{i=1}^{n} z_i x_i\right) \underline{\theta}^T + s_i \beta\right\}
\end{aligned}
$$

where s_i is defined in Equation 3.24 and is *a priori* either 0 or 1. This distribution is intractable due to the normalizing constant in the denominator.

The Gibbs sampling algorithm uses the full conditional distribution for *each* parameter, so we are interested in the full conditional distribution for $\{x_i\}$. The full conditional distribution for each x_i is the likelihood function for the observations at each site multiplied by the conditional prior distribution for the truth, so

$$
p\left(x_i \mid y_i, \underline{x}_{s_i}, a_i, \underline{\lambda}, \beta, \underline{\theta}\right) \propto p\left(y_i \mid x_i, a_i, \underline{\lambda}\right) \pi\left(x_i \mid \underline{x}_{s_i}, \beta, \underline{\theta}\right). \tag{5.6}
$$

for $i = 1, \ldots, N$. The conditional prior distribution, $\pi(x_i \mid \underline{x}_{s_i}, \beta, \underline{\theta})$, is given in Equation 3.9. In particular there are three possible observed data combinations corresponding to nonsampled sites, observation of presence for sampled sites, and observation of absence for sampled sites. These are fully outlined in Equation 5.7. We first present the cases of presence and absence separately for clarity. The full conditional distribution for the probability that a species is truly present given observed presence/absence, true presence/absence in the neighborhood, and model

parameters is defined

$$
\begin{aligned}
p \quad & \left(x_i = 1 \mid y_i, a_i, \underline{x}_{S_i}, \underline{\lambda}, \beta, \underline{\theta}\right) \\
& = p(y_i \mid x_i = 1, a_i, \underline{\lambda})\, \pi\left(x_i = 1 \mid \underline{x}_{S_i}, \beta, \underline{\theta}\right) \\
& \propto \begin{cases} \exp\left(\underline{z}_i \underline{\theta}^T + s_i\beta\right), & a_i = 0, \\ \frac{1}{1+\exp(\underline{w}_i \underline{\lambda}^T)} \exp\left(\underline{z}_i \underline{\theta}^T + s_i\beta\right), & a_i = 1 \text{ and } y_i = 0, \\ \frac{\exp(\underline{w}_i \underline{\lambda}^T)}{1+\exp(\underline{w}_i \underline{\lambda}^T)} \exp\left(\underline{z}_i \underline{\theta}^T + s_i\beta\right), & a_i = 1 \text{ and } y_i = 1. \end{cases}
\end{aligned}
$$

So for a site which was not searched, the probability that the species is truly present is a function of the habitat and spatial covariates and if the site was searched, the probability that the species is truly present is also a function of the probability of detection, depending in turn on observed presence/absence. The probability that a species is truly absent given observed presence/absence, true presence/absence in the neighborhood and model parameters is proportional to

$$
\begin{aligned}
p\left(x_i = 0 \mid y_i, a_i, \underline{x}_{S_i}, \underline{\lambda}, \beta, \underline{\theta}\right) \quad & = \quad p(y_i \mid x_i = 0, a_i, \underline{\lambda})\, \pi\left(x_i = 0 \mid \underline{x}_{S_i}, \beta, \underline{\theta}\right) \\
& \propto \begin{cases} 1, & a_i = 0 \text{ or } a_i = 1 \text{ and } y_i = 0, \\ 0, & a_i = 1 \text{ and } y_i = 1. \end{cases}
\end{aligned}
$$

So for a site which was not searched, or absence was observed, the probability that the species is truly absent is proportional to one and for a site which was searched and presence observed, the probability that the species is truly absent is zero. We combine the above proportional full conditional distributions to fully specify the full

conditional distribution for species presence at site i

$$p\left(x_i = 1 \mid y_i, a_i, \underline{x}_{\delta_i}, \underline{\lambda}, \beta, \underline{\theta}\right) = \begin{cases} \frac{\exp\left(\underline{z}_i \underline{\theta}^T + s_i \beta\right)}{1+\exp\left(\underline{z}_i \underline{\theta}^T + s_i \beta\right)}, & a_i = 0. \\ \frac{\frac{1}{1+\exp\left(\underline{w}_i \underline{\lambda}^T\right)} \exp\left(\underline{z}_i \underline{\theta}^T + s_i \beta\right)}{1+\frac{1}{1+\exp\left(\underline{w}_i \underline{\lambda}^T\right)} \exp\left(\underline{z}_i \underline{\theta}^T + s_i \beta\right)}, & y_i = 0, a_i = 1. \\ \frac{\frac{\exp\left(\underline{w}_i \underline{\lambda}^T\right)}{1+\exp\left(\underline{w}_i \underline{\lambda}^T\right)} \exp\left(\underline{z}_i \underline{\theta}^T + s_i \beta\right)}{0+\frac{\exp\left(\underline{w}_i \underline{\lambda}^T\right)}{1+\exp\left(\underline{w}_i \underline{\lambda}^T\right)} \exp\left(\underline{z}_i \underline{\theta}^T + s_i \beta\right)} = 1, & y_i = 1, a_i = 1. \end{cases} \quad (5.7)$$

for $i = 1, \ldots, N$. The full conditional distribution for true presence at site i, $x_i = 1$, is a function of covariates. For sites which were **not** searched the probability of nondetection is one so only covariates related to species presence are used. For sites that were searched and absence observed the probability of nondetection contributes to the posterior probability of presence through covariates related to detectability. Large probability of nondetection at site i increases the posterior probability of presence at site i. Small probability of nondetection at site i decreases the posterior probability of presence at site i. If the species was observed at the site then nondetection is zero and the predicted probability of presence is one.

Equation 5.7 fully specifies the full conditional distributions for $\{x_i\}$. Sampling from this distribution is straightforward. The routine described in Chapter 3, Section 3.3.2, can be used to carry out the sampling.

5.4.2 Full Conditional Distribution for Nondetection Coefficients

To obtain estimates for $\underline{\lambda}$, the nondetection coefficients, we sample from the full conditional distribution for each λ_j, $j = 0, \ldots, q$. We use the likelihood function

(Eqn 5.1), the normal prior distribution for $\underline{\lambda}$, and the *a priori* mutual independence of $\underline{y}$, β, and $\underline{\theta}$ to obtain the full conditional distributions

$$\begin{aligned} p\left(\lambda_j \mid \underline{y}, \underline{a}, \underline{x}, \underline{\lambda}_{-j}, \beta, \underline{\theta}\right) &\propto f\left(\underline{y}, \underline{a} \mid \underline{x}, \underline{\lambda}, \beta, \underline{\theta}\right) \pi(\lambda_j) \\ &= f\left(\underline{y}, \underline{a} \mid \underline{x}, \underline{\lambda}\right) \pi(\lambda_j) \\ &= \left(\prod_{i=1}^{n} p(y_i, a_i \mid x_i, \underline{\lambda})\right) \pi(\lambda_j) \\ &= \left(\prod_{\substack{a_i=1 \\ y_i=1}} \frac{\exp\left(y_i\left(\underline{w}_i \underline{\lambda}^T\right)\right)}{1+\exp\left(\underline{w}_i \underline{\lambda}^T\right)}\right) \\ &\quad (2\pi\tau_{\lambda_j}^2)^{-1/2} \exp\left(-\frac{\lambda_j^2}{2\tau_{\lambda_j}^2}\right) \end{aligned} \tag{5.8}$$

for $j = 0, \ldots, q$. The product in this full conditional distribution is over sites which were searched and presence observed. If the species is not present at site i there is no possibility of nondetection and hence no contribution to the likelihood function. Sites which are not searched have probability of nondetection equal to one.

This full conditional distribution for $\lambda_j, j = 0, \ldots, q$ is fully specified but is difficult to sample from due to the normalizing constant so we use a Hastings-Metropolis step to obtain sample values. The Hastings-Metropolis step requires the target distribution which is the full conditional distribution and a proposal distribution. We use a normal proposal distribution where the parameters are equal to the maximum likelihood estimates from the target distribution.

The full conditional distribution for $\underline{\lambda}$ is a nonlinear function so we utilize a quasi-Newton minimization algorithm to obtain the MLE's for use in the proposal distribution. We minimize the negative of the log of the full conditional distribution to obtain the MLE for $\underline{\lambda}$, $\hat{\underline{\lambda}}_{\text{MLE}}$. The negative of the log of the full conditional distribution is

$$-\log p\left(\lambda_j \mid \underline{y}, \underline{a}, \underline{x}, \underline{\lambda}_{-j}, \beta, \underline{\theta}\right) = -\log\left[l(\underline{y}, \underline{a} \mid \underline{x}, \underline{\lambda}) \pi(\lambda_j)\right]$$

$$
\begin{aligned}
&= -\log\left[\prod_{\substack{a_i=1\\ y_i=1}} \frac{\exp(y_i(\underline{w}_i\,\underline{\lambda}^T))}{1+\exp(\underline{w}_i\,\underline{\lambda}^T)} \frac{1}{\sqrt{2\pi\tau^2_{\lambda_j}}} \exp\left(\frac{-\lambda_j^2}{2\tau^2_{\lambda_j}}\right)\right] \\
&= \sum_{\substack{a_i=1\\ y_i=1}} \left[-y_i\,\underline{w}_i\,\underline{\lambda}^T + \log(1+\exp(\underline{w}_i\,\underline{\lambda}^T))\right] \\
&\quad + \frac{1}{2}\log(2\pi\tau^2_{\lambda_j}) + \frac{\lambda_j^2}{2\tau^2_{\lambda_j}}
\end{aligned}
$$

for $j = 0, \ldots, q$.

The variance for the normal proposal distribution is obtained as a function of the Fisher information. We derive this variance as follows.

$$
\begin{aligned}
\sigma^2_{\hat{\lambda}_{j,\mathrm{MLE}}} &= \left[\left[-\frac{\partial^2 \log p(\lambda_j \mid \cdot)}{\partial \lambda_j^2}\right]^{-1}\right]_{\lambda_j=\hat{\lambda}_{j,\mathrm{MLE}}} \\
&= \left[\left[-\frac{\partial^2}{\partial \lambda_j^2}\left[\sum_{\substack{a_i=1\\ y_i=1}} \left[y_i \underline{w}_i\,\underline{\lambda}^T - \log(1+\exp(\underline{w}_i\,\underline{\lambda}^T)\right]\right.\right.\right. \\
&\quad \left.\left.\left. -\frac{1}{2}\log(2\pi\tau^2_{\lambda_j}) - \frac{\lambda_j^2}{\tau^2_{\lambda_j}}\right]\right]^{-1}\right]_{\lambda_j=\hat{\lambda}_{j,\mathrm{MLE}}} \\
&= \left[\left[-\frac{\partial}{\partial \lambda_j}\left[\sum_{\substack{a_i=1\\ y_i=1}} \left(y_i w_{ji} - \frac{w_{ji}\exp(\underline{w}_i\,\underline{\lambda}^T)}{1+\exp(\underline{w}_i\,\underline{\lambda}^T)}\right) - \frac{\lambda_j}{\tau^2_{\lambda_j}}\right]\right]^{-1}\right]_{\lambda_j=\hat{\lambda}_{j,\mathrm{MLE}}} \\
&= \left[\left[\sum_{\substack{a_i=1\\ y_i=1}} \frac{w_{ij}^2 \exp(\underline{w}_i\,\underline{\lambda}^T)}{(1+\exp(\underline{w}_i\,\underline{\lambda}^T))^2} + \frac{1}{\tau^2_{\lambda_j}}\right]^{-1}\right]_{\lambda_j=\hat{\lambda}_{j,\mathrm{MLE}}} \qquad (5.9)
\end{aligned}
$$

At iteration t the Hastings-Metropolis step for obtaining sample values for $\lambda_j, j = 0, \ldots, q$ is as follows. We obtain maximum likelihood estimates $\hat{\lambda}_{j,\mathrm{MLE}}$ and $\sigma^2_{\hat{\lambda}_{j,\mathrm{MLE}}}$. We then sample a value from the proposal distribution with these parameters. This sampled value is accepted with probability

$$
\alpha_{\lambda_j^{(t)}} = \frac{p\left(\lambda_j^{(t)}\right) g\left(\lambda_j^{(t-1)}\right)}{p\left(\lambda_j^{(t-1)}\right) g\left(\lambda_j^{(t)}\right)}
$$

$$= \frac{\prod_{i=1}^{N} p\left(y_i \mid x_i, \underline{\lambda}^{(t)}\right) \pi\left(\lambda_j^{(t)}\right) \phi\left(\frac{\lambda_j^{(t-1)} - \hat{\lambda}_{j,\mathrm{MLE}}}{\sigma_{\hat{\lambda}_{j,\mathrm{MLE}}}}\right)}{\prod_{i=1}^{N} p\left(y_i \mid x_i, \underline{\lambda}^{(t-1)}\right) \pi\left(\lambda_j^{(t-1)}\right) \phi\left(\frac{\lambda_j^{(t)} - \hat{\lambda}_{j,\mathrm{MLE}}}{\sigma_{\hat{\lambda}_{j,\mathrm{MLE}}}}\right)}. \tag{5.10}$$

If $\lambda_j^{(t)}$ is not accepted, the previous value, $\lambda_j^{(t-1)}$ is used.

5.4.3 Full Conditional Distribution for Model Parameters

Next we present the Hastings-Metropolis steps used to obtain sample values for model parameters β and $\underline{\theta}$ at iteration t. This step is similar to the Metropolis-Hastings step described in Section 3.3.3. The full conditional distributions for β and $\underline{\theta}$ are given by

$$\begin{aligned} p\left(\beta \mid \underline{y}, \underline{a}, \underline{x}, \underline{\lambda}, \underline{\theta}\right) &\propto p\left(\underline{x} \mid \underline{y}, \underline{a}, \underline{\lambda}, \beta, \underline{\theta}\right) p\left(\beta \mid \underline{y}, \underline{a}, \underline{\lambda}, \underline{\theta}\right) \\ &= p\left(\underline{x} \mid \underline{y}, \underline{a}, \underline{\lambda}, \beta, \underline{\theta}, \right) \pi(\beta). \end{aligned} \tag{5.11}$$

and

$$\begin{aligned} p\left(\theta_k \mid \underline{y}, \underline{a}, \underline{x}, \underline{\lambda}, \underline{\theta}_{-k}\right) &\propto p\left(\underline{x} \mid \underline{y}, \underline{a}, \underline{\lambda}, \beta, \underline{\theta}, \right) p\left(\theta_k \mid \underline{y}, \underline{a}, \underline{\lambda}, \underline{\theta}_{-k}\right) \\ &= p\left(\underline{x} \mid \underline{y}, \underline{a}, \underline{\lambda}, \beta, \underline{\theta}, \right) \pi(\theta_k). \end{aligned} \tag{5.12}$$

for $k = 0, \ldots, p$. These distributions are intractable due to the intractable full conditional distribution for $\underline{x}$.

Since the information from border sites is limited, we restrict the pseudolikelihood to include only interior sites (as in Section 3.3.3.2). We define b_i for site i that is 1 if site i is interior and was sampled and 0 if site i is a border site or was not sampled. We present the full conditional pseudolikelihood distributions for β and $\underline{\theta}$ which we use as the target distributions in the Hastings-Metropolis steps. We then

present the proposal distributions which are normal with means equal to the maximum pseudolikelihood estimates and variances derived from the Fisher information of the pseudolikelihood distributions.

We use the full conditional distribution for x_i given in Equation 5.7 to obtain the full conditional pseudolikelihood distribution for $\underline{x}$

$$\begin{aligned} \mathrm{pl}\left(\underline{x} \mid \underline{y}, \underline{b}, \underline{\lambda}, \beta, \underline{\theta}\right) &= \prod_{i=1}^{N} p\left(x_i \mid y_i, b_i, \underline{x}_{s_i}, \underline{\lambda}, \beta, \underline{\theta}\right) \\ &= \prod_{i=1}^{N} \left[\frac{\left[\left(\frac{1}{1+\exp(\underline{w}_i \underline{\lambda}^T)}\right)^{b_i} \exp(\underline{z}_i \underline{\theta}^T + s_i \beta)\right]^{x_i}}{1 + \left(\frac{1}{1+\exp(\underline{w}_i \underline{\lambda}^T)}\right)^{b_i} \exp(\underline{z}_i \underline{\theta}^T + s_i \beta)} \right]^{(1-y_i)} \end{aligned} \quad .(5.13)$$

This full conditional pseudolikelihood distribution is used in the full conditional pseudolikelihood distribution for β

$$\begin{aligned} \mathrm{pl}\left(\beta \mid \underline{y}, \underline{b}, \underline{x}, \underline{\lambda}, \underline{\theta}\right) &= \mathrm{pl}\left(\underline{x} \mid \underline{y}, \underline{b}, \underline{\lambda}, \beta, \underline{\theta}\right) \pi(\beta) \\ &= \prod_{y_i=0} \frac{\left[\left(\frac{1}{1+\exp(\underline{w}_i \underline{\lambda}^T)}\right)^{b_i} \exp(\underline{z}_i \underline{\theta}^T + s_i \beta)\right]^{x_i}}{1 + \left(\frac{1}{1+\exp(\underline{w}_i \underline{\lambda}^T)}\right)^{b_i} \exp(\underline{z}_i \underline{\theta}^T + s_i \beta)} \frac{\beta^{\psi-1} \exp(-\beta/\alpha)}{\alpha^{\psi} \Gamma(\psi)}. \end{aligned}$$

Similarly, the full conditional pseudolikelihood for θ_k is

$$\begin{aligned} \mathrm{pl}\left(\theta_k \mid \underline{y}, \underline{b}, \underline{x}, \underline{\lambda}, \beta, \underline{\theta}_{-k}\right) &= \mathrm{pl}\left(\underline{x} \mid \underline{y}, \underline{b}, \underline{\lambda}, \beta, \underline{\theta}\right) \pi(\theta_k) \\ &= \prod_{y_i=0} \frac{\left[\left(\frac{1}{1+\exp(\underline{w}_i \underline{\lambda}^T)}\right)^{b_i} \exp(\underline{z}_i \underline{\theta}^T + s_i \beta)\right]^{x_i}}{1 + \left(\frac{1}{1+\exp(\underline{w}_i \underline{\lambda}^T)}\right)^{b_i} \exp(\underline{z}_i \underline{\theta}^T + s_i \beta)} \\ &\quad (2\pi\tau_k^2)^{-1/2} \exp\left(\frac{-\theta_k^2}{2\tau_k^2}\right). \end{aligned}$$

for $k = 0, 1, \ldots, p$.

Next we obtain the MLE's which we use as the parameters in the proposal distributions. The pseudolikelihood distributions are nonlinear functions so we use

a quasi-Newton minimization algorithm on the negative of the log of the full conditional pseudolikelihoods to obtain the maximum pseudolikelihood estimates $\hat{\beta}_{\text{MPLE}}$ and $\hat{\underline{\theta}}_{\text{MPLE}}$. The negative log full conditional pseudolikelihood distribution for β is

$$
\begin{aligned}
-\log \text{pl}\left(\beta \mid \underline{y}, \underline{b}, \underline{x}, \underline{\lambda}, \underline{\theta}\right) & \\
&= -\sum_{y_i=0} \left[x_i \left\{ b_i \log\left[1+\exp(\underline{w}_i \underline{\lambda}^T)\right] + \underline{z}_i \underline{\theta}^T + s_i \beta \right\} \right. \\
&\quad \left. - \log\left[1 + \left\{1/\left[1+\exp(\underline{w}_i \underline{\lambda}^T)\right]\right\}^{b_i} \exp(\underline{z}_i \underline{\theta}^T + s_i \beta)\right]\right] \\
&\quad -(\nu-1)\log\beta + \frac{\beta}{\alpha} + \nu \log\alpha + \log\Gamma(\nu).
\end{aligned}
$$

The negative log full conditional pseudolikelihood distribution for θ_k is

$$
\begin{aligned}
-\log \quad & \text{pl}\left(\theta_k \mid \underline{y}, \underline{b}, \underline{x}, \underline{\lambda}, \beta, \underline{\theta}_{-k}\right) \\
&= -\sum_{y_i=0} \left[x_i \left[b_i \log\left(1+\exp(\underline{w}_i \underline{\lambda}^T)\right) + \underline{z}_i \underline{\theta}^T + s_i \beta \right] \right. \\
&\quad \left. - \log\left(1 + \left[1/\left(1+\exp(\underline{w}_i \underline{\lambda}^T)\right)\right]^{b_i} \exp(\underline{z}_i \underline{\theta}^T + s_i \beta)\right)\right] \\
&\quad + \frac{1}{2}\log(2\pi\tau_k^2) + \frac{\theta_k^2}{2\tau_k^2}
\end{aligned}
$$

for $k = 0, \ldots, p$.

The Fisher information is used to obtain maximum likelihood estimators for the variances in the proposal distributions. We present the derived variances below.

$$
\begin{aligned}
\sigma^2_{\hat{\beta}_{\text{MPLE}}} &= \left[\left[-\frac{\partial^2 \log pl\,(\beta)}{\partial\beta^2}\right]^{-1}\right]_{\beta=\hat{\beta}_{\text{MPLE}}} \\
&= \left[\left[\sum_{y_i=0} \frac{s_i^2 \left(\frac{1}{1+\exp(\underline{w}_i \underline{\lambda}^T)}\right)^{b_i} \exp(\underline{z}_i \underline{\theta}^T + s_i\beta)}{\left(1+\left(\frac{1}{1+\exp(\underline{w}_i \underline{\lambda}^T)}\right)^{b_i} \exp(\underline{z}_i \underline{\theta}^T + s_i\beta)\right)^2} + \frac{\nu-1}{\beta^2}\right]^{-1}\right]_{\beta=\hat{\beta}_{\text{MPLE}}} \qquad (5.14)
\end{aligned}
$$

$$\sigma^2_{\hat{\theta}_k,\text{MPLE}} = \left[\left[-\frac{\partial^2 \log pl\,(\theta_k)}{\partial \theta_k^2}\right]^{-1}\right]_{\theta_k=\hat{\theta}_{k,\text{MPLE}}}$$

$$= \left[\left[\sum_{y_i=0} \frac{z_{ki}^2 \left(\frac{1}{1+\exp(\underline{w}_i \underline{\lambda}^T)}\right)^{b_i} \exp(\underline{z}_i \underline{\theta}^T + s_i \beta)}{\left(1 + \left(\frac{1}{1+\exp(\underline{w}_i \underline{\lambda}^T)}\right)^{b_i} \exp(\underline{z}_i \underline{\theta}^T + s_i \beta)\right)^2} - \frac{2}{\tau_k^2}\right]^{-1}\right]_{\theta_k=\hat{\theta}_{k,\text{MPLE}}} \quad (5.15)$$

for $k = 0, \ldots, p$.

The Hastings-Metropolis steps at iteration t for obtaining sample values for β and $\underline{\theta}$ are as follows. We obtain maximum likelihood estimates and then sample values from the normal proposal distributions

$$N\left(\hat{\beta}_{\text{MPLE}}, \sigma^2_{\hat{\beta}_{\text{MPLE}}}\right). \quad (5.16)$$

and

$$N\left(\hat{\theta}_{k,\text{MPLE}}, \sigma^2_{\hat{\theta}_{k,\text{MPLE}}}\right). \quad (5.17)$$

for $k = 0, \ldots, p$. The values are accepted with probabilities analogous to Equations 3.22 and 3.23. If a new value is rejected the previously sampled value is used.

5.4.4 Gibbs-Hastings Sampler

We review the steps in the Gibbs-Hastings estimation algorithm for the autologistic model with covariates and detection parameter for sample data. We initialize the values of $\underline{s}^{(0)}$, $\beta^{(0)}$, and $\underline{\theta}^{(0)}$ as described in Chapter 3. The detection coefficients, $\underline{\lambda}^{(0)}$, are initialized in the same manner, using estimates from a logistic regression on the sample data. At iteration t one cycle of the Gibbs sampler is described as follows:

1. Calculate the predicted probability of species presence, $\underline{\hat{p}}^{(t)}$, given in Equation 5.7 and the spatial covariate, $\underline{s}^{(t)}$, given in Equation 3.24 for groups of independent sites as described in Section 3.3.2

2. Sample $\underline{x}^{(t)}$, the proposed estimate of the truth, from Bernoulli ($\underline{\hat{p}}^{(t)}$) (Eqn 5.7)

3. Sample each $\lambda_j^{(t)}$ for $j = 0, \ldots, q$ using the Hastings-Metropolis step described in Section 5.4.2

4. Sample each $\theta_k^{(t)}$for $k = 0, \ldots, p$ and $\beta^{(t)}$ using a Hastings-Metropolis step as described in Section 5.4.3

These steps are iterated until convergence as described in Section 3.3.5.

5.5 Examples

We present two examples to demonstrate the improvement in prediction over the autologistic model with covariates for sample data made by using the modified likelihood function to estimate detectability. The simulation setup is the same as Chapter 4.

For the first example we use one covariate, $\underline{z}$, which is pictured in the top left of Figure 5.2. This covariate is confusing since it displays good habitat in some areas of species absence. We specify a covariate related to detection, $\underline{w}$, which is the top right scene in Figure 5.2. In our simulations, high values of $\underline{w}$ (darker pixels) correspond to lower detection of the species. The observed presence/absence data is obtained as follows. We generate site numbers from a simple random sample of size 160. If the species is present at a sample site we observe presence with probability $\frac{\exp(2-3w_i)}{1+\exp(2-3w_i)}$, otherwise we observe absence. This probability is the defined functional

form of the likelihood function with intercept 2 and slope -3. Using this detection function, we observed absence at 14 sites which actually contained the species. The bottom left scene of Figure 5.2 shows the truth, the sample sites, and the observed presence sites. The sites with observed presence are black, the sites with observed absence are medium grey, sites with true presence are light grey, and sites with true species absence are very light grey. Hence medium grey sites which are within the light grey areas are sampled sites at which the species was 'missed'. The prevalence of nondetection is greater toward the top where the detectability covariate is large; for example corresponding to high percent ground cover.

The resulting predicted probability of presence from the autologistic model with covariates and estimated detection for sample data is shown in the bottom right. These results are very similar to Example 2 in Chapter 4 where there are no observations in the top portion although there is a swath of good habitat. Although no presence was observed in the top right corner, the predicted probability of presence there is high. This area of high predicted probability corresponds to good habitat. The sensitivity and specificity (Eqn 4.2) are 0.96 and 0.83 respectively. These measures indicate that the model is predicting correctly most of the time but is less accurate at predicting true absence. The results from the model without detectability are 0.79 and 0.97 for sensitivity and specificity. Thus it seems there is a tradeoff of sensitivity and specificity in prediction by using detection information in this case. Histograms of the predicted probability of presence for groups of absence and presence sites is given in Figure 5.3. There are a few absence sites with high predicted probability of presence but overall the predictions show two groups of high and low predicted probabilities which match true presence and absence.

In Example 2 we use the same covariate related to species presence, $\underline{z}$. The detectability covariate, $\underline{w}$, now corresponds to better detection toward the top of the lattice. The observations were formulated as in Example 1 using probability of observing presence $\frac{\exp(-2-0.5w_i)}{1+\exp(-2-0.5w_i)}$. We used a different detection parameter here to accommodate the altered detection covariate values. Here, we mistakenly observed absence at 9 sites. The observed scene is set up as in Example 1. Note that there are no observations of presence in the bottom half of the lattice. This is similar to the observed presence/absence data in Example 3 from Chapter 4. The sample sizes are similar, 180 for Example 3 from Chapter 4 and 160 for these examples. The predicted probability of presence is shown in the bottom right of Figure 5.4. The predicted probability of presence for the area of 'missed' species at the bottom of the area of interest is higher than from the predicted probability of presence from the model without detection. The top portion indicates a loss in specificity when using detection. The sensitivity and specificity for Example 2 are 0.7 and 0.84, respectively. The sensitivity and specificity for the model without detection was 0.66 and 0.98. Thus showing the same tradeoff we observed in Example 1. The histograms of predicted probability of presence (Figure 5.5) show that there are two groups of predicted probabilities, and that there is some error in their matching with true presence and absence due to the lack of observed presence information in the bottom half.

5.5.1 Conclusions

The addition of a detectability covariate has made a further improvement in sensitivity of prediction for situations with poor or misleading information. There is a slight reduction in specificity by using the detectability covariate. The importance

of the tradeoff will vary from application to application. The amount of observed presence/absence information we used in the estimation of the detection parameters for these examples is small. More covariate information from observations with species presence would be necessary for precise estimates of $\underline{\lambda}$.

Figure 5.1: Example of Nondetection Function $\frac{1}{1+\exp(2-2*\text{Ground Cover})}$

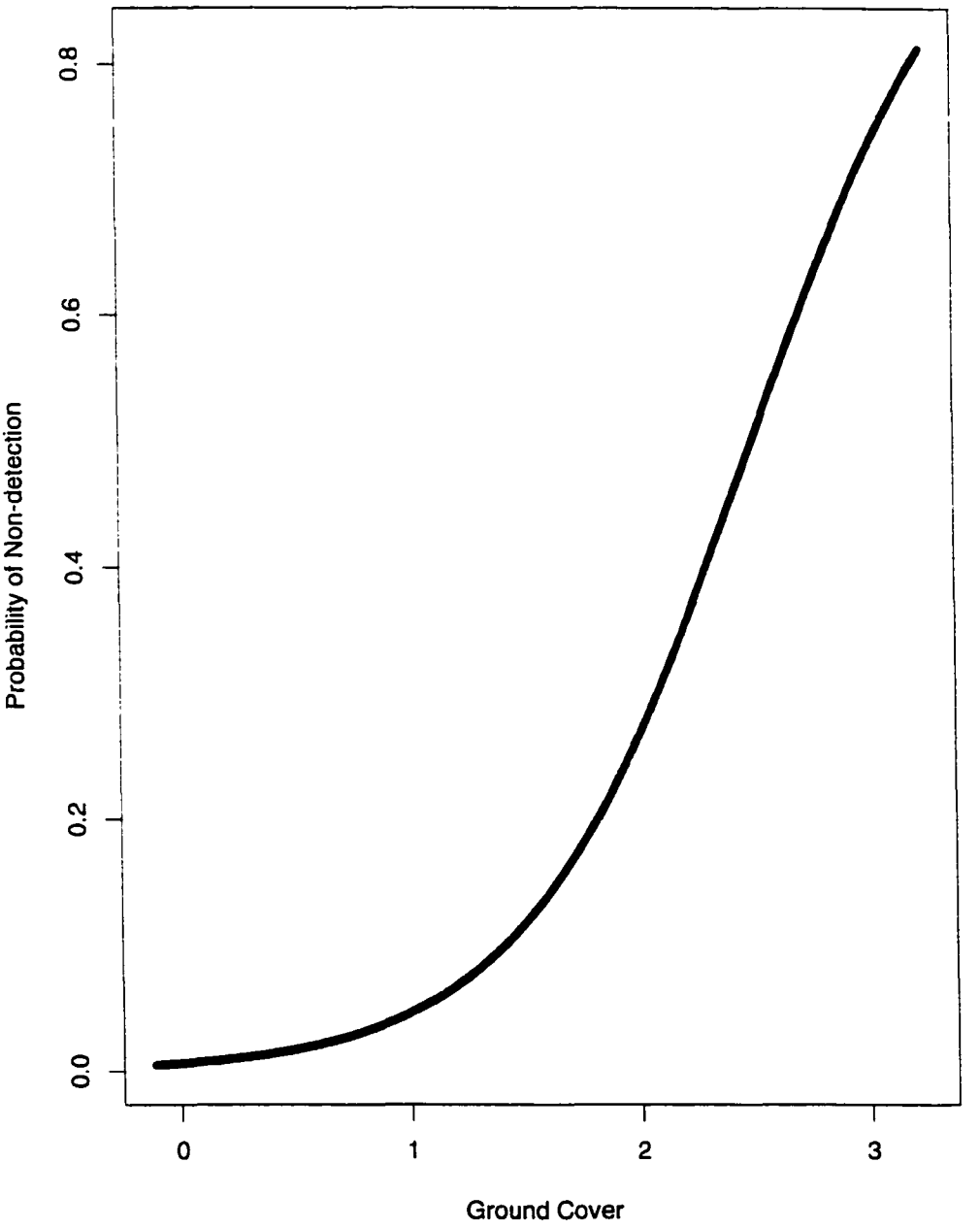


Figure 5.2: Example 1: Demonstration of Detectability Estimation

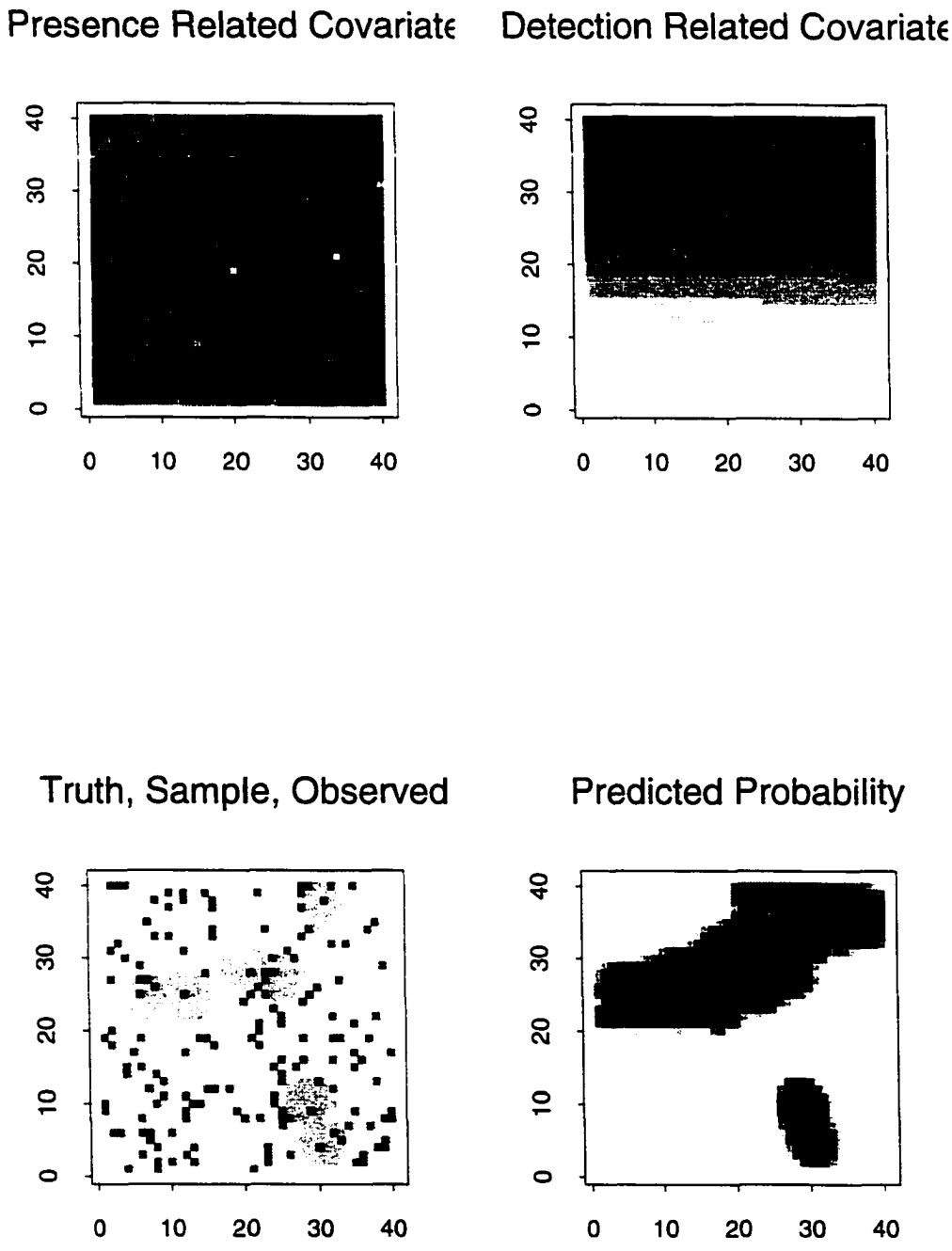


Figure 5.3: Example 1: Histograms of Predicted Probability of Presence for Absence and Presence Sites.

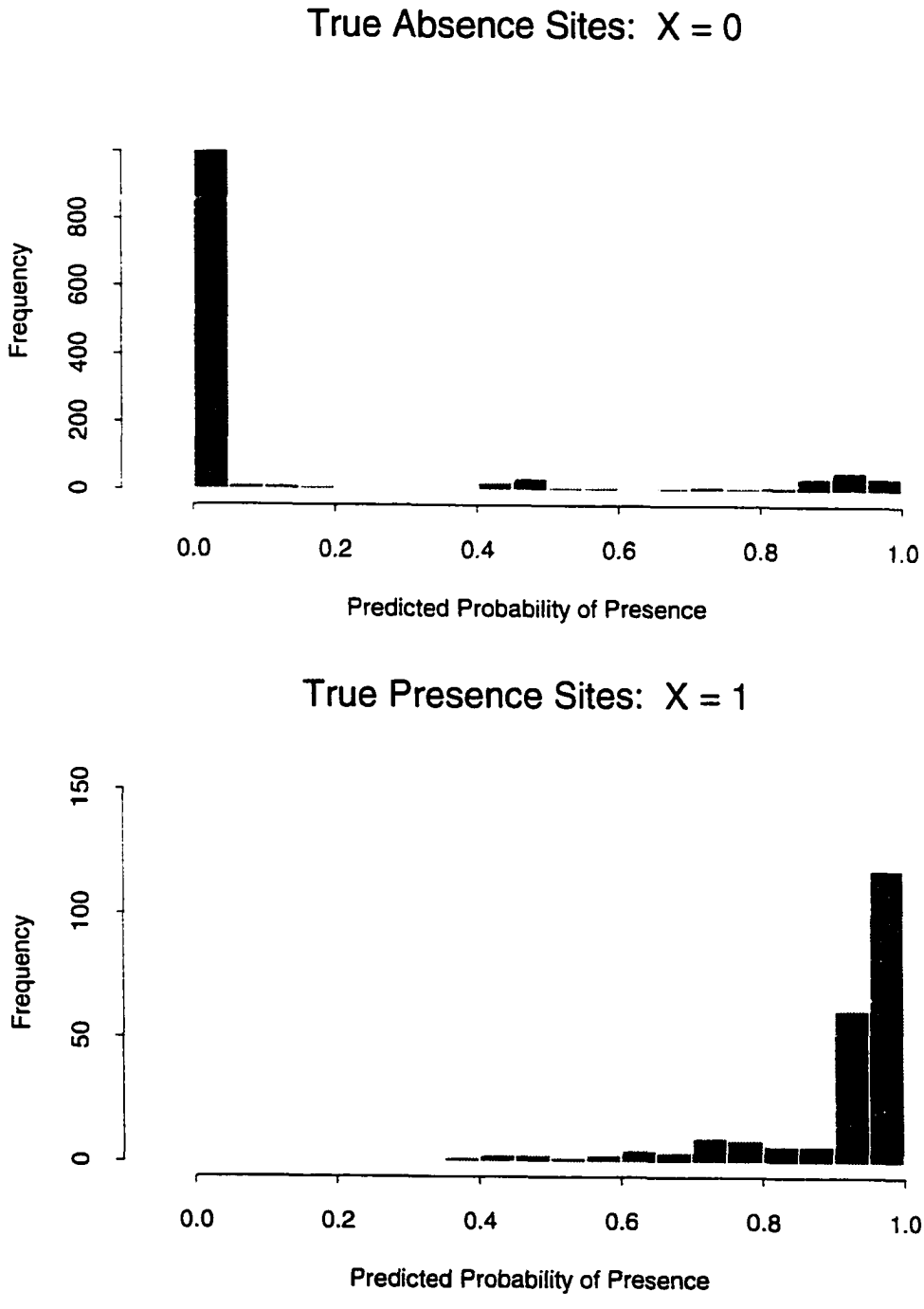


Figure 5.4: Example 2: Demonstration of Detectability Estimation

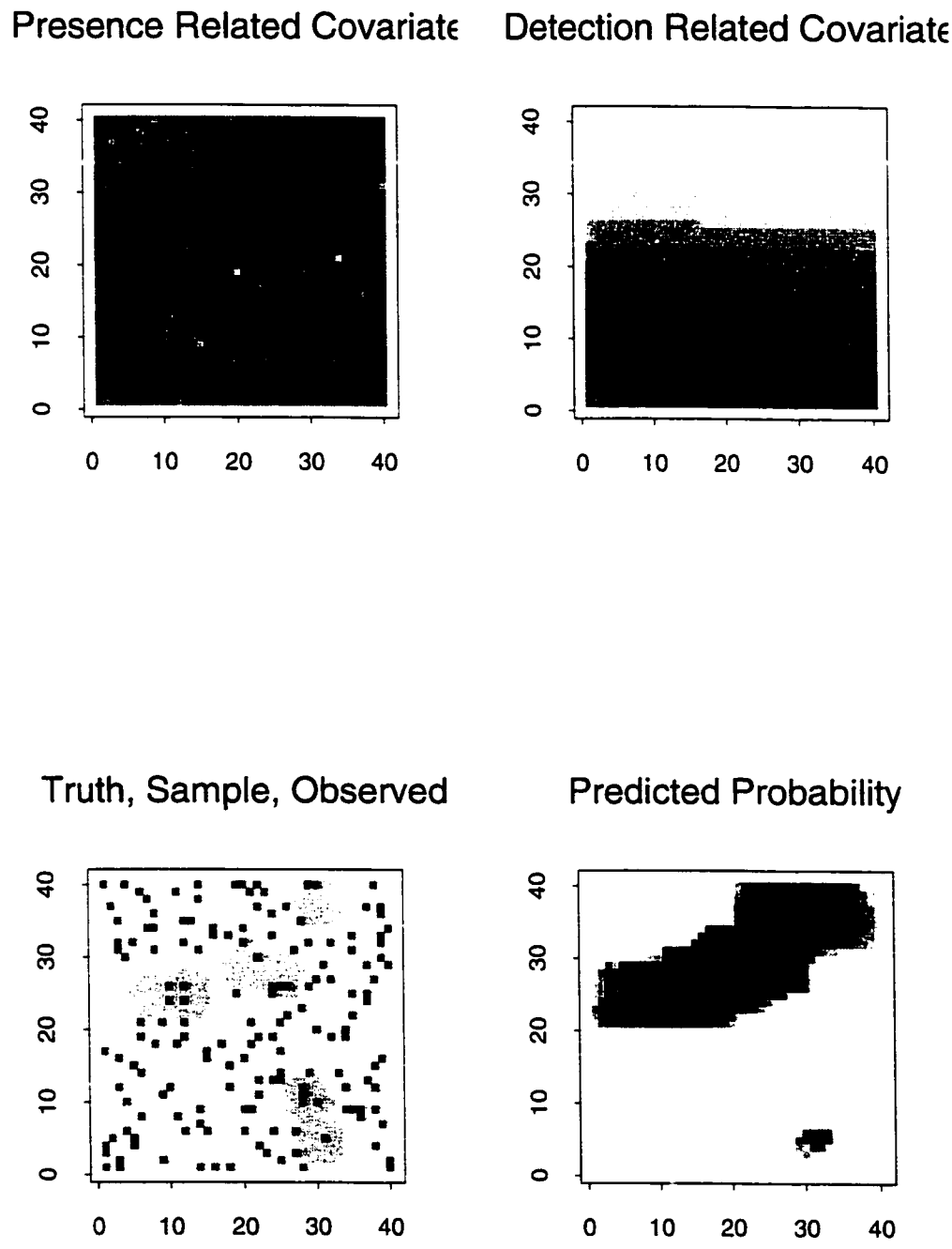
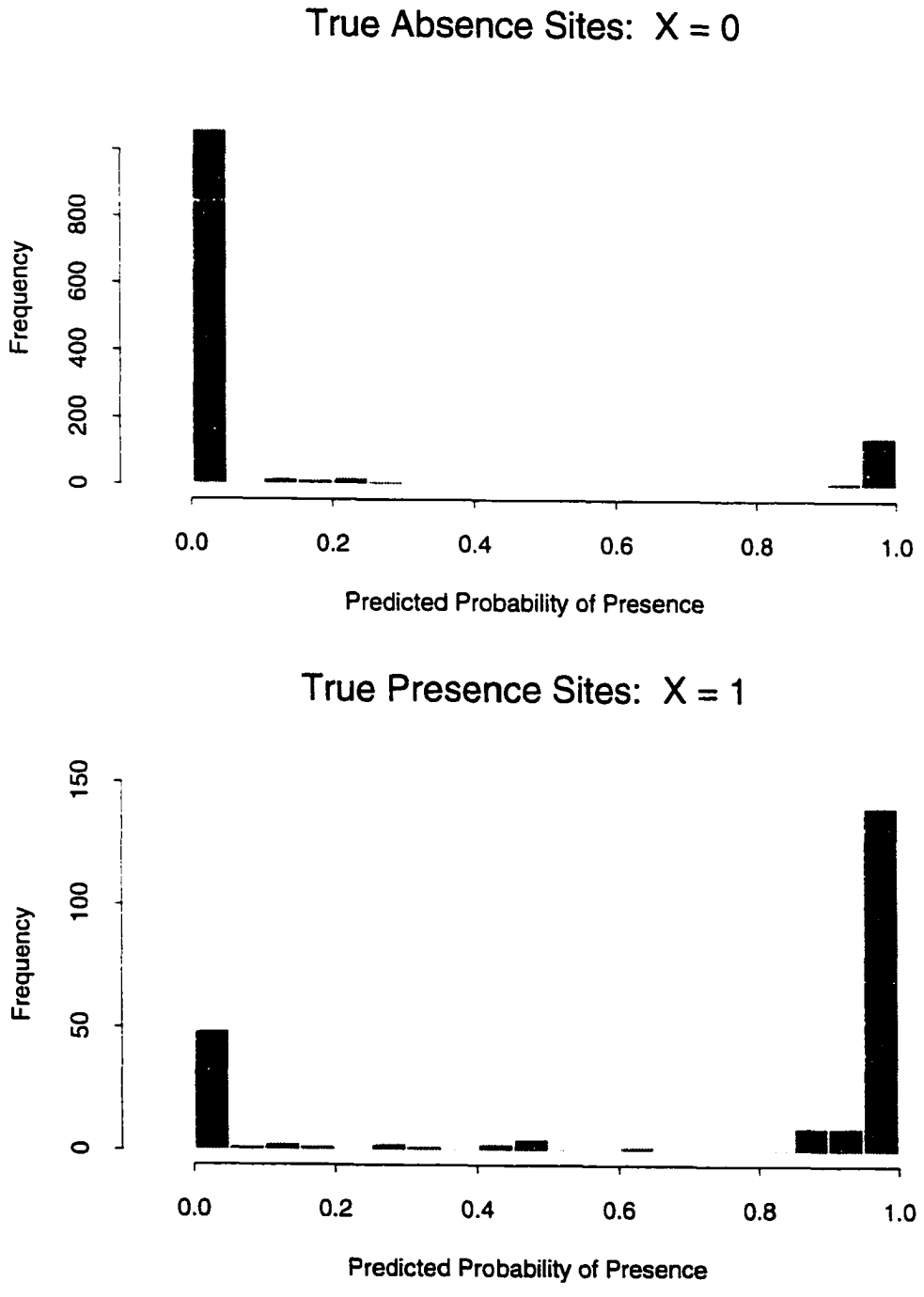


Figure 5.5: Example 2: Histogram of Predicted Probability of Presence for Absence and Presence Sites



Chapter 6

DEVELOPMENT OF SAMPLING DESIGNS

The second goal of the United States Forest Service (USFS) project is to design sampling plans to improve detection of rarely observed species. The sampling plan needs to satisfy certain cost or sample size constraints and maximize the expected number of sample sites which are occupied by the species of interest. At the same time the sampling plan should provide a design unbiased and efficient estimator of the total number of occupied sites. We investigate sampling plans with these goals in mind. The sampling plans we consider use the predicted probability of presence from the autologistic model with covariates for sample data. We introduce sampling plans which incorporate predicted probability of presence in the sample selection scheme or estimation procedure.

There are two main considerations for sampling plan development in the USFS project. The first goal is to maximize detection of the species of interest. We will utilize the predicted probability of presence from the autologistic model with covariates for sample data to increase the probability of sampling occupied sites. A good fitting model will serve to maximize detections, but the predictions are not without error. The sampling plan must be robust to possible errors in prediction.

The other goal in sample design for the USFS project is to provide an estimator of the total number of occupied sites which has an acceptable mean squared error. We consider unbiased estimators since for this single objective a 'best' estimator

would attain minimum variance. Different sample plans are required to achieve these two goals. We investigate several sampling plans which provide a reasonable compromise between the two goals.

A naive sampling plan would cover the area of interest evenly. Ignoring available information, such as the predicted probability of presence, sites would be equally likely to contain the species *a priori*. We have developed the autologistic model with covariates for sample data to obtain predicted probability of presence for the species of interest for each site. We investigate three sampling plans which use predicted probability of presence to select sample sites: sampling with inclusion probabilities proportional to the predicted probability of presence, stratified sampling, and fixed top stratum sampling (see Section 6.4 for definition). We also consider simple random sampling using the estimated probability in estimation via the ratio estimator.

The chapter begins by reviewing and introducing notation. There are three classes of sampling design which are introduced, sampling with probability proportional to estimated probability, stratified sampling, and fixed top stratum sampling. We introduce each along with their respective design parameter alternatives, and make comparisons along the way. Finally we demonstrate the level of utility for each of the three models, the logistic regression model, autologistic model, and the autologistic model with covariates for sample data, in designing sample plans with respect to expected number of sample detections and estimation variance for chosen designs and estimators.

6.1 Notation

We review notation for the discussion of sampling plans which is consistent with the previous chapters and introduce further necessary notation.

The presence or absence of a species at site i is denoted by x_i. The number of sites on the lattice, which is the population size, is N. The predicted posterior probability of presence for site i is denoted by $\hat{p}_i = \hat{P}(x_i = 1 \mid \underline{y}, \underline{a})$.

The single draw sample inclusion probability is denoted by p_i. The definition of this will change depending on the specified design.

In the sampling plans described below, the sample size is denoted by n. For a stratified population, we use h as a stratum index to identify the different strata, and H as the total number of strata. The number of units in stratum h, is denoted by N_h. The sample size for stratum h is n_h.

The total number of occupied sites in the population is denoted by $X = \sum_{i=1}^{N} x_i$. The total number present in each stratum is denoted by $X_h = \sum_{i=1}^{N_h} x_{h,i}$, the sum of indicators for sites i in stratum h. Note that $\sum_{h=1}^{H} X_h = X$.

6.2 Sampling with Probability Proportional to Estimated Likelihood

We assume the predicted probability of presence for each site from a good fitting model is closely related to presence of the species. We propose assigning sample inclusion probabilities to sites based on the predictions from the autologistic model with covariates for sample data. We refer to the predicted probability of presence as the estimated probability and the sampling plans based on the estimated probability as sampling with probability proportional to estimated probability (SPL).

In SPL plans, sites are selected using one of three general approaches. The procedures for choosing sample points differ depending on whether the sites are "draw" sequentially selected from the population of sites, or "list" sequentially selected where each site in turn is selected with a given inclusion probability (Särndal *et al.*,

1991). We introduce three sample selection approaches and compare their estimators.

6.2.1 Sampling With Replacement

To sample with replacement each site is first assigned a single draw probability of inclusion which is proportional to the estimated probability, $p_1, \ldots, p_N$, where

$$p_i = \frac{\hat{p}_i}{\sum_{j=1}^{N} \hat{p}_j}. \tag{6.1}$$

Next the single draw inclusion probabilities, p_i, and the cumulative single draw inclusion probabilities, $\sum_{k=1}^{i} p_k$, are calculated for all sites. A uniform random variable between 0 and 1, u, is chosen and site i, $i > 1$, is chosen to be in the sample if $\sum_{k=1}^{i-1} p_k \leq u < \sum_{k=1}^{i} p_k$. Site 1 is chosen if $u < p_1$. The selection of n independent uniform variate values $[0, 1)$ then identifies the n site labels composing the sample.

The expected number of detections for SPL with replacement is easily determined. Let t_i be a random variable which counts the number of times the i^{th} site is included in the sample. The expected value of t_i is $E(t_i) = np_i$. Then the sample number of observed occupied sites is $D^{*}_{\text{WR}} = \sum_{i=1}^{N} x_i t_i$ where x_i is the true presence/absence at site i. Hence the expected number of detections for SPL with replacement is

$$\begin{aligned} \mathrm{E}(D^{*}_{\text{WR}}) &= \sum_{i=1}^{N} x_i E(t_i) \\ &= \sum_{i=1}^{N} x_i n p_i. \end{aligned}$$

Sampling with replacement may result in observing and counting an occupied site more than once. What is of interest, however, is maximizing the number of *unique* occupied sites, D. The expected number of unique detections is obtained

using an indicator, $I_{[t_i \geq 1]}$, which equals one when site i is included in the sample at least once. The expected number of unique sample detections is then

$$\begin{aligned} \mathrm{E}(D_{\mathrm{WR}}) &= \sum_{i=1}^{N} x_i E(I_{[t_i \geq 1]}) \\ &= \sum_{i=1}^{N} x_i P(t_i \geq 1) \\ &= \sum_{i=1}^{N} x_i (1 - P(t_i = 0)) \\ &= \sum_{i=1}^{N} x_i (1 - (1 - p_i)^n) \\ &= X - \sum_{i=1}^{N} x_i (1 - p_i)^n. \end{aligned}$$

The expected number of unique detections is always less than or equal to the number of expected detections when sampling is done with replacement.

The Hansen-Hurwitz (1943) estimator of the number of occupied sites is

$$\begin{aligned} \hat{X}_{\mathrm{HH}} &= \sum_{i=1}^{n} \frac{x_i}{np_i} \\ &= \sum_{i=1}^{N} \frac{x_i t_i}{np_i} \\ &= \sum_{i=1}^{N} x_i \frac{t_i}{E(t_i)}. \end{aligned}$$

which is obviously an unbiased estimator of X. The variance of this estimator is

$$V(\hat{X}_{\mathrm{HH}}) = \frac{1}{n} \sum_{i=1}^{N} p_i \left(\frac{x_i}{p_i} - X \right)^2. \tag{6.2}$$

This variance simplifies to

$$\begin{aligned} V(\hat{X}_{\mathrm{HH}}) &= \frac{1}{n} \sum_{i=1}^{N} p_i \left(\frac{x_i^2}{p_i^2} - 2X \frac{x_i}{p_i} + X^2 \right) \\ &= \frac{1}{n} \sum_{i=1}^{N} \left(\frac{x_i^2}{p_i} - 2X x_i + X^2 p_i \right) \\ &= \frac{1}{n} \left(\sum_{i=1}^{N} \frac{x_i^2}{p_i} - 2X^2 + X^2 \right) \\ &= \frac{1}{n} \left(\sum_{i=1}^{N} \frac{x_i^2}{p_i} - X^2 \right) \end{aligned} \tag{6.3}$$

We use this form of the variance for comparisons with other SPL methods.

6.2.2 Sampling Without Replacement

In sampling without replacement, the inclusion probabilities for unselected sites change after each draw. The calculation of all inclusion probabilities for sampling without replacement is generally computationally prohibitive. For a summary of many unequal probability methods see Brewer and Hanif (1983) and Cochran (1977) We consider only one technique, the Rao, Hartley, Cochran method (Rao *et al.*, 1962) which has some desirable properties. The Rao, Hartley, Cochran (RHC) method randomly divides the population into n groups, indexed by g, where group g is of size N_g, $g = 1, \ldots, n$ and $\sum_{g=1}^{n} N_g = N$. It is recommended that N_g, $g = 1, \ldots, n$ be made as close to equal as possible. One unit is selected from each group with single draw inclusion probability which is proportional to the estimated probability, normed within the group. The single draw inclusion probability for site i within group g is

$$p_{g,i} = \frac{\hat{p}_i}{\sum_{j=1}^{N} \hat{p}_j t_{g,j}},$$

where $t_{g,j}$ is the indicator for inclusion of site j in group g. The number of observed occupied sites for RHC is equal to

$$D_{\text{RHC}} = \sum_{g=1}^{n} \sum_{j=1}^{N_g} t_{g,j} x_{g,j}, \tag{6.4}$$

where $x_{g,j}$ denotes presence or absence of site j in group g. The expected number of sample detections for RHC can then be determined using conditional expectation.

$$\begin{aligned} E(D_{\text{RHC}}) &= E\left[E\left(\sum_{i=1}^{n} \sum_{j=1}^{N_g} t_{g,j} x_{g,j} \mid x_{g,j} \right) \right] \\ &= \sum_{i=1}^{n} E\left\{ \sum_{j=1}^{N_g} x_{g,j} p_{g,j} \right\} \end{aligned}$$

$$
\begin{aligned}
&= \sum_{g=1}^{n} E\left\{\frac{\sum_{j=1}^{N_g} x_{g,j}\hat{p}_{g,j}}{\sum_{j=1}^{N_g}\hat{p}_{g,j}}\right\} \\
&= \sum_{g=1}^{n} E\left\{\hat{R}_g\right\} \\
&\approx n\sum_{i=1}^{N} x_i p_i. \qquad (6.5)
\end{aligned}
$$

where $\hat{R}_g$ is a ratio estimator in a SRS without replacement of size N_g from N. Hence this is a first order approximation of the expectation. The RHC unbiased estimator of the number of occupied sites is

$$\hat{X}_{\mathrm{RHC}} = \sum_{g=1}^{n}\sum_{i=1}^{N_g}\frac{x_{gi}}{p_{gi}}.$$

The variance of this estimator, for the case $N_g = N/n, g = 1,\ldots,n$, is

$$V(\hat{X}_{\mathrm{RHC}}) = \left(1 - \frac{n-1}{N-1}\right)\frac{1}{n}\left(\sum_{i=1}^{N}\frac{x_i^2}{p_i} - N^2\right).$$

Thus, sampling can be performed without replacement and estimation of variance can be calculated easily using only the first order single draw inclusion probabilities (Särndal *et al.*, 1991). Note that since $(n-1) < (N-1)$, $V(\hat{X}_{\mathrm{RHC}}) < V(\hat{X}_{\mathrm{HH}})$. Thus SPL without replacement, using the Rao, Hartley, Cochran method, always attains smaller estimation variance than SPL with replacement. Note also that sampling without replacement achieves greater expected unique detections.

6.2.3 Poisson Sampling

Poisson sampling (Särndal *et al.*, 1991, Chapter 3.5) selects sites by sampling list sequentially. Site i is selected with inclusion probability np_i where it is assumed that $np_i \leq 1$, or $p_i \leq 1/n$ for all sites. Since each site is independently selected or not selected with probability np_i, the number of sites selected is a random number.

The expected sample size is n. This is easily shown since the expected number sampled is the sum of all np_i,

$$\sum_{i=1}^{N} np_i = \sum_{i=1}^{N} n\frac{\hat{p}_i}{\sum_{j=1}^{N}\hat{p}_j} = n\frac{\sum_{j=1}^{N}\hat{p}_j}{\sum_{j=1}^{N}\hat{p}_j} = n.$$

The expected number of sample detections is the sum of the inclusion probabilities over sites with presence.

$$\mathrm{E}(D_{\mathrm{PO}}) = \sum_{i=1}^{N} x_i np_i.$$

and hence is approximately equal to that for SPL without replacement using RHC. An unbiased estimator of the number of occupied sites is

$$\hat{X}_{\mathrm{PO}} = \sum_{i=1}^{n} \frac{x_i}{np_i}.$$

The variance of the estimator of number of occupied sites using Poisson sampling for binary data is

$$\begin{aligned} V(\hat{X}_{\mathrm{PO}}) &= \sum_{i=1}^{N} np_i(1-np_i)\frac{x_i^2}{(np_i)^2} \\ &= \sum_{i=1}^{N}\left(\frac{1}{np_i}-1\right)x_i^2 \\ &= \sum_{i=1}^{N}\frac{x_i}{np_i} - X \end{aligned} \tag{6.6}$$

Särndal *et. al.* (1991) present an alternative to this estimator used with Poisson sampling, a ratio estimator of the number of occupied sites.

$$\hat{X}_{\mathrm{ALT}} = N\frac{\sum_{i=1}^{N}\frac{t_i x_i}{np_i}}{\sum_{i=1}^{N}\frac{t_i}{np_i}}. \tag{6.7}$$

They obtain an approximation for the variance of this ratio estimator from a general regression estimator framework as

$$V(\hat{X}_{\mathrm{ALT}}) \approx \sum_{i=1}^{N}\frac{x_i}{np_i} - \frac{2X}{N}\sum_{i=1}^{N}\frac{x_i}{np_i} + \frac{X^2}{N^2}\sum_{i=1}^{N}\frac{1}{np_i} - X + \frac{X^2}{N}.$$

The variances of these two Poisson sampling estimators compare as follows. $V(\hat{X}_{\text{ALT}}) < V(\hat{X}_{\text{PO}})$ if

$$\frac{\sum_{i=1}^{N} \frac{1}{np_i}}{N} < \left\{ 2\frac{\sum_{i=1}^{N} \frac{x_i}{np_i}}{X} - 1 \right\}. \tag{6.8}$$

This is true for instance if all sites with presence have the same inclusion probability, p_1, and all sites with absence have the same inclusion probability, p_0, with $p_1 > p_0$, and $X < N/2$. Thus the predicted probability of presence is greater for all presence sites and no more than half the sites are occupied. We could expect this situation, at least approximately, from the output from the autologistic model with covariates for sample data.

6.2.4 Comparison of SPL Methods

Next we compare the four SPL methods. Since the expected number of detections are approximately equal for RHC, PO and ALT, but is less for sampling with replacement, and all estimators are unbiased, we compare only the estimation variance obtained from each method. We note that $V_{\text{RHC}} < V_{\text{HH}}$ always. We compare the SPL methods to simple random sampling (SRS) with a ratio estimator as well. This estimator has the following form

$$\hat{X}_{\text{SRS-Ratio}} = \sum_{i=1}^{N} \hat{p}_i \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} \hat{p}_i}. \tag{6.9}$$

The variance of the ratio estimator for SRS (Cochran, 1977) is approximated by

$$V(\hat{X}_{\text{SRS-Ratio}}) \approx \frac{N-n}{n} \left[X - 2\frac{\sum_{i=1}^{N} x_i}{\sum_{i=1}^{N} \hat{p}_i} \sum_{i=1}^{N} x_i \hat{p}_i + \left(\frac{\sum_{i=1}^{N} x_i}{\sum_{i=1}^{N} \hat{p}_i} \right)^2 \sum_{i=1}^{N} \hat{p}_i^2 \right].$$

We make some general comparisons among variances and then make comparisons based on simulated populations.

First

$$\begin{aligned} V(\hat{X}_{\text{PO}}) - V(\hat{X}_{\text{HH}}) &= \sum_{i=1}^{N} \frac{x_i}{np_i} - X - \frac{1}{n}\left(\sum_{i=1}^{N} \frac{x_i}{p_i} - X^2\right) \\ &= -X + \frac{X^2}{n} \\ &= X\left(\frac{X}{n} - 1\right). \end{aligned}$$

If $X < n$, then $V(\hat{X}_{\text{PO}}) < V(\hat{X}_{\text{HH}})$. If instead, $X > n$, then $V(\hat{X}_{\text{HH}}) < V(\hat{X}_{\text{PO}})$. For rare populations, we might expect $X < n$ and would prefer PO over HH.

We also compare the PO variance to the RHC variance. Since for $X > n$, we already know that $V(\hat{X}_{\text{RHC}}) < V(\hat{X}_{\text{HH}}) < V(\hat{X}_{\text{PO}})$, we concentrate on the case $X < n$. Again we consider the difference in their variances.

$$V(\hat{X}_{\text{PO}}) - V(\hat{X}_{\text{RHC}}) = \frac{1}{n}\sum_{i=1}^{N} \frac{x_i}{p_i} - X - \left(1 - \frac{n-1}{N-1}\right)\frac{1}{n}\left(\sum_{i=1}^{N} \frac{x_i}{p_i} - X^2\right).$$

Assuming $\frac{n-1}{N-1} \approx \frac{n}{N}$, we find

$$\begin{aligned} V(\hat{X}_{\text{PO}}) - V(\hat{X}_{\text{RHC}}) &< \frac{1}{X}\sum_{i=1}^{N} \frac{x_i}{p_i} - X - \left(1 - \frac{n}{N}\right)\frac{1}{X}\left(\sum_{i=1}^{N} \frac{x_i}{p_i} - X^2\right) \\ &= \frac{1}{X}\sum_{i=1}^{N} \frac{x_i}{p_i} - X - \left(1 - \frac{n}{N}\right)\left(\frac{1}{X}\sum_{i=1}^{N} \frac{x_i}{p_i} - X\right) \\ &= \frac{1}{X}\sum_{i=1}^{N} \frac{x_i}{p_i} - X - \frac{1}{X}\sum_{i=1}^{N} \frac{x_i}{p_i} + X + \frac{n}{N}\left(\frac{1}{X}\sum_{i=1}^{N} \frac{x_i}{p_i} - X\right) \\ &= \frac{n}{N}\left(\frac{1}{X}\sum_{i=1}^{N} \frac{x_i}{p_i} - X\right). \end{aligned}$$

If $p_i = \frac{1}{X}$, for all i such that $x_i = 1$, then $\sum_{i=1}^{N} \frac{x_i}{p_i} = X^2$, and $V(\hat{X}_{\text{PO}}) = V(\hat{X}_{\text{RHC}})$. But, if $p_i \geq \frac{1}{X}$, for i such that $x_i = 1$, and strictly greater for some i, then in fact $\sum_{i=1}^{N} \frac{x_i}{p_i} < X^2$, and $V(\hat{X}_{\text{PO}}) < V(\hat{X}_{\text{RHC}})$. Thus, if the model predicts presence with fairly high probability for actual presence sites, the SPL Poisson sampling method may have smaller variance than RHC. Note that as X increases, the p_i generally decrease, and vice versa, so that in general $V(\hat{X}_{\text{PO}}) < V(\hat{X}_{\text{RHC}})$.

We make no other general comparisons for the SPL methods, but use numerical studies to compare their efficiency. In this simulation study we compare the variance of the estimators by varying three characteristics of the population. We use only three values for estimated probability of presence for simplicity: the values are 0.8, 0.5 and 0.1. The first population characteristic we consider is the relationship between species presence/absence, X, and the estimated probability, $\hat{p}$. We consider three scenarios of this relationship The first possibility is that species presence corresponds to large values of $\hat{p}$, 'good' prediction. The second possibility is that species presence corresponds to small values of $\hat{p}$, 'poor' prediction. The third possibility is that species presence is not related to $\hat{p}$. We discuss the simulation of the data for these three setups below.

The second characteristic we investigate is the distribution of the value of $\hat{p}$. We consider a lattice with 1600 sites. The first situation, denoted P1, we consider is equal number of sites with $\hat{p} = .8$, $\hat{p} = .5$, and $\hat{p} = .1$. The second situation, P2, assigns moderate estimated probabilities to 1200 sites, high estimated probabilities to 100 sites and low estimated probabilities to 300 sites. The third situation, P3, has low estimated probabilities at 1200 sites, moderate estimated probabilities at 300 sites and high estimated probabilities at 100 sites. These are summarized in Table 6.1.

The third population characteristic we consider is abundance of the species. We use three values for number of occupied sites, $X = 500, 200,$ and 50. We consider performance at three sample sizes n= 50, 250, and 400.

The data are generated as follows. The sites are assigned estimated probabilities according to one of the setups described above. For the good prediction scenario, presence is attached to sites sequentially from highest estimated probability to lower

estimated probability up to the abundance number specified. For the poor prediction scenario, presence is assigned sequentially to sites with lowest estimated probability to higher estimated probability up to the abundance number specified. Note that the order within estimated probability 'classes' is inconsequential since the values are equal. The third scenario has independence between estimated probability and presence, so that presence is attached to sites through random number generation. The reported variance is the mean of 100 runs of the random number generation.

The results are given in Tables 6.2, 6.3, and 6.4, for the three presence/estimated probability relationship scenarios. Bold values indicate the design and estimator with the smallest computed variance within a comparison group. Generally RHC is best, but for rare (X = 50) presence situations, RHC, PO, and ALT are very similar with PO better in half of the six cases of Table 6.2. If presence is related to small estimated probability (Table 6.3) or when presence is not at all related to estimated probability (Table 6.4) then SRS with the ratio estimator is best. For presence independent of estimated probability, with larger samples and rare presence, the other estimators approach the variance of SRS. Since we expect the model to perform well, so that presence is related to large values of $\hat{p}$, we recommend the RHC estimator. In cases where PO is better, the difference is minimal. We use the RHC method for the SPL sampling results in the rest of our investigations.

6.3 Stratified Sampling

Our population, or area of interest is believed to be heterogeneous with respect to presence of any particular species. The goals of this sampling design are to maximize detection of the species and provide an estimator of total number of occupied

Table 6.1: Definition of P1, P2, and P3 for SPL Comparisons. Number of Sites with Specified $\hat{p}$

Distribution of Estimated Likelihood		
P1	P2	P3
533: $\hat{p} = 0.8$	100: $\hat{p} = 0.8$	100: $\hat{p} = 0.8$
533: $\hat{p} = 0.5$	1200: $\hat{p} = 0.5$	300: $\hat{p} = 0.5$
534: $\hat{p} = 0.1$	300: $\hat{p} = 0.1$	1200: $\hat{p} = 0.1$

sites with minimum mean squared error. We use the estimated probability as auxiliary information in an attempt to define strata which are homogeneous with respect to the estimated probability. Stratification addresses both of our goals. Detection of species is maximized when the estimated probabilities correspond to presence of the species. The estimated probabilities have error which may result in decreased detection. Using a portion of the sampling effort in low estimated probability sites builds robustness into the sampling design. Definition of homogeneous strata also increases the efficiency of the estimator.

In this section we explore the elements of stratified sampling (Str). We first consider two methods for defining strata, a computational approach and a subjective one. We explore two sampling selection plans within strata, the SRS and RHC to see which makes more sense in our setup. We then consider various allocation of sample size methods, some involving only one goal, and many which consider a compromise of cost, detection, and efficiency goals. We make some comparisons of strata definition, sampling design, and allocation in the sections they are introduced and make overall comparisons at the end. The comparisons we make are with respect

Table 6.2: Variance of SPL Estimators: Presence Corresponds to Large $\hat{p}$

		P1			P2			P3		
Sample Size	Method	X=500	X=200	X=50	X=500	X=200	X=50	X=500	X=200	X=50
n=50	PO	4169	1668	417	6068	2108	394	5537	938	**169**
	RHC	**2034**	**1377**	**415**	**3816**	**1789**	**383**	**3318**	**692**	182
	ALT	6560	2050	441	5950	2169	406	6270	1374	200
	SRS	5156	2625	727	5156	2625	727	5156	2625	727
n=250	PO	1368	547	**137**	2127	723	**128**	1915	255	**38**
	RHC	**732**	**196**	149	**1374**	**644**	141	**1195**	**249**	65
	ALT	2418	715	147	2174	762	133	2302	445	51
	SRS	1856	945	262	1856	945	262	1856	945	262
n=400	PO	667	**267**	**67**	1142	377	**61**	1009	**84**	**5**
	RHC	**407**	275	83	**763**	**358**	79	**664**	138	36
	ALT	1382	381	74	1230	411	65	1310	212	13
	SRS	1031	525	145	1031	525	145	1031	525	145

Table 6.3: Variance of SPL Estimators: Presence Corresponds to Small $\hat{p}$

		P1			P2			P3		
Sample Size	Method	X=500	X=200	X=50	X=500	X=200	X=50	X=500	X=200	X=50
n=50	PO	36850	14740	3685	23640	14000	3500	17000	6800	1700
	RHC	32692	13640	3480	20300	12946	3307	14071	6191	1618
	ALT	18815	11854	3505	12540	11088	3318	10568	5771	1636
	SRS	**5156**	**2625**	**727**	**5156**	**2625**	**727**	**5156**	**2625**	**727**
n=250	PO	14440	5776	1444	9156	5480	1370	6500	2600	650
	RHC	11769	4910	1253	7308	4660	1190	5066	2229	583
	ALT	7320	4637	1373	4810	4330	1298	4021	2203	625
	SRS	**1856**	**945**	**262**	**1856**	**945**	**262**	**1856**	**945**	**262**
n=400	PO	8837	3535	884	5535	3350	838	3875	1550	388
	RHC	6538	2728	696	4060	940	661	2811	1238	324
	ALT	4446	2832	840	2877	2611	793	2381	1311	373
	SRS	**1031**	**525**	**145**	**1031**	**525**	**145**	**1031**	**525**	**145**

Table 6.4: Variance of SPL Estimators: Presence is Random with Respect to $\hat{p}$

		P1			P2			P3		
Sample Size	Method	X=500	X=200	X=50	X=500	X=200	X=50	X=500	X=200	X=50
n=50	PO	16117	6189	1500	11906	4481	1120	13610	5311	1277
	RHC	13166	5859	1506	9142	4264	1135	10809	4998	1351
	ALT	11047	5517	1573	8081	4066	1118	9313	4739	1237
	SRS	**5156**	**2625**	**727**	**5156**	**2625**	**727**	**5156**	**2625**	**727**
n=250	PO	6146	2355	570	4463	1673	418	5144	2004	481
	RHC	4740	2109	542	3291	1535	409	3891	1799	486
	ALT	4213	2102	600	3026	1521	418	3519	1791	466
	SRS	**1856**	**945**	**262**	**1856**	**945**	**262**	**1856**	**945**	**262**
n=400	PO	3654	1397	337	2602	970	242	3027	1178	282
	RHC	2633	1172	301	1828	853	227	2162	1000	270
	ALT	2504	1248	357	1762	885	243	2070	1054	273
	SRS	**1031**	**525**	**145**	**1031**	**525**	**145**	**1031**	**525**	**145**

to the expected number of sample detections and the variance of the estimator of the number of occupied sites.

6.3.1 Notation

We review the estimators and notation here to provide reference for the upcoming derivations and comparisons. We use h to denote the stratum, where H is the number of strata. For simple random sampling (SRS) within strata, we utilize the ratio estimator

$$\hat{X}_{\text{SRS}}^{\text{Str}} = \sum_{h=1}^{H} \left[\sum_{i=1}^{N_h} \hat{p}_{h,i} \frac{\sum_{i=1}^{n_h} x_{h,i}}{\sum_{i=1}^{n_h} \hat{p}_{h,i}} \right],$$

whose variance is

$$V(\hat{X}_{\text{SRS}}^{\text{Str}}) = \sum_{h=1}^{H} \frac{N_h - n_h}{n_h} \left[\hat{X}_{\text{SRS},h}^{\text{Str}} - 2 \frac{\sum_{i=1}^{N_h} x_{h,i}}{\sum_{i=1}^{N_h} \hat{p}_{h,i}} \sum_{i=1}^{N_h} x_{h,i} \hat{p}_{h,i} + \left(\frac{\sum_{i=1}^{N_h} x_{h,i}}{\sum_{i=1}^{N_h} \hat{p}_{h,i}} \right)^2 \sum_{i=1}^{N_h} \hat{p}_{h,i}^2 \right].$$

The other sampling plan used is the Rao, Hartley, Cochran method for sampling without replacement within strata which utilizes its corresponding estimator,

$$\hat{X}_{\text{RHC}}^{\text{Str}} = \sum_{h=1}^{H} \sum_{g=1}^{n_h} \sum_{i=1}^{N_{h,g}} \frac{x_{h,g,i}}{p_{h,g,i}},$$

and variance

$$V(\hat{X}_{\text{RHC}}^{\text{Str}}) = \sum_{h=1}^{H} \left(1 - \frac{n_h - 1}{N_h - 1} \right) \frac{1}{n_h} \left[\sum_{i=1}^{N_h} \frac{x_{h,i}^2}{p_{h,i}} - \left(\sum_{i=1}^{N_h} x_{h,i} \right)^2 \right].$$

6.3.2 Construction of Strata

We present two strategies for constructing strata and compare them using the expected number of sample detections and the estimation variance.

One strategy for constructing strata from Dalenius and Hodges (1959) was developed to obtain optimal estimation of a population total, which is in our case

the number of occupied sites, from SRS without replacement within strata. The estimator of number of occupied sites in this case is given by

$$D_{\text{SRS}}^{\text{Str}} = \sum_{h=1}^{H} \sum_{j=1}^{N_h} t_{h,j} x_{h,j}. \tag{6.10}$$

The method forms strata that are equally spaced on the scale of the cumulant square root of the density of an auxiliary variable. We use the estimated probability from the autologistic model with covariates for sample data as the auxiliary variable. The method is as follows. A frequency (F) and relative frequency (R.F.) distribution of the estimated probability are constructed. Next, the square root of the relative frequency is cumulatively summed (Cum$\sqrt{\text{R.F.}}$). The strata are formed so that the cumulant square root densities have approximately the same range within each stratum.

We demonstrate this method using Example 2 from Section 1.4.2. We display the relative frequency distribution, and formation of strata in Table 6.5. The columns for 3 Strata and 6 Strata display the partitioning into strata.

Another approach is to subjectively form strata which are homogeneous relative to estimated probability. We demonstrate using estimated probabilities for Example 2 to form strata. The histogram of estimated probabilities in Figure 1.6 seems to be either large (greater than 0.65) or small (smaller than 0.1). We therefore consider strata formed by,

$$\begin{aligned}
&\text{Stratum 1:} \quad 0 \quad &&< \hat{p} < \quad 0.1 \\
&\text{Stratum 2:} \quad 0.1 \quad &&< \hat{p} < \quad 0.65 \\
&\text{Stratum 3:} \quad 0.65 \quad &&< \hat{p} < \quad 1
\end{aligned}$$

We discuss sampling among strata, allocation of the sample within strata and then compare these two methods of stratification.

Table 6.5: Formation of Strata Using Cum$\sqrt{R.F.}$.

$\hat{p}$	F	R.F.	$\sqrt{\text{R.F.}}$	Cum$\sqrt{\text{R.F.}}$	3 Strata	6 Strata
0 - 0.03	489	0.306	0.553	0.553	1	1
0.03 - 0.05	615	0.384	0.620	1.173	1	2
0.05 - 0.10	151	0.094	0.307	1.480	2	3
0.10 - 0.15	45	0.028	0.168	1.648	2	3
0.15 - 0.20	24	0.015	0.122	1.770	2	4
0.20 - 0.25	20	0.013	0.112	1.882	2	4
0.25 - 0.30	15	0.009	0.097	1.979	2	4
0.30 - 0.35	13	0.008	0.090	2.069	2	4
0.35 - 0.40	20	0.013	0.112	2.181	2	4
0.40 - 0.45	16	0.010	0.100	2.281	2	5
0.45 - 0.50	14	0.009	0.094	2.374	3	5
0.50 - 0.55	8	0.005	0.071	2.445	3	5
0.55 - 0.60	11	0.007	0.083	2.528	3	5
0.60 - 0.65	11	0.007	0.083	2.611	3	5
0.65 - 0.70	16	0.010	0.100	2.711	3	5
0.70 - 0.75	21	0.013	0.115	2.826	3	6
0.75 - 0.80	30	0.019	0.137	2.963	3	6
0.80 - 0.85	22	0.014	0.117	3.080	3	6
0.85 - 0.90	19	0.012	0.109	3.189	3	6
0.90 - 0.95	0	0	0	3.189	3	6
0.95 - 1.00	40	0.025	0.158	3.347	3	6

6.3.3 Sampling Within Strata

We consider two methods for sampling within strata: RHC and SRS. SRS involves simple sample selection criteria and estimation formulas and is appropriate for homogeneous populations. We explore the possibility of using SRS within strata and the ratio estimator instead of RHC since we construct strata which form homogeneous groups of sites relative to the estimated probability. We first compare expected number of sample detections and estimation variance for SRS and RHC and then consider whether the assumption of homogeneity within strata is appropriate based on the examples in Chapter 4.

The expected number of sample detections using SRS is

$$\mathrm{E}(D_{\mathrm{SRS}}^{\mathrm{Str}}) = \sum_{h=1}^{H} n_h \frac{X_h}{N_h},$$

since the inclusion probability is n_h/N_h for all sites is stratum h. The estimator for occupied sites using RHC within strata is

$$D_{\mathrm{RHC}}^{\mathrm{Str}} = \sum_{h=1}^{H} \sum_{g_h=1}^{n_h} \sum_{j=1}^{N_{h,g}} t_{h,g,j} x_{h,g,j}. \tag{6.11}$$

where there are g_h groups in stratum h, $x_{h,g,j}$ is presence or absence of site j in group g_h, and $t_{h,g,j}$ is the indicator of a site being included in the group g_h. The expected number of sample detections using RHC, following the argument of Equation 6.5 is

$$\mathrm{E}(D_{\mathrm{RHC}}^{\mathrm{Str}}) \approx \sum_{h=1}^{H} n_h \sum_{i=1}^{N_h} x_{h,i} p_{h,i}.$$

where $p_{h,i} = \frac{\hat{p}_{h,i}}{\sum_{i=1}^{N_h} \hat{p}_{h,i}}$ is the single draw inclusion probability for site i within stratum h. The difference in the expected number of sample detections for the two sampling methods is

$$\begin{aligned} \mathrm{E}(D_{\mathrm{SRS}}^{\mathrm{Str}}) - \mathrm{E}(D_{\mathrm{RHC}}^{\mathrm{Str}}) &= \sum_{h=1}^{H} \left[n_h \frac{X_h}{N_h} - n_h \sum_{x_{h,i}=1} p_{h,i} \right] \\ &= \sum_{h=1}^{H} \left[n_h \sum_{i=1}^{N_h} x_{h,i} \left(\frac{1}{N_h} - p_{h,i} \right) \right]. \end{aligned}$$

If $p_{h,i} = p_h = 1/N_h$ for all h then the expected detections for SRS and RHC are equal. Consider the case where all sites with species presence have $p_{h,i} \geq 1/N_h$, with strict inequality for at least one site. This requires that all sites with species absence have $p_{h,i} \leq 1/N_h$. In this case the expected number of sample detections within stratum h are greater for RHC than for SRS. We expect $p_{h,i}$ to be large for presence sites for a good fitting model in the whole population but this relationship may not necessarily hold within each stratum.

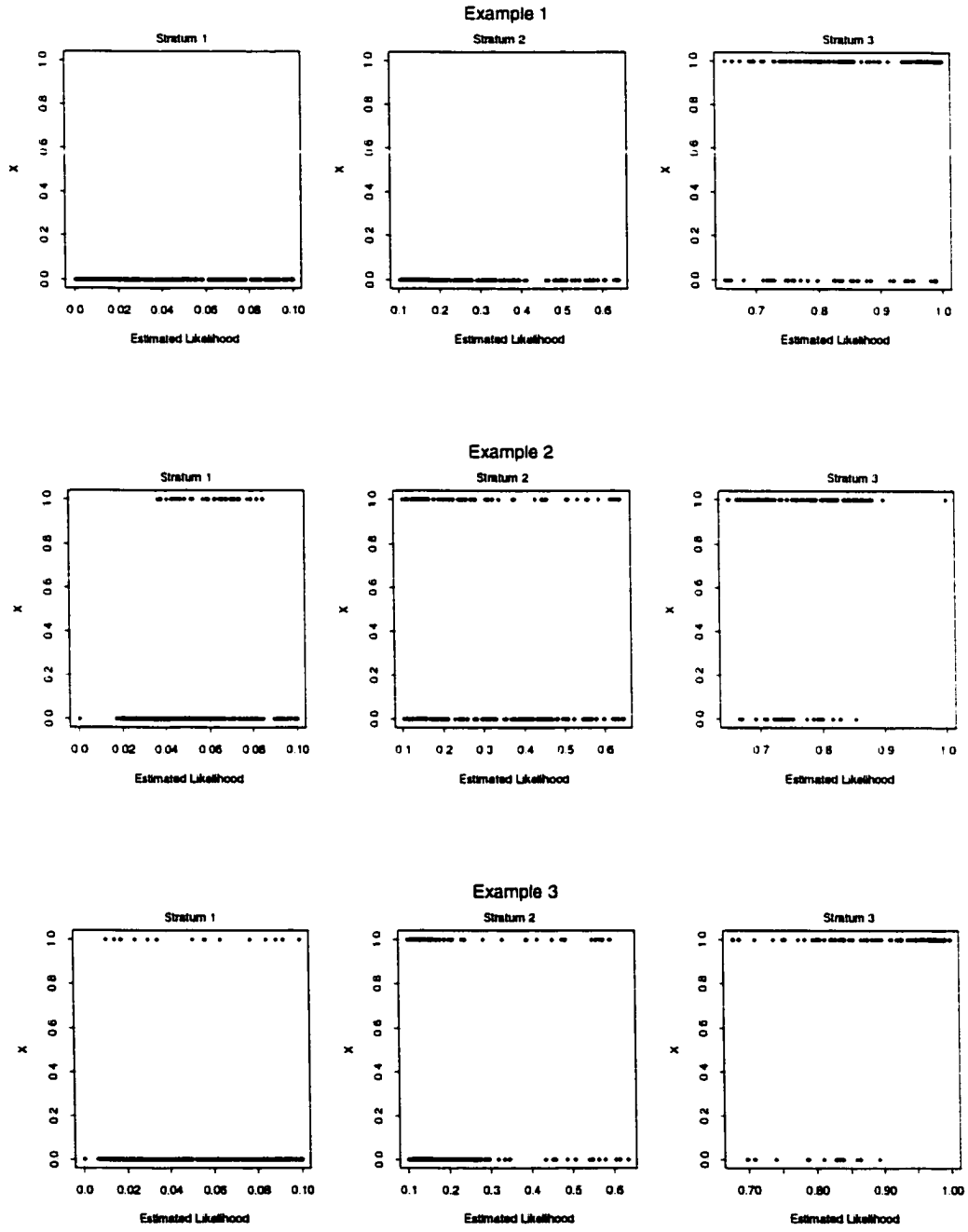
Based on the simulations in Tables 6.2, 6.3, and 6.4, SRS is better for the scenario of independence between presence and estimated probability but the difference is not great for rare species if sample size is large.

Now we investigate the validity of the assumption of homogeneity of the estimated probability within strata. We plot the estimated probabilities versus the true presence/absence for 3 strata in Examples 1, 2, and 3 (Figure 6.1). The strata were formed using the subjective cutoffs of 0.1, and 0.65. The estimated probabilities appear to be fairly random across the stratum range for the groups of presence and absence sites but are generally spread acorss the entire range from zero to one. This indicates heterogeneity within strata and thus we chose RHC within strata for further explorations of sampling designs.

6.3.4 Allocation of Sites to Strata

The sample size, which is generally limited by cost or other factors, is allocated among strata to fulfill various, often conflicting, goals. The goals of the current project are to maximize the number of detections, and to minimize the estimation variance. Allocation techniques are either simple allocation techniques to achieve one goal or a compromise between two or three goals. We present allocation methods and compare their performance.

Figure 6.1: Estimated Likelihood Versus True Presence/Absence for Examples 1, 2, 3.



6.3.4.1 Simple Allocation

We present three simple allocation methods; proportional allocation, Neyman allocation, and auxiliary variable allocation.

In proportional allocation, the total sample, n, is allocated to strata in proportion to the strata size. For stratum h, the sample size is

$$n_h = n\frac{N_h}{N},$$

for $h = 1, \ldots, H$. This technique allows for approximately equal coverage of the whole population.

Neyman allocation (Cochran, 1977) allocates the sample in order to minimize the variance of an estimator for a given sample size n. To minimize the variance of the estimator of the number of occupied sites, the Neyman allocation strategy results in stratum sample sizes

$$n_h = n\frac{\sqrt{\mathrm{Var}(\hat{X}_{n_h,h})}}{\sum_{h=1}^{H}\sqrt{\mathrm{Var}(\hat{X}_{n_h,h})}}.$$

for $h = 1, \ldots, H$, where $\hat{X}_{n_h,h}$ is some estimator of occupied sites for stratum h based on a sample of size n_h. For RHC within strata this is,

$$n_h = n\frac{\sqrt{\sum_{i=1}^{N_h}\frac{x_{h,i}^2}{p_{h,i}} - X_h^2}}{\sum_{h=1}^{H}\sqrt{\sum_{i=1}^{N_h}\frac{x_{h,i}^2}{p_{h,i}} - X_h^2}}. \tag{6.12}$$

Auxiliary variable allocation (Yadava and Singh, 1984) utilizes a variable related to the response of interest to partition the population into strata that are homogeneous within and heterogeneous among. The response of interest is presence of the species and the auxiliary variable is the estimated probability, $\hat{p}$. This

allocation results in stratum sample sizes of

$$n_h = n \frac{\sum_{i=1}^{N_h} \hat{p}_{h,i}}{\sum_{h=1}^{H} \sum_{i=1}^{N_h} \hat{p}_{h,i}}.$$

for $h = 1, \ldots, H$.

6.3.4.2 Compromise Allocation

We introduce allocation methods which achieve a compromise between two or more conflicting goals. The three goals we consider are to minimize estimation variance, maximize detection, and minimize cost.

We make comparisons using RHC sampling plan and estimator and introduce the following notation used in formulating allocations. The variance is denoted by.

$$\begin{aligned} V &= \mathrm{Var}(\hat{X}^{\mathrm{Str}}_{\mathrm{RHC}_n}) \\ &= \sum_{h=1}^{H} \frac{N_h - n_h}{N_h n_h} \left(\sum_{i=1}^{N_h} \frac{x_{h,i}}{p_{h,i}} - X_h^2 \right) \\ &= \sum_{h=1}^{H} \left[\frac{1}{n_h} \left(\sum_{i=1}^{N_h} \frac{x_{h,i}}{p_{h,i}} - X_h^2 \right) - \frac{1}{N_h} \left(\sum_{i=1}^{N_h} \frac{x_{h,i}}{p_{h,i}} - X_h^2 \right) \right] \\ &= \sum_{h=1}^{H} \frac{v_h^*}{n_h} - \sum_{i=1}^{H} \frac{v_h^*}{N_h}. \end{aligned} \tag{6.13}$$

Note that only the first term of V depends on allocation $n_h, h = 1, \ldots, H$. We assume that the variable cost due to sampling can be approximated as

$$C = \sum_{h=1}^{H} c_h n_h,$$

where c_h is the average cost per unit in stratum h. Expected number of sample detections is denoted by

$$E = E(D^{\mathrm{Str}}_{\mathrm{RHC}})$$

$$= \sum_{h=1}^{H} n_h \sum_{i=1}^{N_h} x_{h,i} p_{h,i}$$
$$= \sum_{i=1}^{H} n_h e_h,$$

where $e_h = \sum_{i=1}^{N_h} x_{h,i} p_{h,i} < 1$. Expected nondetections are denoted by

$$\begin{aligned} M &= n - E \\ &= \sum_{h=1}^{H} (n_h - n_h e_h) \\ &= \sum_{h=1}^{H} n_h (1 - e_h). \end{aligned}$$

There are four compromise allocation setups we consider. We begin with allocations which involve compromises between minimum variance and either minimum cost or maximum detection. In the second setup we minimize variance subject to fixed cost and detection, minimize cost subject to fixed variance and detection, and maximize detection subject to fixed variance and cost. The third setup is to find a compromise between two goals subject to the third being fixed. The final three goal compromise setup is to attempt to optimize all three, variance, cost and detection together.

The first allocation method minimizes a function of the goal quantities of interest (Cochran, 1977). For illustration with 2 goals of optimizing one goal subject to constraint on another consider the estimation variance, V, and the sampling cost, C. These two quantities are written as the product of two sums of squared terms. The Cauchy-Schwartz inequality is utilized to obtain a bound on the product. This bound is then solved for the allocation, n_h, which achieves the bound. Either of the constraints is then fixed to obtain the necessary total sample size, n.

We now consider in detail a compromise between V and M in the two goal compromise allocation. The allocation is obtained using the Cauchy-Schwartz inequality on the product of V and M, and solving for either one fixed.

The goal is to minimize the variance and non-detection. The product is

$$V \times M = \sum_{h=1}^{H} \left[\frac{1}{n_h} v_h^* - \frac{1}{N_h} v_h^*\right] \times \sum_{h=1}^{h} n_h(1 - e_h).$$

where $v_h^* = \sum_{i=1}^{N_h} \frac{x_{h,i}}{p_{h,i}} - X_h^2$ and $e_h = \sum_{i=1}^{N_h} x_{h,i} p_{h,i}$. The function to be minimized involves only those terms with n_h. So the function to be minimized is

$$\left(\sum_{h=1}^{H} \sqrt{\frac{v_h^*}{n_h}}^2\right) \left(\sum_{h=1}^{H} \sqrt{n_h(1 - e_h)}^2\right).$$

By Cauchy-Schwartz, the two goal compromise rule is

$$n_h = n \frac{\sqrt{\frac{v_h^*}{(1-e_h)}}}{\sum_{h=1}^{H} \sqrt{\frac{v_h^*}{(1 - e_h)}}}.$$

To determine sample size for a fixed goal, substitute n_h into formulas for V and M set at constants V_0 and M_0.

- For fixed variance, V_0,

$$n = \frac{\sum_{h=1}^{H} \sqrt{\frac{v_h^*}{(1 - e_h)}} \sum_{h=1}^{H} \sqrt{v_h^*(1 - e_h)}}{\left[V_0 + \sum_{h=1}^{H} \frac{1}{N_h} v_h^*\right]}.$$

- For fixed detections, M_0,

$$n = \frac{M_0 \sum_{h=1}^{H} \sqrt{\frac{v_h^*}{(1 - e_h)}}}{\sum_{h=1}^{H} \sqrt{v_h^*(1 - e_h)}}.$$

The corresponding approach and solution pertain to optimizing the product of V and C, which is the classic allocation problem (Cochran, 1977). We do not include it here, since it mirrors the above results.

The second allocation method is to optimize one goal while the other two are held fixed. This involves forming a function using two Lagrange multipliers, taking derivatives, and solving. We do this for each of the three goals being optimized, of which only optimizing for the estimation variance results in a direct solution. The other two might be solved numerically.

- Minimize V subject to $M = M_0$ and $C = C_0$

  The function to be optimized is

$$F(n_1, n_2, \ldots, n_H) = \sum_{h=1}^{H} \frac{v_h^*}{n_h} + \lambda_1 \left(\sum_{h=1}^{H} c_h n_h - C_0 \right) + \lambda_2 \left(\sum_{h=1}^{H} (1 - e_h) n_h - M_0 \right).$$

  We take the derivative with respect to n_h to get

$$\frac{\partial F(n_1, n_2, \ldots, n_H)}{\partial n_h} = -\frac{v_h^*}{n_h^2} + \lambda_1 c_h + \lambda_2 (1 - e_h).$$

  We set this equal to zero and solve for n_h to get

$$n_h = \sqrt{\frac{v_h^*}{\lambda_1 c_h + \lambda_2 (1 - e_h)}}.$$

  Since $\sum_{h=1}^{H} n_h = n$, we find the allocation rule to be

$$n_h = n \frac{\sqrt{\frac{v_h^*}{\lambda_1 c_h + \lambda_2 (1 - e_h)}}}{\sum_{h=1}^{H} \sqrt{\frac{v_h^*}{\lambda_1 c_h + \lambda_2 (1 - e_h)}}}. \tag{6.14}$$

  This is difficult to solve for λ_1 and λ_2 in general, so we solve for the special case $H = 3$. To solve for λ_1 we first substitute n_h into M_0.

$$
\begin{aligned}
M_0 &= (1-e_1)n_1 + (1-e_2)n_2 + (1-e_3)n_3 \\
M_0 &= (1-e_1)\sqrt{\frac{v_1^*}{\lambda_1 c_1 + \lambda_2(1-e_1)}} + (1-e_2)\sqrt{\frac{v_2^*}{\lambda_1 c_2 + \lambda_2(1-e_2)}} + \\
&\quad (1-e_3)\sqrt{\frac{v_3^*}{\lambda_1 c_3 + \lambda_2(1-e_3)}} \\
\sqrt{\frac{v_3^*}{\lambda_1 c_3 + \lambda_2(1-e_3)}} &= \frac{M_0 - (1-e_1)\sqrt{\frac{v_1^*}{\lambda_1 c_1 + \lambda_2(1-e_1)}} - (1-e_2)\sqrt{\frac{v_2^*}{\lambda_1 c_2 + \lambda_2(1-e_2)}}}{(1-e_3)}.
\end{aligned}
\tag{6.15}
$$

Now we substitute this into C_0.

$$
\begin{aligned}
C_0 &= c_1 n_1 + c_2 n_2 + c_3 n_3 \\
&= c_1\sqrt{\frac{v_1^*}{\lambda_1 c_1 + \lambda_2(1-e_1)}} + c_2\sqrt{\frac{v_2^*}{\lambda_1 c_2 + \lambda_2(1-e_2)}} + c_3\sqrt{\frac{v_3^*}{\lambda_1 c_3 + \lambda_2(1-e_3)}} \\
&= c_1\sqrt{\frac{v_1^*}{\lambda_1 c_1 + \lambda_2(1-e_1)}} + c_2\sqrt{\frac{v_2^*}{\lambda_1 c_2 + \lambda_2(1-e_2)}} \\
&\quad + c_3\left(\frac{M_0 - (1-e_1)\sqrt{\frac{v_1^*}{\lambda_1 c_1 + \lambda_2(1-e_1)}} - (1-e_2)\sqrt{\frac{v_2^*}{\lambda_1 c_2 + \lambda_2(1-e_2)}}}{(1-e_3)}\right) \\
&= \sqrt{\frac{v_1^*}{\lambda_1 c_1 + \lambda_2(1-e_1)}}\left(c_1 - \frac{(1-e_1)c_3}{(1-e_3)}\right) + \\
&\quad \sqrt{\frac{v_2^*}{\lambda_1 c_2 + \lambda_2(1-e_2)}}\left(c_2 - \frac{(1-e_2)c_3}{(1-e_3)}\right) + \frac{c_3}{(1-e_3)}.
\end{aligned}
\tag{6.16}
$$

Now substituting M_0 (6.15) and n_h (6.14) into $n = \sum_{h=1}^{H} n_h$ we get

$$
\begin{aligned}
n &= \sqrt{\frac{v_1}{\lambda_1 c_1 + \lambda_2(1-e_1)}} + \sqrt{\frac{v_2}{\lambda_1 c_2 + \lambda_2(1-e_2)}} + \sqrt{\frac{v_3}{\lambda_1 c_3 + \lambda_2(1-e_3)}} \\
&= \sqrt{\frac{v_1}{\lambda_1 c_1 + \lambda_2(1-e_1)}} + \sqrt{\frac{v_2^*}{\lambda_1 c_2 + \lambda_2(1-e_2)}}
\end{aligned}
$$

$$+\frac{M_0-(1-e_1)\sqrt{\frac{v_1^*}{\lambda_1 c_1+\lambda_2(1-e_1)}}-(1-e_2)\sqrt{\frac{v_2^*}{\lambda_1 c_2+\lambda_2(1-e_2)}}}{(1-e_3)}$$

$$= \sqrt{\frac{v_1^*}{\lambda_1 c_1+\lambda_2(1-e_1)}}\left(1-\frac{(1-e_1)}{(1-e_3)}\right)+\sqrt{\frac{v_2^*}{\lambda_1 c_2+\lambda_2(1-e_2)}}\left(1-\frac{(1-e_2)}{(1-e_3)}\right)+ \frac{M_0}{(1-e_3)}. \tag{6.17}$$

Now substituting (6.17) above into the C_0 function above (6.16). we get

$$C_0 = \sqrt{\frac{v_1^*}{\lambda_1 c_1+\lambda_2(1-e_1)}}\left(c_1-\frac{(1-e_1)c_3}{(1-e_3)}\right) + \left(\frac{n-\sqrt{\frac{v_1^*}{\lambda_1 c_1+\lambda_2(1-e_1)}}\left(1-\frac{(1-e_1)}{(1-e_3)}\right)-\frac{M_0}{(1-e_3)}}{1-\frac{(1-e_2)}{(1-e_3)}}\right)\left(c_2-\frac{(1-e_2)c_3}{(1-e_3)}\right)+\frac{c_3}{(1-e_3)}M_0$$

$$= \sqrt{\frac{v_1^*}{\lambda_1 c_1+\lambda_2(1-e_1)}}\left(c_1-\frac{(1-e_1)c_3}{(1-e_3)}-\frac{\left(1-\frac{(1-e_1)}{(1-e_3)}\right)\left(c_2-\frac{(1-e_2)c_3}{(1-e_3)}\right)}{1-\frac{(1-e_2)}{(1-e_3)}}\right) + \frac{n-\frac{M_0}{(1-e_3)}}{1-\frac{(1-e_2)}{(1-e_3)}}\left(c_2-\frac{(1-e_2)c_3}{(1-e_3)}\right)+\frac{c_3}{(1-e_3)}M_0. \tag{6.18}$$

Rearranging to solve for λ_1 we get

$$\frac{\lambda_1 c_1+\lambda_2(1-e_1)}{v_1^*} = \left[\frac{c_1-\frac{(1-e_1)c_3}{(1-e_3)}-\frac{\left(1-\frac{(1-e_1)}{(1-e_3)}\right)\left(c_2-\frac{(1-e_2)c_3}{(1-e_3)}\right)}{1-\frac{(1-e_2)}{(1-e_3)}}}{C_0-\left(\frac{n-\frac{M_0}{(1-e_3)}}{1-\frac{(1-e_2)}{(1-e_3)}}\right)\left(c_2-\frac{(1-e_2)c_3}{(e_3)}\right)-\frac{c_3}{(1-e_3)}M_0}\right]^2$$

$$\lambda_1 = \frac{\left[\frac{c_1-\frac{(1-e_1)c_3}{(1-e_3)}-\frac{\left(1-\frac{(1-e_1)}{(1-e_3)}\right)\left(c_2-\frac{(1-e_2)c_3}{(1-e_3)}\right)}{1-\frac{(1-e_2)}{(1-e_3)}}}{C_0-\left(\frac{n-\frac{M_0}{(1-e_3)}}{1-\frac{(1-e_2)}{(1-e_3)}}\right)\left(c_2-\frac{(1-e_2)c_3}{(1-e_3)}\right)-\frac{c_3}{(1-e_3)}M_0}\right]^2 v_1^*-\lambda_2(1-e_1)}{c_1}$$

$$= \frac{Rv_1^*-\lambda_2(1-e_1)}{c_1}. \tag{6.19}$$

In order to solve for λ_2 we first substitute λ_1 (6.19) into n (6.17).

$$n = \sqrt{\frac{v_1^*}{\lambda_1 c_1+\lambda_2(1-e_1)}}+\sqrt{\frac{v_2^*}{\lambda_1 c_2+\lambda_2(1-e_2)}}+\sqrt{\frac{v_3^*}{\lambda_1 c_3+\lambda_2(1-e_3)}}$$

$$
\begin{aligned}
&= \sqrt{\frac{v_1^*}{\left(\frac{Rv_1^* - \lambda_2(1-e_1)}{c_1}\right)c_1 + \lambda_2(1-e_1)}} + \sqrt{\frac{v_2^*}{\left(\frac{Rv_1^* - \lambda_2(1-e_1)}{c_1}\right)c_2 + \lambda_2(1-e_2)}} \\
&\quad + \sqrt{\frac{v_3^*}{\left(\frac{Rv_1^* - \lambda_2(1-e_1)}{c_1}\right)c_3 + \lambda_2(1-e_3)}} \\
&= \sqrt{\frac{v_1^*}{Rv_1^* - \lambda_2(1-e_1) + \lambda_2(1-e_1)}} + \sqrt{\frac{v_2^*}{R\frac{v_1^* c_2 - \lambda_2 c_2(1-e_1)}{c_1} + \lambda_2(1-e_2)}} + \\
&\quad \sqrt{\frac{v_3^*}{R\frac{v_1^* c_3 - \lambda_2 c_3(1-e_1)}{c_1} + \lambda_2(1-e_3)}} \\
&= \frac{1}{\sqrt{R}} + \sqrt{\frac{v_2^*}{\lambda_2\left((1-e_2) - \frac{(1-e_1)c_2}{c_1}\right) + R\frac{v_1^* c_2}{c_1}}} + \sqrt{\frac{v_3^*}{\lambda_2\left((1-e_3) - \frac{(1-e_1)c_3}{c_1}\right) + R\frac{v_1^* c_3}{c_1}}}.
\end{aligned}
\tag{6.20}
$$

Now substituting this into C_0 (6.18) and then using the fact that $n = n_1 + n_2 + n_3$ in the next step.

$$
\begin{aligned}
C_0 &= c_1 n_1 + c_2 n_2 + c_3 n_3 \\
&= \frac{c_1}{\sqrt{R}} + c_2\sqrt{\frac{v_2^*}{\lambda_2\left((1-e_2) - \frac{(1-e_1)c_2}{c_1}\right) + \frac{Rv_1^* c_2}{c_1}}} + c_3\sqrt{\frac{v_3^*}{\lambda_2\left((1-e_3) - \frac{(1-e_1)c_3}{c_1}\right) + \frac{Rv_1^* c_3}{c_1}}} \\
&= \frac{c_1}{\sqrt{R}} + c_2\left(n - \frac{1}{\sqrt{R}} - \sqrt{\frac{v_3^*}{\lambda_2\left((1-e_3) - \frac{(1-e_1)c_3}{c_1}\right) + \frac{Rv_1^* c_3}{c_1}}}\right) + \\
&\quad c_3\sqrt{\frac{v_3^*}{\lambda_2\left((1-e_3) - \frac{(1-e_1)c_3}{c_1}\right) + \frac{Rv_1^* c_3}{c_1}}} \\
&= \frac{c_1 - c_2}{\sqrt{R}} + c_2 n + (c_3 - c_2)\sqrt{\frac{v_3^*}{\lambda_2\left((1-e_3) - \frac{(1-e_1)c_3}{c_1}\right) + \frac{Rv_1^* c_3}{c_1}}}.
\end{aligned}
\tag{6.21}
$$

Now we can solve for λ_2

$$
C_0 = \frac{c_1 - c_2}{\sqrt{R}} + c_2 n + (c_3 - c_2)\sqrt{\frac{v_3^*}{\lambda_2\left((1-e_3) - \frac{(1-e_1)c_3}{c_1}\right) + \frac{Rv_1^* c_3}{c_1}}}
$$

$$
\left[\frac{C_0 - \frac{c_1 - c_2}{\sqrt{R}} - c_2 n}{c_3 - c_2}\right]^2 = \frac{v_3^*}{\lambda_2\left((1-e_3) - \frac{(1-e_3)(1-e_1)}{c_1}\right) + R\frac{v_1^* c_3}{c_1}}
$$

$$\lambda_2 = \frac{\frac{v_3^*}{\left[\frac{C_0 - \frac{c_1 - c_2}{\sqrt{R}} - c_2 n}{c_3 - c_2}\right]^2} - R\frac{v_1^* c_3}{c_1}}{(1 - e_3) - \frac{c_3(1 - e_1)}{c_1}}.$$

Thus for a 3 strata setup we have the exact solution. There does not seem to be a general rule which we can apply for H strata.

We now set up the scenario for the other two compromise allocations and show that the solution may only be obtained numerically.

- Maximize M subject to $V = V_0$ and $C = C_0$

The function to be optimized using Lagrange multipliers is

$$F(n_1, n_2, \ldots, n_H) = \sum_{h=1}^{H}(1 - e_h)n_h + \lambda_1\left(\sum_{h=1}^{H}\frac{v_h^*}{n_h} - V_0\right) + \lambda_2\left(\sum_{h=1}^{H} c_h n_h - C_0\right).$$

Again, taking the derivative with respect to n_h,

$$\frac{\partial F(n_1, n_2, \ldots, n_H)}{\partial n_h} = (1 - e_h) - \lambda_1\frac{v_h^*}{n_h^2} + \lambda_2 c_h.$$

Setting equal to zero and solving, as before, results in

$$n_h = n\frac{\sqrt{\frac{v_h^*}{(1-e_h)+\lambda_2 c_h}}}{\sum_{h=1}^{H}\sqrt{\frac{v_h^*}{(1 - e_h) + \lambda_2 c_h}}}. \qquad (6.22)$$

Although this seems a more simple equation than (6.14), the solution has no closed form even for the special case $H = 3$. To illustrate, we begin the substitution and simplification which should lead to the solution.

Based on the allocation rule,

$$
\begin{aligned}
n &= \sum_{h=1}^{H} n_h &&= \sum_{h=1}^{H} \sqrt{\frac{v_h^*}{(1-e_h)+\lambda_2 c_h}} \\
c &= \sum_{h=1}^{H} c_h n_h &&= \sum_{h=1}^{H} c_h \sqrt{\frac{v_h^*}{(1-e_h)+\lambda_2 c_h}} \\
M &= \sum_{h=1}^{H} (1-e_h) n_h &&= \sum_{h=1}^{H} (1-e_h) \sqrt{\frac{v_h^*}{(1-e_h)+\lambda_2 c_h}} \\
V &= \sum_{h=1}^{H} \frac{v_h^*}{\sqrt{\frac{v_h^*}{(1-e_h)+\lambda_2 c_h}}} &&= \sum_{h=1}^{H} \sqrt{v_h^* ((1-e_h)+\lambda_2 c_h)}
\end{aligned}
$$

Let us again consider the simpler case of $H = 3$. We first use n above to solve for a piece of the formula

$$
\begin{aligned}
n &= n_1 + n_2 + n_3 \\
&= \sqrt{\frac{v_1^*}{(1-e_1)+\lambda_2 c_1}} + \sqrt{\frac{v_2^*}{(1-e_2)+\lambda_2 c_2}} + \sqrt{\frac{v_3^*}{(1-e_3)+\lambda_2 c_3}}
\end{aligned}
$$

So that

$$
\sqrt{(1-e_3)+\lambda_2 c_3} = \left[\frac{\sqrt{v_3^*}}{n - \sqrt{\frac{v_1^*}{(1-e_1)+\lambda_2 c_1}} - \sqrt{\frac{v_2^*}{(1-e_2)+\lambda_2 c_2}}} \right].
$$

We can now substitute this into V_0

$$
\begin{aligned}
V_0 &= \sqrt{v_1^* ((1-e_1)+\lambda_2 c_1)} + \sqrt{v_2^* ((1-e_2)+\lambda_2 c_2)} + \\
&\quad \sqrt{v_3^*} \left[\frac{\sqrt{v_3^*}}{n - \sqrt{\frac{v_1^*}{(1-e_1)+\lambda_2 c_1}} - \sqrt{\frac{v_2^*}{(1-e_2)+\lambda_2 c_2}}} \right] \\
&= \sqrt{v_1^* ((1-e_1)+\lambda_2 c_1)} + \sqrt{v_2^* ((1-e_2)+\lambda_2 c_2)} + \frac{v_3^*}{n - \frac{\sqrt{v_1^*}}{(1-e_1)+\lambda_2 c_1}} - \\
&\quad \frac{\sqrt{v_2^*}}{\sqrt{(1-e_2)+\lambda_2 c_2}}.
\end{aligned}
$$

We now use C_0 to solve for another term to further simplify the formula.

$$\begin{aligned}
C_0 &= c_1 n_1 + c_2 n_2 + c_3 n_3 \\
&= c_1 \sqrt{\frac{v_1}{(1-e_1)+\lambda_2 c_1}} + c_2 \sqrt{\frac{v_2}{(1-e_2)+\lambda_2 c_2}} \\
&\quad + c_3 \left(n - \sqrt{\frac{v_1^*}{(1-e_1)+\lambda_2 c_1}} - \sqrt{\frac{v_2^*}{(1-e_2)+\lambda_2 c_2}} \right) \\
&= \sqrt{\frac{v_1^*}{(1-e_1)+\lambda_2 c_1}} (c_1 - c_3) \\
&\quad + \sqrt{\frac{v_2^*}{(1-e_2)+\lambda_2 c_2}} (c_2 - c_3) + c_3 n \\
C_0 - c_3 n - \sqrt{\frac{v_1^* (c_1 - c_3)^2}{(1-e_1)+\lambda_2 c_1}} &= \sqrt{\frac{v_2^*}{(1-e_2)+\lambda_2 c_2}} (c_2 - c_3) \\
\sqrt{(1-e_2)+\lambda_2 c_2} &= \frac{\sqrt{v_2^*}(c_2 - c_3)}{C_0 - c_3 n - \sqrt{\frac{v_1^*}{(1-e_1)+\lambda_2 c_1}} (c_1 - c_3)}.
\end{aligned}$$

Now we can substitute this back into V_0.

$$\begin{aligned}
V_0 &= \sqrt{v_1^* ((1-e_1)+\lambda_2 c_1)} + \frac{\sqrt{v_2^*}(c_2 - c_3)}{C_0 - c_3 n - \sqrt{\frac{v_1^*}{(1-e_1)+\lambda_2 c_1}} (c_1 - c_3)} \\
&\quad + \frac{v_3^*}{n - \sqrt{v_1^*}(1-e_1)+\lambda_2 c_1} \left(\frac{c_1 - c_3}{c_2 - c_3} - 1 \right) + n - \frac{C_0 - c_3 n}{c_2 - c_3}.
\end{aligned}$$

Although there is no algebraic solution, since only one Lagrange multiplier remains, it is possible to determine an appropriate value for λ_2 with respect to the other parameters. Next we sketch the other scenario.

- Minimize C subject to $M = M_0$ and $V = V_0$

The function to be minimized is

$$F(n_1, n_2, \ldots, n_H) = \sum_{h=1}^{H} c_h n_h + \lambda_1 \left(\sum_{h=1}^{H} \frac{v_h^*}{n_h} - V_0 \right) + \lambda_2 \left(\sum_{h=1}^{H} (1 - e_h) n_h - M_0 \right).$$

Taking the derivative with respect to n_h, we get

$$\frac{\partial F(n_1, n_2, \ldots, n_H)}{\partial n_h} = c_h - \lambda_1 \frac{v_h^*}{n_h^2} + \lambda_2 (1 - e_h).$$

Setting equal to zero and solving results in the following allocation rule.

$$n_h = n \frac{\sqrt{\frac{v_h^*}{c_h + \lambda_2(1 - e_h)}}}{\sum_{h=1}^{H} \sqrt{\frac{v_h^*}{c_h + \lambda_2(1 - e_h)}}}.$$

This looks extremely similar to the allocation rule for the previous scenario (Equation 6.22). Any attempt to solve for λ_2 would result in a similar situation, with no closed form.

The third method of compromise allocation is a three goal comparison. We consider the product of two factors and minimize subject to the third goal being fixed. We explore the only possible three goal compromise allocation. We accomplish the optimization by using Lagrange multipliers, taking derivatives, setting equal to 0 and solving for the Lagrange weight λ. The three goal compromise allocation rule might be obtained as follows.

- Minimize V and M subject to $C = C_0$

The function to minimize is

$$F(n_1, n_2, \ldots n_H) = \left(\sum_{h=1}^{H} \frac{v_h^*}{n_h} \right) \left(\sum_{h=1}^{H} n_h (1 - e_h) \right) + \lambda \left(\sum_{h=1}^{H} c_h n_h - C_0 \right).$$

We take the derivative with respect to n_h,

$$\frac{\partial F(n_1, n_2, \ldots, n_H)}{\partial n_h} = \frac{-v_h^*}{n_h} \left(\sum_{h=1}^{H} n_h (1 - e_h) \right) + \left(\sum_{h=1}^{H} \frac{v_h^*}{n_h} \right) (1 - e_h) + \lambda c_h.$$

By setting equal to zero, we attempt to solve for n_h. There is no direct solution for this problem.

The last approach to allocation with three conflicting goals is to optimize the function which is the product of the three goals,

$$F(n_1, n_2, \ldots, n_H) = V \times M \times C = \left(\sum_{h=1}^{H} \frac{v_h^*}{n_h}\right)\left(\sum_{h=1}^{H}(1-e_h)n_h\right)\left(\sum_{h=1}^{H} c_h n_h\right).$$

Taking the derivative results in

$$\begin{aligned}\frac{\partial F(n_1, n_2, \ldots, n_H)}{\partial n_h} &= -\frac{v_h^*}{n_h^2}\left(\sum_{h=1}^{H}(1-e_h)n_h\right)\left(\sum_{h=1}^{H} c_h n_h\right) \\ &\quad +(1-e_h)\left(\sum_{h=1}^{H}\frac{v_h^*}{n_h}\right)\left(\sum_{h=1}^{H} c_h n_h\right) + c_h\left(\sum_{h=1}^{H}\frac{v_h^*}{n_h}\right)\left(\sum_{h=1}^{H}(1-e_h)n_h\right).\end{aligned}$$

Since the n_h are all involved in sums, there will be no direct solution here either. Again, a numerical solution may be possible for specific design criteria.

6.3.4.3 Comparisons Among Allocation Methods

We use simulated populations to compare estimation variance and expected detections for the six allocation methods described above. The simple allocation methods are proportional, Neyman, and auxiliary variable. The compromise allocations are compromises for variance and detection, variance and cost, and minimum variance subject to constraints on detection and cost. The strata are determined according to a subjective rule which results in the cutoffs in the predicted probability of presence of 0.1, and 0.65. We compare the sample size requirements and allocations for various sampling levels, determined by compromise allocation between variance and cost and variance and detection.

There are three simulated populations. The population consists of 1600 sites. Each population has different estimated probabilities and presence patterns. The

actual values were generated according to Table 6.6. The Beta distributions for setups 1 and 3 are quite peaked for the top and lower strata and broad for the middle strata. The Beta distributions for setup 2 are uniform from 0 to 1. The binomial distributions provide for expected number present to be 200 overall. There were 1000 simulations run to solve for allocation and sample size.

The first population has those characteristics that the autologistic model with covariates is designed to model. About 150 of the presence sites are in the top stratum of size 500, 45 in the middle stratum of size 700 and 5 in the lower stratum of size 400. The estimated probability of presence matches well with the presence of species in the strata, high values for the top stratum, values of between 0.2 and 0.7 for the middle stratum and low values. The second population has the same presence setup, but the estimated probability is rather random with respect to presence. The third has evenly dispersed presence among the strata which have estimated probabilities as in setup one.

Two allocation methods depend on total sampling cost and sampling cost per strata. We used two levels of costs for the simulated populations. The first level was equal sampling cost in all strata, the second level defined a cost of 10 for the top stratum, 5 for the middle stratum, and 2 for the lower stratum.

The resulting sample size requirements are in Table 6.7. There are a few general findings from this simulation study. As the expected number of nondetections increases, so does sample size. Although this seems counter-intuitive, having a fixed number of nondetections *requires* the design to, at least on average, sample that number of absence sites. Notice that as the number of nondetections and required sample size increase, so do the number of detections. The expected detections is the required sample size minus the fixed nondetections.

Table 6.6: Simulation Setup for Comparison of Allocation Methods

Setup Number	Strata	Presence	Predicted Likelihood
1	1	Binomial (500, 0.3)	Beta (4, 0.5)
	2	Binomial (700, 0.064)	Beta (10,10)
	3	Binomial (400, 0.0125)	Beta (0.5, 4)
2	1	Binomial (500, 0.3)	Beta (1, 1)
	2	Binomial (700, 0.064)	Beta (1,1)
	3	Binomial (400, 0.0125)	Beta (1, 1)
3	1	Binomial (500, 0.125)	Beta (4, 0.5)
	2	Binomial (700, 0.125)	Beta (10,10)
	3	Binomial (400, 0.125)	Beta (0.5, 4)

Table 6.7: Sample Size Requirement for Compromise Between Variance and Detection for Simulated Populations

Fixed Nondetects	Setup1	Setup2	Setup3	Fixed Variance	Setup1	Setup2	Setup3
50	60	63	58	200	704	896	635
150	176	189	172	400	503	738	589
250	296	314	286	600	398	632	554
350	415	440	400	800	333	554	525
450	533	565	515	1000	288	494	505
550	652	691	629	1200	255	447	480
650	770	816	742	1400	229	408	462
750	888	942	858	1600	209	376	446

Table 6.8: Allocation for Simulated Populations, % of Sample Per Strata

Allocation Method	Setup1	Setup2	Setup3
Proportional	31, 44, 25	31, 44, 25	31, 44, 25
Neyman	41, 30, 29	56, 36, 8	10, 14, 76
Auxiliary	53, 42, 5	31, 44, 25	53, 42, 5
Variance - Nondetect	44, 29, 27	60, 33, 7	10, 14, 6
Variance - Cost (equal costs)	41, 39, 29	56, 36, 8	10, 14, 76
Variance - Cost (unequal costs)	34, 30, 36	52, 38, 10	7, 13, 80

Table 6.9: Allocation to Minimize Variance Subject to Cost and Detection for Simulated Populations for Equal Costs Among Strata

Fixed Nondetects	Fixed Cost	Allocation		
		Setup 1	Setup 2	Setup 3
250	400	45, 20, 36	30, 23, 47	49, 6, 4
	1500	72, 12, 16	40, 27, 33	54, 5, 41
	3000	69, 13, 18	57, 26, 17	57, 5, 38
450	400	43, 20, 37	31, 23, 46	49, 6 45
	1500	51, 19, 30	35, 25, 40	51, 5 44
	3000	69, 13, 18	40, 27, 33	53, 5, 42
650	400	42, 20, 38	30, 23 47	49, 6, 45
	1500	48, 19, 33	33, 25, 42	50, 5 45
	3000	56, 18 26	37, 26, 37	52, 5, 43

Table 6.10: Allocation to Minimize Variance Subject to Cost and Detection for Simulated Populations for Unequal Costs Among Strata

Fixed Nondetects	Fixed Cost	Allocation		
		Setup 1	Setup 2	Setup 3
250	400	30, 21, 49	21, 17, 62	36, 6, 58
	1500	32, 22, 46	23, 19, 58	38, 5, 57
	3000	36, 22, 42	24, 20, 56	40, 5, 55
450	400	30, 21, 49	21, 17, 62	36, 6, 58
	1500	31, 21, 48	22, 18, 60	37, 5, 58
	3000	33, 22, 45	22, 19, 59	38, 5, 57
650	400	30, 21, 49	21, 17, 62	36, 6, 58
	1500	31, 21, 48	22, 18, 60	37, 5, 58
	3000	32, 22, 46	22, 18, 60	38, 5, 57

The resulting allocations are in Tables 6.8, 6.9, 6.10. For large fixed cost (essentially no restriction), limiting the number of nondetections results in an allocation that favors the first strata which is more likely to have presence (Table 6.9).

Over all setups, the unequal costs among strata case (Table 6.10) puts less effort in the first strata (high estimated probability) and the various allocation rules in this case show very little variability. Hence the difference in cost seems to be the driving force in this case.

The underlying true population characteristics greatly effect the allocation rules and hence how any sampling design will perform. We emphasize the situation where the predictions of likely presence are good and presence is rare and compare it to only two other situations, but many more possibilities will exist in environmental data. In the first setup, which corresponds to a good fitting model, there is consistently

more effort allocated to the first strata (where there is more presence and higher predicted probabilities of presence). Also, for allocation methods which emphasize detection, the effort afforded this strata is the largest. For the other two setups, both representing a poor fitting model setup, there are no definite patterns in the allocation rules. Setup 2 corresponds to having most of the presence sites contained in the first stratum, so minimization of variance implies taking many samples there. But since the predictions don't match up with the presence sites, other allocation goals are difficult to meet, as is the case overall for setup 3.

6.3.5 Comparisons Among Strata Construction and Allocation Methods

We now compare stratification methods within the comparison of allocation methods. We restrict the comparison to Example 2 (Figure 4.4), using three strata and a fixed sample size of 180. We compare these methods based on the variance of the estimator and the expected detections. We use the RHC estimator and sampling without replacement with probability proportional to estimated probability of presence.

The stratification methods are the subjective stratification based on looking at histograms of estimated probabilities for presence and absence sites, and the cumulant method of Dalenius and Hodges (1959). We recall that the subjective stratification rule makes breaks at 0.1 and 0.65. The method of Dalenius and Hodges results in breaks at 0.05 and 0.45.

We use six allocation rules, three simple rules and three compromise rules. The simple methods are proportional allocation, Neyman allocation, and allocation based on an auxiliary variable, in our case the estimated probability of presence. The two goal compromise methods are allocation for compromise between variance

Table 6.11: Comparison of Variance and Detection for Stratification and Some Allocation Methods for Example 2

Allocation	Subjective Stratification		Cumulant Stratification	
	Variance	Detects	Variance	Detects
Proportional	4612	18	6858	21
Neyman	2671	24	2412	25
Auxiliary	71978	74	54561	83
V-M	2737	30	2490	31
V-C Equal Costs	2671	24	2412	25
V-C Unequal Costs	3057	15	2774	16

and detections (V-M), and allocation for compromise between variance and cost (V-C), both for the equal and unequal costs among strata setups. The three goal compromise allocation method is to minimize variance subject to fixed cost and detections (V-C-M), again for both equal and unequal costs among strata.

We find a few general comparisons hold overall based on our results (Tables 6.11, 6.12, 6.13). It is no surprise that method combinations which result in large variance also result in large expected detections. Also of no surprise is that the method which results in the largest expected detections is the auxiliary variable method, since estimated probability of presence is indeed closely related to actual presence. Also, that the Neyman allocation results in minimum variance, which is its goal. The compromise allocations result in various levels of variance and detection. So overall, the Neyman and auxiliary methods address each goal separately and the V-M and three goal compromises address both goals together.

Most allocation methods perform better, in terms of variance and detections, using the cumulant method of Dalenius and Hodges for stratification. There are 10

Table 6.12: Comparison of Variance and Detection for Stratification and V-C-M Allocation Methods for Equal Costs Among Strata for Example 2

Fixed Cost	Fixed Nondetects	Subjective Stratification		Cumulant Stratification	
		Variance	Detects	Variance	Detects
400	250	4653	56	2428	28
	450	3678	47	3196	46
	650	3540	46	4164	56
1500	250	2892	35	2441	29
	450	7936	69	2591	35
	650	11735	74	3332	48
3000	250	2768	19	2441	29
	450	2968	37	2511	32
	650	4215	52	2774	39

Table 6.13: Comparison of Variance and Detection for Stratification and V-C-M Allocation Method for Unequal Costs Among Strata for Example 2

Fixed Cost	Fixed Nondetects	Subjective Stratification		Cumulant Stratification	
		Variance	Detects	Variance	Detects
400	250	4758	31	4164	56
	450	2779	32	2774	39
	650	2779	32	3703	12
1500	250	3349	43	4164	56
	450	2929	36	8158	72
	650	2907	34	2737	38
3000	250	25732	81	4164	56
	450	3467	45	29952	85
	650	3084	39	3945	54

cases for various fixed values using the three compromise allocation of V-C-M, where the subjective approach is better for detections and 7 cases where it is better for variance. One other exception is that variance for proportional allocation is larger for cumulant stratification than for subjective stratification.

The results for the V-C-M compromise allocation need more interpretation for the fixed values and what they represent (Table 6.12, 6.13). The values used for fixed cost accomplish three goals. The lowest value of 400 is restrictive of sample size, especially for expensive strata in the unequal costs case. The middle value of 1500 is somewhat restrictive for the unequal case. The large value of 3000 is not restrictive at all. The nondetection level actually specifies detection. As discussed earlier, the larger the nondetection level, the larger also the detections. The nondetection level serves to increase the proportion of the sample allocated to the more likely stratum.

The values of cost and detection have great effect on the three goal compromise results. Many middle values of detection and variance can be obtained for both the equal and unequal costs setups. For example, for equal costs, medium values for both variance and detection can be obtained for fixed cost of 400 and fixed nondetects of 250 with subjective stratification, or either stratification for fixed nondetects of 450 or 650. These goals are difficult to specify without constraints related to a real problem. We present the method for comparison since in practice these parameters can be specified.

6.4 Fixed Top Stratum Sampling

In the comparison of stratification and allocation methods, we find that the top stratum contributes most to large expected detection as well as contributing a large part to the variance of the estimator. We can accomplish the goals of maximizing

detection in this high probability stratum and eliminate its influence on variance at the same time by specifying a top strata and sampling all of its members. We call this design fixed top stratum sampling (Fst).

The fixed top stratum sampling plan is designed to maximize detection by forming a 'top' stratum which contains sites with the highest estimated probability and performing a census there. We then sample the rest of the population with the remaining sample resources.

We restrict out attention to the case H=3 for simplicity. Application to more strata is straightforward. We use the RHC estimator and sampling without replacement in order to compare with stratified designs from the previous section. We use stratified sampling for the remainder of the population based on Neyman allocation to minimize variance. We begin by deriving an estimator for expected number of sample detections and the variance of the estimator of number of occupied sites. We then describe the design of the sample plan by investigating the construction of the top stratum. We finally compare design alternatives using Example 2 data.

6.4.1 Estimation and Variance

Using the result in Equation 6.4 the number of detections for the Fst design is

$$D_{\mathrm{RHC}}^{\mathrm{Fst}} = X_1 + \sum_{g=1}^{n_2} \sum_{j=1}^{N_{2,g}} t_{2,g,j} x_{2,g,j} + \sum_{g=1}^{n_3} \sum_{j=1}^{N_{3,g}} t_{3,g,j} x_{3,g,i}. \tag{6.23}$$

where there are g_h groups in stratum h. $x_{h,g,j}$ is presence or absence of site j in group g_h, and $t_{h,g,j}$ is the indicator of a site being included in the group g_h.

Using further the expectation from Equation 6.5, the expected number of sample detections is

$$\mathrm{E}(D_{\mathrm{RHC}}^{\mathrm{Fst}}) = X_1 + n_2 \sum_{i=1}^{N_2} x_{2,i} p_{2,i} + n_3 \sum_{i=1}^{N_3} x_{3,i} p_{3,i}. \tag{6.24}$$

where n_2 and n_3 are determined from Neyman allocation with sample size $n - N_1$, from Equation 6.12

$$n_h = (n - N_1)\frac{\sqrt{\sum_{i=1}^{N_h} \frac{x_{h,i}}{p_{h,i}} - X_h^2}}{\sum_{h=2}^{3} \sqrt{\sum_{i=1}^{N_h} \frac{x_{h,i}}{p_{h,i}} - X_h^2}}. \tag{6.25}$$

for $h = 2, 3$.

The estimator for number of occupied sites is

$$\hat{X}_{\mathrm{RHC}}^{\mathrm{FSt}} = N_1 + \sum_{g=1}^{n_2} \frac{x_{2,g,i}}{p_{2,g,i}} + \sum_{g=1}^{n_3} \frac{x_{3,g,i}}{p_{3,g,i}}.$$

with variance,

$$V(\hat{X}_{\mathrm{RHC}}^{\mathrm{FSt}}) = \sum_{h=2}^{3} \left(1 - \frac{n_h - 1}{N_h - 1}\right) \frac{1}{n_h} \left(\sum_{i=1}^{N_h} \frac{x_{h,i}}{p_{h,i}} - X_h^2\right).$$

Note that the expected number of detections includes *all* sites with presence from the top stratum, the sample allocation for the lower strata result from the 'leftover' sample size $n - N_1$, and the estimation variance from the top stratum is zero.

6.4.2 Design Construction

Construction of the top stratum determines the overall utility of the sample design since improper definition of a top stratum may enable presence sites to be missed by sampling too few with large estimated probabilities, or result in under-sampling of the rest of the population by choosing a top stratum that is too large.

We investigate two methods for defining the top stratum. The first is to specify a percent of the sample to devote to those sites which have the highest estimated probability of presence. We consider using either 50% of the sample or 80% of the sample. The second method for constructing the top stratum is to define a cutoff

such that all sites with estimated probability greater than the cutoff are sampled with probability one. We consider using the top stratum defined by Dalenius and Hodges' cumulant rule and an arbitrary value of, say, 0.99.

We compare results of these methods based on the data in Example 2 from Chapter 4 where the estimated probability is closely related to species presence (Table 6.14). We find that using the stratification cutoffs from a stratified sample results in large variance, and cannot always be carried out due to the size of the top stratum and sample size. We also notice that the method of determining the top strata is virtually unimportant for larger sample sizes. So the method only matters for smaller sample sizes and for this example, 50% of the sample is the best way to allocate resources to the sites with highest estimated probability.

6.5 Comparison of Design Performance with Three Modeling Methods

In this section we use the designs which have shown themselves to perform well and compare them with respect to the three estimation approaches from three models to predict probability of presence. We make these comparisons for the three examples from Chapter 4.

Using the predictions from the examples in Chapter 4, we calculated the variance and expected detections for the various designs and estimators (Table 6.15). Note that for each example, the optimal design parameters were chosen based on the specific design. For instance, the cumulant stratification rule of Dalenius and Hodges was calculated for each model for each example.

In general, the predictions from the autologistic model with covariates for sample data formed a better basis for sampling than did the logistic regression or the autologistic model. The variance of the estimator was smaller and the expected

Table 6.14: Comparison of Variance and Detection for Three Top Stratum Designs

Design	n	n1	n2	n3	Variance	Detects
Top 50%	180	90	52	38	555	92
Top 80%		128	28	24	862	102
$\hat{p} > 0.45$		371				
$\hat{p} > 0.99$		117	38	25	735	98
Top 50%	400	90	179	131	100	181
Top 80%		128	146	126	102	185
$\hat{p} > 0.45$		371	8	21	665	200
$\hat{p} > 0.99$		117	169	114	95	182
Top 50%	500	90	236	174	55	221
Top 80%		128	199	173	54	223
$\hat{p} > 0.45$		371	34	95	119	213
$\hat{p} > 0.99$		117	229	154	47	220

detections were larger. This does not hold for the SPL design. Since the predictions from the logistic regression are spread more evenly over the range from zero to one, the variance will be less than when the predictions are either very high or very low as in the autologistic model with covariates for sample data. Also for example 2, the variance for SRS and Strat-Neyman are very close.

The best design overall is the fixed top stratum sampling design. This design had the highest detections for all examples and models, and the smallest variance for four of the seven examples and models. The designs which had lower variance were the stratified design using Neyman allocation for example 1 predictions from the autologistic model with covariates for sample data and simple random sampling for examples 2 and 3 predictions from the logistic regression model.

6.6 Conclusions

After making many interim comparisons, we find that a new design achieves both of the goals in surveying for rare species, fixed top stratum sampling. This design which forms a census of sites with the highest estimated probability of presence, achieves both the most detections and also generally achieves the smallest or at least a small variance of the estimator.

Table 6.15: Comparison of Sampling Designs Based on Predictions From the Three Models in Chapter 4

Design	Estimator	Model	Example 1		Example 2		Example 3	
			Variance	Detects	Variance	Detects	Variance	Detects
SRS	Ratio	Log. Regr.	1197	27	1230	27	1230	27
		Autolog.	1715	27	2100	27	588	27
		Autolog. Cov.	707	27	1497	27	588	27
Strat. Neyman	RHC	Log. Regr.	2038	28	2528	27	2528	27
		Autolog.	12932	7	10900	8	3355	20
		Autolog. Cov.	48	126	2412	25	3355	20
Strat. Auxiliary	RHC	Log. Regr	4321	56	4131	32	4131	32
		Autolog.	83946	34	74927	30	8512	76
		Autolog. Cov.	82	101	93679	84	8512	76
Top Stratum	RHC	Log. Regr.	713	104	1466	62	1466	62
		Autolog.	4219	60	3074	59	955	109
		Autolog. Cov	81	149	527	95	955	109
SPL	RHC	Log. Regr.	143519	19	16093	8	16093	8
		Autolog.	3000000	10	28000000	12	31000000	16
		Autolog. Cov.	7000000	24	23000000	21	31000000	16

Chapter 7

CONCLUSIONS

This work has achieved two goals: development and initial verification of a new spatial model for binary data and investigation of sampling designs based on the output from our improved model. We derived the model within a Bayesian framework, and developed an estimation procedure. We have shown that the model improves upon both the non-spatial logistic regression model and the spatial non-covariate autologistic model. We also extended the model to account for a general issue of detectability. Through examples we have shown some improvements with this extension. Through investigations of sampling designs, we discover that information from a good fitting model can be used to improve detection and to decrease variance. We also discover which designs perform better, with respect to various criteria, under differing circumstances. We conclude this work with a summary of these three main topics and our ideas for future work based on the current progress.

7.1 Autologistic Model with Covariates for Sample Data

The autologistic model with covariates for sample data uses three types of information to predict presence or absence over a lattice. The three types of information are the observed presence or absence at sampled sites, the covariate information which is available for all lattice grid areas, and sample inclusion information for all lattice grid areas. We make some assumptions which primarily define the spatial

relationship among lattice grid areas, namely the Markov random field assumption. This assumption defines that the spatial relationship is limited to a neighborhood of a site. This is reasonable since we *a priori* assume that the species of interest were found in clusters.

These sources of information are utilized in a Bayesian framework to produce estimates of predicted probability of presence over the lattice. The setup of the Bayesian model follows the framework of Heikkinen and Högmander (1994, HH hereafter). Our model is an extension of this which accounts for covariates related to species presence and also accounts for the observation data which are from a sample rather than a census of the area of interest. The extension using sample data required a modification of the likelihood function and the extension for covariates required a more complex set of prior distributions.

This complex model requires alternative means of estimation. We develop a hybrid Gibbs sampling algorithm to sample from all full conditional likelihood functions to obtain these estimates. The algorithm is extended from HH to account for more complicated sampling of the estimated likelihood functions. Sampling for $\underline{x}$, and $\underline{p}$, the predicted presence/absence status and the predicted probability respectively, involve a more complex function as well as an improved neighborhood characterization, $\underline{s}$.

We use simulated setups in lieu of any real data examples, since the USFS data was not available. The three setups we presented represent common data issues that would come up in real data; limited sample data, and covariate data which although correlated with species presence, may not be consistently related over the area of interest. In the overall study, we used two sampling plans, simple random and a systematic grid sample of clusters, two coverages, partial and whole,

and large and small sample sizes. Although cluster sampling provides improved neighborhood information over simple random sampling, simple random sampling provides for more areal coverage. This increased coverage of the area of interest leads to an improvement in discriminating presence and absence, since true absence is known more evenly over the area of interest. We presented three examples which represent differing types of data and are representative of all examples. Based on these simulations, we find that our autologistic model with covariates for sample data improves on both the logistic model and the autologistic model without covariates. Both sensitivity and specificity are improved using our model as well as separation of predicted probability for presence and absence sites as displayed in histograms. We assured convergence of the estimation procedure through various indicators.

7.2 Detectability

The autologistic model with covariates for sample data uses two levels of detectability for the sample data. If the site was sampled, then probability of detection was one and if the site was not sampled, the probability of detection was zero. In many applications this is not general enough. Data sets often contain information related to detectability of the species of interest. This information may be in the form of habitat values such as percent ground cover or experience of the site observer. The inclusion of this information is a further improvement of our model. This added information can be used to predict if an unseen species might actually be there but was missed due to thick vegetation or inexperience of an observer.

The addition of the detectability term adds much to our ability to predict presence and absence without unduly complicating the model. The detectability, in the functional form of a logit function of covariates, enters the likelihood function.

The likelihood function is now of the data given the truth and these detection variables. Although this complicates the likelihood function, the estimation procedure is not unduly complicated by its inclusion. For each iteration of the Gibbs-Hastings sampler, the step to estimate presence/absence and estimated likelihood of presence is changed and there are additional steps required to estimate the parameters associated with the detection variables.

We demonstrate with two simulated examples the utility of the extension. The examples are set up to be similar to two from the autologistic model with covariates for sample data. The sensitivity of the extended model is greater when using detectability, but the specificity drops slightly. This may be a product of the examples used which have only a small amount of sample data in which to estimate parameters associated with detection.

7.3 Sampling Designs

The production of maps with predicted probability of presence is interesting, but is also a very useful tool for developing sampling plans and monitoring programs. The predicted probabilities can be used toward this end in many ways. We explore three ways to use the information in developing sampling plans. The first possible use of the information is to select sites to sample based on the predicted probabilities in an unequal probability sampling design. Another possibility is to use the predicted probabilities to stratify the population into more homogeneous sections. The final use we explore is to use the information only in the estimation of occupied sites and use a simple random sample to select sites. We'll summarize the two sample design uses and their various elements and then summarize the comparisons we've made on their performance.

Sampling with probability proportional to predicted probability (SPL) can be accomplished in at least three different ways. Selection of samples can be achieved with replacement, without replacement, and using Poisson sampling with unspecified sample size. We considered at least one estimator for each of the three selection strategies and compared their performance. The design and estimator that performed at least as well as all others under the good fitting model setup was sampling without replacement following the Rao, Hartley, Cochran (Rao *et al.*, 1962) selection procedure with the accompanying estimator (RHC).

Stratified sampling involves specifying three elements: how the strata are formed, how to sample within strata, and how to allocate the sample among strata. We suspect that using a stratified sample will achieve three goals. The first is to maximize detection by forming subpopulations according to predicted probability. The stratum with the highest predicted probability should contain the most species presence, so that sampling there intensely will maximize species detection. The second goal is to minimize variance. Variance estimates from stratified sampling, being the sum of variance estimates of more homogeneous regions, should be less than SPL variance estimates. The third goal is to verify the model which produces the estimated likelihoods. In sampling some from sites which have medium and low predicted probabilities, we ensure that a poor fitting model does not lead us totally astray and we also have information to verify the utility of the model.

We compare two methods of stratification. The method of Dalenius and Hodges (Dalenius and Hodges, 1959) specifies strata that have equal ranges on the cummulant square root of the relative frequency of an auxiliary variable. Another technique we tried was to subjectively determine the strata for a simulated example based on histograms of predicted probabilities for presence and absence sites. The method of

Dalenius and Hodges performed better overall than the subjective approach. This is somewhat unexpected since we used the true data to determine the strata for the subjective approach. This result is favorable though, since the subjective approach it is not easily applied in practice when the truth is unknown.

We considered two sample selection methods for within strata, simple random sampling with the ratio estimator and RHC sampling and estimation. We suspect that even stratification will not produce homogeneous subpopulations in variable environmental examples. Thus a simple random sample, even using the ratio estimator, may not perform as well as selecting sample sites according to the predicted probability. We used the RHC approach and estimator for our other comparisons.

There are numerous allocation strategies to consider and we make no recommendations on one best rule. Allocation rules we consider give differing weight to three goals. The three goals are maximization of detection, minimization of variance, and minimization of cost. There are allocation rules which optimize one of the goals, two goals, and three goals. The compromises between two of the goals may also involve optimization subject to constraints on the uninvolved goal. Allocations which minimize variance tend to also minimize detection, and, inversely, allocations which maximize detections tend to result in large variance. There are some allocation rules which result in average values for both detection and variance. These parameters need to be specified on a case by case basis. We demonstrate that certain complicated allocation rules can be worked out, at least in the three strata case, and that others cannot be determined in general at all.

We also considered a modified stratified sampling design we term fixed top stratum sampling. This design takes a census of a stratum defined to have the highest predicted probabilities and samples the rest of the population. The three

goals here are the same as in a stratified sample. The goals are to maximize detection, minimize variance and develop a sampling design that will be both robust to a faulty model and useful for model verification. In taking a census of the top strata, we maximize detection, as long as we have a good fitting model and reduce variance, since there is no sampling variance associated with a census. By sampling the rest of the population we have a design that can be used for model verification and is also robust. Using an example from Chapter 4, we compare one element of this design, specification of the top stratum, a stratum of the sites with the largest predicted probability of presence. The rest of the population we sample using two strata formed by the method of Dalenius and Hodges, allocation based on Neyman allocation, and RHC selection and estimation within strata.

The specification of the top strata really determines the utility of the design. We compare two top stratum determination rules at two levels each. One rule is to specify that all sites with predicted probability greater than some cutoff be assigned to the top stratum. The other rule specifies a percent of the total available sample to be assigned to the top stratum. Our comparisons show that using 50% of the sample in the top stratum is most efficient for detection and precision.

7.4 Model Results and Sampling Design Comparison

Finally we compare the utility of the three models, logistic regression, autologistic model, and autologistic model with covariates for sample data, for their use in sampling designs. The results from the examples in Chapter 4 are used for each of the three models. The sampling designs we consider are first, sampling proportional to predicted probability using the RHC approach and estimator, second, stratified

sampling using stratification rule of Danlenius and Hodges, RHC sampling and estimation within strata, and both Neyman and auxiliary variable allocation, third fixed top stratum sampling using 50% of the sample in top stratum, and fourth simple random sampling with the ratio estimator. We find that the autologistic model with covariates for sample data produces better results as far as detection and precision than either the logistic regression or autologistic model over all sampling designs. We also find that the fixed top stratum sampling design generally improves detection and precision over other designs.

7.5 Future Work

In our progression through this work, we have come up with many leads into further research. The autologistic model with covariates for sample data is a basic model which is open for extensions. The first extension we would consider is to multiple species. As much as species presence is related to habitat covariates and has spatial dependence, species depend on each other. The presence of certain other species may promote or inhibit the presence of a species of interest. It may be useful to model two species together if both are of equal interest. Species may have similar habitat requirements or very different, they may live well together or not. These relationships will certainly improve prediction of presence for both species upon predicted presence for the species individually.

The second extension to the model is to incorporate a temporal element. The model was developed to be part of a monitoring program, which will be carried out over years. The answer to "How the species is doing?" involves both its range and the trend in abundance and range over time.

In that same vein, the utility of a Bayesian framework is to continually incorporate newfound information. We would like to explore the improvements that take place in prediction when better prior information is available.

One extension to sampling designs that we haven't explored yet is adaptive sampling (Thompson and Seber, 1996) In adaptive sampling, when a site is found to have a species present, neighboring sites are also sampled. Thus if a species is found, the whole cluster of it will be found and counted. This should improve estimation of abundance as well as spatial extent. Also the information pertaining to the edges of species presence will be very valuable in future models.

Bibliography

Arnold, B. and Strauss, D. (1991). Pseudolikelihood estimation: some examples. *The Indian Journal of Statistics*, 53:233–243.

Augustin, N., Mugglestone, M., and Buckland, S. (1996). An autologistic model for the spatial distribution of wildlife. *Journal of Applied Ecology*, 33:339–347.

Augustin, N., Mugglestone, M., and Buckland, S. (1998). The role of simulation in modeling spaially correlated data. *Environmetrics*, 9:175–196.

Besag, J. (1972). Nearest-neighbor systems and the auto-logistic model for binary data. *Journal of the Royal Statistics Society, Series B*, 36:75–83.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistics Society, Series B*, 36:192–236.

Besag, J. (1986). On the statistical analysis of dirty pictures. *Journal of the Royal Statistics Society, Series B*, 48:259–302.

Besag, J. (1989). Towards Bayesian image analysis. *Journal of Applied Statistics*, 16:395–407.

Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43:1–20.

Brewer, K. and Hanif, M. (1983). *Sampling with Unequal Probabilities*. Springer-Verlag, New York.

Cochran, W. G. (1977). *Sampling Techniques*. John Wiley & Sons, New York, NY.

Cox, D. and Hinkley, D. (1974). *Theoretical Statistics*. Chapman and Hall, London.

Cressie, N. A. (1993). *Statistics for Spatial Data*. John Wiley & Sons, Inc.

Dalenius, T. and Hodges, J. (1959). Minimum variance stratification. *Journal of the American Statistical Association*, 54:88–101.

Gelman, A., Roberts, G., and Gilks, W. (1995). Efficient Metropolis jumping rules. In *Bayesian Statistics 5*. Oxford University Press, New York.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741.

Geyer, C. and Thompson, E. (1992). Constrained monte carlo maximum likelihood for dependent data (with discussion). *Journal of the Royal Statistical Society, Series B*, 54:657–699.

Gray, A. J., Kay, J., and Titterington, D. (1992). On the estimation of noisy binary markov random fields. *Pattern Recognition*, 25:749–768.

Gumpertz, M. L., Graham, J. M., and Ristaino, J. B. (1997). Autologistic model of spatial pattern of Phytophthora epidemic in bell pepper: Effects of soil variables on disease presence. *Journal of Agricultural, Biological, and Environmental Statistics*, 2:131–156.

Hansen, M. and Hurwitz, W. (1943). On the theory of sampling from finite populations. *Annals of Mathematical Statistics*, 14:333–362.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97–109.

Heikkinen, J. and Högmander, H. (1994). Fully Bayesian approach to image restoration with an application in biogeography. *Applied Statistics*, 43:569–582.

Högmander, H. and Møller, J. (1995). Estimating distribution maps from atlas data using methods of statistical image analysis. *Biometrics*, 51:393–404.

Huffer, F. W. and Wu, H. (1995). Markov chain monte carlo for autologistic regression models with application to the distribution of plant species. submitted for publication.

Possolo, A. (1986). Estimation of binary markov random fields. Technical Report 77, Department of Statistics, University of Washington.

Preisler, H. K. (1993). Modelling spatial patterns of trees attacked by bark-beetles. *Applied Statistics*, 42:501–514.

Propp, J. G. and Wilson, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, 9:223–252.

Raftery, A., Madigan, D., and Hoeting, J. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, 92:179–191.

Raftery, A. E. and Lewis, S. (1992). How many iterations in the Gibbs sampler? In *Bayesian Statistics 4*. Oxford University Press, New York.

Rao, J., Hartley, H., and Cochran, W. (1962). A simple procedure of unequal probability sampling without replacement. *Journal of the Royal Statistical Society, Series B*, 24:482–491.

Särndal, C., Swensson, B., and Wretman, J. (1991). *Model assisted survey sampling*. Springer-Verlag, New York.

Thompson, S. K. and Seber, G. A. F. (1996). *Adaptive Sampling*. John Wiley & Sons, New York, NY.

Wu, H. and Huffer, F. W. (1997). Modelling the distribution of plant species using the autologistic regression model. *Environmental and Ecological Statistics*, 4:49–64.

Yadava, S. S. and Singh, R. (1984). Optimum stratifications for allocation proportional to strata totals for simple random sampling scheme. *Communications in Statistics - Theory and Methods*, 13:2793–2806.