Spatial and Temporal Modelling of Hoki Distribution using Gaussian Markov Random Fields

by

Lindsay Robert Morris

A thesis
submitted to the Victoria University of Wellington
in fulfilment of the
requirements for the degree of
Master of Science
in Statistics.

Victoria University of Wellington
2017

Abstract

In order to carry out assessment of marine stock levels, an accurate estimate of the current year's population abundance must be formulated. Standardized catch per unit of effort (CPUE) values are, in theory, proportional to population abundance. However, this only holds if the species catchability is constant over time. In almost all cases it is not, due to the existence of spatial and temporal variation. In this thesis, we fit various models to test different combinations and structures of spatial and temporal autocorrelation within hoki (*Macruronus novaezelandiae*) CPUE. A Bayesian approach was taken, and the spatial and temporal components were modelled using Gaussian Markov random fields. The data was collected from summer research trawl surveys carried out by the National Institute of Water and Atmospheric Research (NIWA) and the Ministry for Primary Industries (MPI). It allowed us to model spatial distribution using both areal and point reference approaches. To fit the models, we used the software Stan (Gelman et al., 2015) which implements Hamiltonian Monte Carlo. Model comparison was carried out using the Watanabe-Akaike information criterion (WAIC, (Watanabe, 2010)). We found that trawl year was the most important factor to explain variation in research survey hoki CPUE. Furthermore, the areal approach provided better indices of abundance than the point reference approach.

ii

Acknowledgments

I would like to acknowledge the following people and organizations, for their continual professional, emotional and financial help with writing this thesis.

First of all, my sincerest thanks go to my primary supervisor, Dr Nokuthaba Sibanda. Nokuthaba has provided incredible support academically, and continues to help me develop as a statistician and researcher. Her advice and ideas have been invaluable to me.

Thank you to both the Ministry for Primary Industries (MPI) and the National Institute of Water and Atmospheric Research (NIWA) for allowing me to use their data, as well as funding my research.

My secondary supervisor Dr Matthew Dunn also deserves thanks for helping me to understand the data and fisheries related queries I had throughout the thesis.

I thank Maddie, Geoff, and Charlotte for their inputs, discussions, support, and friendship throughout the year. I also thank Maddie for her emotional support, as well as proof reading this thesis.

A special thanks go to Sam and Katie, who have been with me from the very beginning. They have kept me sane and grounded with their emo-

iv

tional support during the writing of this thesis.

Lastly, I thank my parents, Christina and Craig, who have always backed me from the beginning, and continue to encourage and love me.

Dedication

"Keep your shoulders back, and take deep breaths. You have no reason to worry"

Wendy Ann Avery (10/03/1943 – 19/08/2016)

I dedicate this thesis to my dear Nana. You have always inspired me to go as far as I can. It gets hard without your love and support, Nana, but you didn't raise a quitter! I wish you were here to see the final product, but I'm glad you are now at peace.

Contents

Chapter 1

Introduction

An important part of modelling the population dynamics of marine species involves using catch data obtained from surveys or commercial fisheries. Information about the catch, abundance, and other features of species is required in order to make estimates about stock productivity, current stock status, and allowable catch levels. Data available to scientists will usually include catch and effort statistics. Catch per unit effort (CPUE) is a common assessment of local abundance, and can be summarized by using simple statistics (such as mean or median CPUE), or by more complex models that provide an index of stock abundance. However, estimates of CPUE can be inaccurate and imprecise if spatial and temporal variation is not accounted for.

Indices of species abundance are most often calculated at present using delta-generalized linear mixed models (delta-GLMM, Cooper et al. (2004)) to model a species CPUE. This involves treating spatial and temporal components as either fixed or random effects. However, previous research (see Shelton et al. (2014)) has shown that models that incorporate the spatial and temporal autocorrelation as random fields can yield more precise and accurate indices of abundance. Such models were investigated by Thorson et al. (2015) using a frequentist approach, however the models were computationally intensive.

1.1   Aim of Research

The focus of this thesis is on improving abundance index estimation by modelling spatial and temporal autocorrelation within a species' CPUE using a Bayesian approach. We propose various models to test different combinations and structures of spatial and temporal autocorrelation within hoki (*Macruronus novaezelandiae*) CPUE. We aim to model the spatial and temporal components using Gaussian Markov random fields in two ways. Data from summer research trawl surveys carried out by the National Institute of Water and Atmospheric Research (NIWA) and the Ministry for Primary Industries (MPI) allowed us to model spatial distribution using both areal and point reference approaches. By accounting for autocorrelation in different ways, we aim to compare each model's ability to explain spatial and temporal variation. We make models of this kind easier to compute by implementing them using a Bayesian approach, thus increasing their attractiveness and usability.

We will now review relevant literature on stock assessment models and types of fisheries data. We review the concepts related to catch per unit effort including standardization and modelling. This is followed by a section on Gaussian Markov random field models and their applications in the wider literature. We then look at model assessment and information criteria as applied to Bayesian models. We conclude this chapter by presenting the layout for this thesis.

1.2   Review of Literature

1.2.1   Fisheries Science and Stock Assessment

To begin this thesis, it is essential to introduce stock assessment modelling in the fisheries sciences.

Stock assessment involves the use of various statistical and mathemati-

cal calculations to make quantitative predictions about the behaviour of fish populations in response to alternative management choices (Hilborn & Walters, 2013). A fish stock can be defined as all fish belonging to a given species that live in a particular geographic area at a particular time. That is, all individuals capable of interbreeding (OSB, 1998). The purposes of stock assessment models are: to monitor the abundance and productivity of exploited fish populations; and to provide fishery managers a quantitative evaluation of the potential consequences of alternative actions.

In the 19th century, it was thought that fishing only had minor impacts on fish stocks (Haddon, 2010). Huxley (1883) claimed that most fish populations were so numerous and fecund that they could not be affected by the limited activities of human fishing. The problem of overfishing gained recognition when Russell (1931) presented a simple expression used to describe stock biomass dynamics.

$$S_{i+1} = S_i + (A + G) - (C + M), \tag{1.1}$$

where S_i is the stock biomass in year i, A is the total weight of all individuals recruiting to the stock each year, G is the total growth in biomass of individuals already recruited to the stock, C is the total weight of all fish caught, and M is the total weight of all fish that die of natural causes during the year.

Many stock assessment models have been formulated from this expression. These include the Baranov catch equation (Baranov, 1918), Ricker model (Ricker, 1954), Beverton-Holt model (Beverton, 1957), and Schaefer model (Schaefer, 1957).

In this thesis, we focus our attention on catch per unit effort, which can be used as an input to a stock assessment model.

We now review dependent and independent fisheries data, with a focus on their benefits and disadvantages.

1.2.2   Commercial Fisheries and Research Surveys

In order to make predictions on stock biomass it is necessary to estimate the number or weight of fish caught in the current year. Abundance indices can be used for this purpose. Catch and effort data provide information on the amount of a species caught, and the effort expended in order to obtain the catch. Abundance indices can be calculated by standardizing catch and effort data to account for differences in fishing gear, vessels, temporal and spatial trends, and biological and ecological factors. The factors that need to be accounted for depend on the type of catch and effort data we are working with. There are two main types of catch and effort data available to fisheries scientists. The first type of data comes from scientific surveys of fish populations. Scientific surveys are meticulously planned and well designed, which lead to observations that have minimal bias. However, these surveys are expensive to implement. The second type of data comes from commercial and recreational fishing vessels. This type of data is cheap and in abundance, however, is heavily biased (Maunder & Punt, 2004). When constructing stock assessment models, we often work with commercial and recreational fishing data, which creates problems due to the biased nature of these observations. Commercial data are 'presence-only' data: they are collected in areas and at times when fish are available (fishing hotspots). There is also no information about the impact that fishing has on the population, which is problematic because data can be influenced by how a certain population reacts to fishing. In this thesis, we restrict ourselves to working with fisheries independent data from scientific surveys, in order to focus only on accounting for temporal and spatial variation.

We now review abundance indices and catch per unit effort.

1.2.3 Catch per Unit Effort

One of the most common pieces of information to assess the status of fish stocks is the relative abundance index. This is usually derived from catch per unit effort (CPUE) data (Maunder & Starr, 2003; Maunder & Punt, 2004; Maunder et al., 2006; West et al., 2014). The interpretation of CPUE rests on a fundamental relationship in fisheries science that relates amount of catch to population abundance and fishing effort. This relationship is given by:

$$C_t = hE_tN_t, \tag{1.2}$$

where C_t is a measure of catch during time period t, E_t is the fishing effort expended in time period t, N_t is the population abundance in time period t, and h is the catchability coefficient which is the portion of stock caught by one unit of effort (Maunder et al., 2006).

Equation 1.2 can be rearranged to give an expression that shows an assumed proportional relationship between catch per unit effort and population abundance:

$$\text{CPUE} = \frac{C_t}{E_t} = hN_t \tag{1.3}$$

In other words,

$$\text{CPUE}_t \propto N_t, \tag{1.4}$$

given that h is constant over time.

However, in almost all cases h is not constant over time. To account for this the raw CPUE needs to be standardized. Standardization attempts to control the effects of factors that may cause the catchability to change over time (Maunder et al., 2006). The factors that affect catchability depend on whether the data is fisheries dependent (from commercial or recreational fishing) or fisheries independent (from research surveys). Fisheries dependent CPUE data needs to be standardized to account for the fact that the catch data come from many different fishing vessels. These different

vessels each come with their own fishing equipment and associated efficiencies/deficiencies. In addition, CPUE should also be standardized to account for year and location effects that may contribute to the variability in catch. Because fisheries independent data come from carefully designed research surveys the only necessary standardization is that of time and spatial effects. This is because research survey data attempt to control bias associated with different fishing vessels and equipment by keeping these constant for each catch sample taken. In a paper written by Maunder & Punt (2004) it is stated that year should always be included in the model, whether it be significant or not. This is because the primary objective of standardizing catch and effort data is to detect trends in abundance over time. In chapter 4, we compare models with and without a temporal effect.

1.2.4   Catch per Unit Effort Modelling

In this section we begin by introducing a variety of models used to standardize CPUE. Each successive model becomes more general in order to incorporate other distributions. In addition to looking at models, we also look at some distributions used to model CPUE, and factors relating to it.

Throughout this thesis, we use 'standardized CPUE' to denote standardized catch per unit effort constructed from fisheries dependent data, and 'modelled/predicted CPUE' to denote standardized catch per unit effort from fisheries independent data. The data used in this thesis comes from fisheries independent research surveys (see Section 2.1) and hence we focus on constructing abundance indices by modelling CPUE.

A variety of methods have been used to standardize and/or model CPUE (Gavaris, 1980; Campbell, 2004; Bolker et al., 2009; Thorson et al., 2015). By far the most common method is to employ a generalised linear model (GLM). GLMs can be used to account for the variation in amount of catch using a linear combination of explanatory variables (Lynch et al., 2012). As

a result they are relatively straightforward and can be easily implemented (Maunder & Punt, 2004). GLMs are attractive in that there is flexibility for the choice in distribution of the abundance index, the covariates that relate to catch (the systematic component), and how the two components are linked (the link function). An example of a GLM would be the classical general linear model. The model equation, below, is taken from Nelder & Baker (1972):

$$y_i = \boldsymbol{x_i}\boldsymbol{\beta} + \varepsilon_i, \tag{1.5}$$

where the continuous response variable y_i is assumed to follow a Normal distribution and is linked to a systematic component of linear predictors $\boldsymbol{x_i} = (x_{i1}, x_{i2}, ..., x_{iK})$ and their coefficients $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_K)$. Here, $i = 1, ..., n$ with n being the number of observations, and K is the total number of predictors.

However, there is a model assumption that the covariates are indeed related to catch linearly. This may not always be the case, and so the use of generalised additive models (GAMs) have gained some popularity (Swartzman et al., 1992; Katsanevakis & Maravelias, 2009). The model is traditionally written in the following form (Hastie & Tibshirani, 1986)

$$g(\mathrm{E}(y_i)) = \sum_{j=1}^{K} f_j(x_{ij}), \tag{1.6}$$

where g is the link function, and $f_j(x_{ij})$ is a function on the jth predictor for the ith observation.

Another assumption of GLMs is the independence of errors. When this assumption is relaxed, more than one error component can be introduced. Such models are called hierarchical generalized linear models (McCullagh & Nelder, 1989). In hierarchical generalized linear models, random effects are introduced into the model in addition to the fixed effects. Furthermore, the random effects can be correlated and that allows us to incorporate structures such as spatial and temporal dependency between observations. Hierarchical generalised linear models have the following model

form

$$g(\mathrm{E}(y_i)) = \boldsymbol{x_i}\boldsymbol{\beta} + v_i, \tag{1.7}$$

where g is the link function, y_i is the ith observation, $\boldsymbol{x_i}$ is the vector of predictors for observation i, $\boldsymbol{\beta}$ is the parameter vector for the predictors, and v_i is the random effect. The assumptions are that the response y_i given the random effect has some distribution, f, and the random effect has some separate distribution f_u (McCullagh & Nelder, 1989).

When the random effects are assumed to be normally distributed, then the hierarchical GLM is called a generalized linear mixed model (GLMM).

The log-Normal distribution is commonly used for modelling CPUE. However, it is only defined for positive inputs, which has been identified as problematic when there is a large proportion of zero catches. A solution is to use a modified GLMM. An example of this the so-called delta-generalized linear mixed model (dGLMM), which assigns distributions to different components of a response variable. An illustration of this is modelling probability of encounter and catch rate separately, with different distributions. The dGLMM allows us to model situations with zero catch for non-zero effort. This makes use of a mixture distribution such that

$$f(y_i|\boldsymbol{\beta}) = (1 - \theta_i)\mathbb{I}_{(y_i=0)} + \theta_i f(y_i|y_i > 0, \boldsymbol{\beta}), \tag{1.8}$$

where θ_i is the probability of encountering the target species.

An example of such a model is given in Thorson et al. (2015). Thorson et al. apply a dGLMM to data for groundfish species. They model the probability of a positive encounter with the species separately from the positive catch rate. The probability of positive catch is is modelled using a logistic GLM, with the number of positive encounters modelled by a Binomial distribution:

$$\log\left(\frac{\theta_i}{1 - \theta_i}\right) = \sum_{j=1}^{K} \beta_j x_{ij} + \nu_i^{(\theta)} + \varepsilon_i^{(\theta)}. \tag{1.9}$$

The distribution of positive catch rate is modelled using a log-Normal distribution, with catch per area swept as the response variable, such that

$$\log y_i = \sum_{j=1}^{K} \beta_j x_{ij} + \nu_i^{(y)} + \varepsilon_i^{(y)}. \tag{1.10}$$

In Equations 1.9 and 1.10, θ_i is the probability of a positive encounter with the species in trawl i, y_i is the catch per area swept of trawl i, $x_{i1}, ..., x_{iK}$ are the K predictors, $\boldsymbol{\beta} = (\beta_1, ..., \beta_K)$ is the parameter vector for the predictors, $\nu_i^{(\theta)}$ is the spatial random effect for the encounter probability, $\nu_i^{(y)}$ is the spatial random effect for catch rate, $\varepsilon_i^{(\theta)}$ is the spatio-temporal interaction for the encounter probability, and $\varepsilon_i^{(y)}$ is the spatio-temporal interaction for the catch rate. To model the random spatial and spatio-temporal components in the model, Thorson et al. (2015) uses multivariate Normal distributions that incorporate Markovian properties. These distributions are called Gaussian Markov random fields, and we review these in the following section.

Considering all of the methods used to standardize and/or model CPUE described Section 1.2.4, it was decided that a hierarchical model would be used to incorporate spatial and temporal dependence between observations. In this thesis, a Bayesian approach is used to estimate the model parameters. This is further discussed in Chapter 3.

1.2.5 Gaussian Markov Random Fields

In this section, we investigate the use of Gaussian Markov random fields (GMRFs) as applied to hierarchical models. We review two applications of GMRFs and assess their suitability for modelling spatial and temporal autocorrelation in this thesis. The first application we review is in time series modelling, while the second is in spatial statistics. Both applications are relevant to this thesis.

A Gaussian Markov random field (GMRF) is a finite-dimensional random vector following a multivariate Normal distribution (Rue & Held,

2005). The term *Markov* comes from satisfying conditional independence assumptions (described further in Section 3.2.2). GMRFs play an important role in complex hierarchical modelling, being used to describe the spatial and temporal dynamics of natural and real systems (Rue & Held, 2005). They have been applied in a wide range of fields including structural time series (Dakos et al., 2008; Mikkonen et al., 2008), longitudinal and survival data (Knorr-Held, 1999), image analysis (Besag et al., 1991), and spatial statistics (Besag et al., 1991; White & Ghosh, 2009; Paciorek et al., 2013; Thorson et al., 2015).

Autoregressive models are GMRFs applied in one dimension. They have been a staple in modelling time series data since the 19th century (Rue & Held, 2005). One of many examples of an autoregressive model can be found in Dakos et al. (2008). A first-order autoregressive model was used to compute the autocorrelation within a time series of the Earth's temperatures during abrupt climate change events. Dakos et al. (2008) were able to test for evidence of decreasing fluctuations in temperature before a major climatic event.

Another example of autoregressive models can be found in Mikkonen et al. (2008). Here, Mikkonen et al. (2008) use autoregressive models to account for temporal autocorrelation within the commercial CPUE of baltic herring. They use the following hierarchical model:

$$
\begin{aligned}
y_{ijkv} =& (\beta_0 + u_i) + \beta_1 x_{1v} + (\beta_2 + v_{1i})x_{2iv} + \beta_3 x_{3i} + \beta_4 x_{4ijkv} \\
& + \beta_5 x_{5ijkv} + (\alpha_k + v_{ik}) + \delta_v + \gamma_{ijkv} + v_{2i}x_{6ijkv} + \varepsilon_{ijkv},
\end{aligned}
\tag{1.11}
$$

where y_{ijkv} is the log CPUE of trawl j made by vessel i in month k of year v, $i = 1, ..., n$, where n is the number of vessels, $j = 1, ..., m_i$, where m_i is the number of trawls made by vessel i, $k = 1, ..., 12$, and $v = 1990, ..., 2002$. The x terms are various predictors, the β terms, δ_v, and γ_{ijkv} are fixed effects, and the u and v terms are random effects. They are all defined in Mikkonen et al. (2008). The parameter ε_{ijkv} is the residual term.

Mikkonen et al. (2008) selected three different covariance structures for the residual term. The covariance structures are of increasing complexity to account for the temporal autocorrelation within baltic herring CPUE. The first model uses a residual covariance that assumes independence between each year. The second model uses a residual covariance that assumes a first order autoregressive structure to model the temporal autocorrelation. The third and final model uses an autoregressive moving average (ARMA) model to account for the temporal autocorrelation. Each model was compared using the Akaike information criterion (Akaike, 1973), and the Bayesian information criterion (Schwarz et al., 1978). The latter two models that accounted for temporal autocorrelation using autoregressive models, were found to perform the best.

We decided to use a GMRF to model temporal autocorrelation within the CPUE of hoki. This is looked at in greater detail in Chapters 2 and 3.

We now look at examples where GMRFs are generalised from one-dimensional autoregressive models to multi-dimensional spatial models.

As mentioned previously in Section 1.2.4, Thorson et al. (2015) constructed a dGLMM to model both positive catch rate (CPUE) and probability of encounter, for multiple fish species off the US West Coast. GMRFs were used to account for spatial and spatio-temporal autocorrelation present in CPUE at a point level. Matèrn distributions were used to model the covariance structures of the GMRFs, which allowed for imposing correlations based on distances between observations. Thorson et al. (2015) compared their GMRF model to a stratified dGLMM model on the same data. The stratified model included a random effect to account for spatial variation across stratified observations. They found that abundance indices were more precise when computed using the GMRF model.

Conditionally autoregressive (CAR) models (Besag et al., 1991) use another specification of the covariance matrix in order to account for spatial autocorrelation. A CAR model assumes that strata or points that neighbour (i.e. share a border, or are in close proximity) will be more correlated

than those that do not. Therefore, the covariance matrix is a function of either the number of neighbours for each stratum, or a function of the distances between each pair of points.

CAR models have been used extensively to model spatial variation and autocorrelation. One such example is found in White & Ghosh (2009), where they construct a variation of the CAR model for the purposes of predicting cesium concentrations at unobserved locations in Switzerland. Another example can be found in Lee (2011), and they use four variations of the CAR model to map the spatial pattern of cancer risk in Greater Glasgow, Scotland. As a final example, Paciorek et al. (2013) presented two CAR models that attempt to account for spatial autocorrelation in both areal and point references observations.

We decided to implement CAR models to account for spatial variation and autocorrelation in hoki CPUE. This is discussed in much more detail in Section 3.2.3.

1.2.6 Model Assessment

An important part of constructing any model involves assessing its ability to make accurate predictions. In this thesis, we focus on constructing models using the Bayesian approach. We first introduce the concept of predictive accuracy. Then we review a range of information criteria that have been used to assess Bayesian models in the literature.

Predictive Accuracy

Once a model has been fit, it is necessary to measure the model's predictive accuracy. Predictive accuracy allows for assessment of a model's goodness of fit, and can be used in the process of model comparison and selection (Vehtari & Gelman, 2014). We can measure a model's predictive accuracy in different ways, which are each tailored toward the model application. Predictive accuracy can be measured for both point predictions and prob-

abilistic predictions. For point predictions, measures of predictive accuracy are called scoring functions. Gneiting (2011) provides a discussion of common scoring functions including the mean squared error, absolute error, and relative error. The most popular scoring function is the mean squared error. This has the form,

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \text{E}(y_i|\theta))^2, \tag{1.12}$$

where y_i is the observed response at index i and $\text{E}(y_i|\theta)$ is the posterior mean.

The MSE is advantageous due to its simplicity in calculation and interpretation, however, it is not appropriate when there is a departure from the assumption of normality of the observations.

A more general scoring function is the logarithmic score. The log predictive density can be used to summarize a model's predictive fit (Gelman et al., 2014a). This is because it has a close connection to the Kullback-Leibler (KL) information measure. In addition, the log predictive density is proportional to the mean squared error if the model is Normal with constant variance. Because log predictive density is derived from the KL information, a model with the lowest KL information has the highest expected log predictive density, and hence highest posterior probability. Therefore it can be used as a measure of overall model fit.

An ideal measure of model fit would be its predictive performance for new data points. This quantity is defined below:

$$\log p(\tilde{y}_i|\boldsymbol{y}) = \log \int p(\tilde{y}_i|\theta) p(\theta|\boldsymbol{y}) d\theta, \tag{1.13}$$

where $\boldsymbol{y} = (y_1, ..., y_n)^T$ is the response vector, $p(\tilde{y}_i|\boldsymbol{y})$ is the posterior predictive distribution and $p(\theta|\boldsymbol{y})$ is the posterior distribution. However, the future data points are themselves unknown, and so we define the **expected**

out-of-sample log predictive density for a new data point,

$$\text{elpd} = \text{E}_f(\log p(\tilde{y}_i|\boldsymbol{y})) = \int \log p(\tilde{y}_i|\boldsymbol{y}) f(\tilde{y}_i) d\tilde{\boldsymbol{y}}, \qquad (1.14)$$

where $f(\tilde{\boldsymbol{y}})$ is a realization of a new dataset from the true data generating distribution, f.

In order to keep comparability with the observed dataset, a measure is defined that takes each data point individually. This is the expected out-of-sample log pointwise predictive density,

$$\text{elppd} = \sum_{i=1}^{n} \text{E}_f(\log p(\tilde{y}_i|\boldsymbol{y})). \qquad (1.15)$$

However, we do not know what the true data generating distribution, f is. There is no general approximation for this. A naive attempt uses the log predictive density for the existing data to approximate the elppd:

$$\text{lppd} = \sum_{i=1}^{n} \log p(y_i|\boldsymbol{y}) = \sum_{i=1}^{n} \log \int p(y_i|\theta) p(\theta|\boldsymbol{y}) d\theta. \qquad (1.16)$$

This quantity can be computed from posterior simulations as shown below,

$$\widehat{\text{lppd}} = \sum_{i=1}^{n} \log \left(\frac{1}{S} \sum_{s=1}^{S} p(y_i|\boldsymbol{\theta}^s) \right), \qquad (1.17)$$

where S is the number of posterior simulations. This is an overestimate of the elppd for future data. To obtain better estimates of the elppd, a range of bias corrections have been used in the literature. The following subsections look at two possible bias corrections, related to the number of parameters in a model. We start with the Akaike Information criterion (Akaike, 1973).

Akaike Information Criterion

In a statistical analysis, inferences on the parameters, θ, can be summarized by either point estimates (such as maximum likelihood estimates) or by posterior distributions. In the case where a maximum likelihood approach is taken, the out-of-sample predictive accuracy is defined by the expected log predictive density given the point estimate $\hat{\theta}$ (Gelman et al., 2014b),

$$\text{elppd}_{\hat{\theta}} = \text{E}_f\left(\sum_{i=1}^{n} \log p(\tilde{y}_i | \hat{\theta}) \right). \tag{1.18}$$

To compute this quantity, it is necessary to use the log posterior density of the observed data, $\boldsymbol{y}$ given a point estimate $\hat{\theta}$, then correct for the bias due to overfitting (Gelman et al., 2014b).

The simplest bias correction is based upon the asymptotic normal posterior distribution. In this limit, subtracting k (the number of parameters estimated in the model) from the log predictive density given the maximum likelihood estimate, is a correction for how much the fitting of k parameters will increase the predictive accuracy, by chance alone (Gelman et al., 2014b). We arrive at the famous formula of Akaike (1973), the so-called Akaike information criterion (AIC):

$$\text{AIC} = -2\widehat{\text{elppd}}_{\text{AIC}} = -2 \sum_{i=1}^{n} \log p(y_i | \hat{\theta}_{\text{MLE}}) + 2k, \tag{1.19}$$

where k is the number of parameters estimated in the model.

In this thesis, we intend to fit Bayesian models that contain hierarchical structure. Because of this, we cannot simply add k to account for model complexity. This is because hierarchical structures and informative priors reduce the amount of overfitting, compared to the maximum likelihood approach. As such, the measure of model complexity (the bias correction term) needs to be altered.

Deviance Information Criterion

An alternative to AIC to account for these hierarchical structures is the deviance information criterion (DIC) conceived by Spiegelhalter et al. (2002). This information criterion substitutes the maximum likelihood estimate for the parameters θ with the posterior mean, $\hat{\theta}_{\text{Bayes}}$. As a result, DIC has Bayesian roots, although it is not strictly Bayesian. In addition to the replacement of the estimate of the parameter, the effective number of parameters is modified to incorporate the data:

$$\text{DIC} = -2\widehat{\text{elppd}}_{\text{DIC}} = -2\sum_{i=1}^{n} \log p(y_i|\hat{\theta}_{\text{Bayes}}) + 2p_{\text{DIC}}. \tag{1.20}$$

Here, p_{DIC} is the effective number of parameters, with two forms given by Gelman et al. (2014b).
The first form is defined below,

$$p_{\text{DIC}} = 2\left(\sum_{i=1}^{n} \log p(y_i|\hat{\theta}_{\text{Bayes}}) - \text{E}_{\text{post}}\left(\sum_{i=1}^{n} \log p(y_i|\theta) \right) \right). \tag{1.21}$$

This can be computed using the posterior simulations of θ, θ^s, where $s = 1, ..., S$:

$$\widehat{p}_{\text{DIC}} = 2\left(\sum_{i=1}^{n} \log p(y_i|\hat{\theta}_{\text{Bayes}}) - \frac{1}{S}\sum_{s=1}^{S}\sum_{i=1}^{n} \log p(y_i|\theta^s) \right). \tag{1.22}$$

The alternative formula is variance based, which has the advantage of only producing positive values:

$$p_{\text{DIC,alt}} = 2\text{Var}_{\text{post}}\left(\sum_{i=1}^{n} \log p(y_i|\theta) \right). \tag{1.23}$$

This is unlike the first formula, which can produce negative values when the posterior mean is far from the mode.
Both AIC and DIC have their disadvantages in the context of model selection. AIC does not work in settings where strong prior information exists

(Gelman et al., 2014a). In this thesis we will be using GMRFs to incorporate our assumptions of spatial and temporal dependence within CPUE. As a result, AIC will be inappropriate as a model selection criterion. DIC is a more likely candidate, however, it fails when a posterior distribution is not well summarized by its mean (Gelman et al., 2014a). In this thesis, we will look at using a fairly new information criteria for the purpose of model selection.

The Watanabe-Akaike information criterion (also known as the widely applicable information criterion, WAIC) is more Bayesian than DIC as described above. It was conceived by Watanabe (2010), and uses the computed log pointwise posterior predictive distribution 1.17, as opposed to the log posterior density evaluated at a point estimate. Its effectiveness in relation to Bayesian hierarchical models has been investigated by Gelman et al. (2014a), and we introduce the criterion in Section 3.5.

We have reviewed stock assessment and the types of data used in relation to fisheries modelling. We decided to construct hoki abundance indices by using Bayesian hierarchical models. The models will use GMRFs to account for spatial and temporal variation and autocorrelation within hoki CPUE. We will then use WAIC to compare the models.

1.3 Thesis Outline

In Chapter 2 we introduce the summer research trawl surveys carried out by the National Institute of Water and Atmospheric Research (NIWA) for the Ministry for Primary Industries (MPI). We discuss the design of the surveys, as well as the variables reported in the dataset. This is followed by an exploration of the data, which includes giving evidence of temporal and spatial autocorrelation within hoki catch per unit of effort. Chapter 3 follows, and presents Bayesian hierarchical models, Gaussian Markov random fields and Hamiltonian Monte Carlo posterior sampling. This is followed by a section on convergence diagnostic methods and model as-

sessment techniques. Chapter 4 introduces the five models we fit to the
hoki CPUE data. A discussion on convergence for the models is given.
The results from each model are given in Chapter 5, and are discussed in
Chapter 6. Lastly, Chapter 6 concludes this thesis by presenting further
considerations for future research.

All statistical programming in this thesis was written in **R**. Code is available on request.

Chapter 2

Data

This chapter outlines the dataset used in this thesis. We begin with an account of the research vessel surveys conducted to obtain the data. A description of the dataset, and each variable of interest then follows. We present a section dedicated to exploration of the data, with a focus on detecting autocorrelation due to time and space.

2.1 Research Vessel Surveys

The data used in this thesis come from summer research trawl surveys of the sub-Antarctic region carried out by the National Institute of Water and Atmospheric Research (NIWA) for the Ministry for Primary Industries, New Zealand (MPI). The data were provided by NIWA and MPI. The dataset is a time series that has been accumulated from 1991 to 1993, and then again from 2000 to 2008 (Bagley et al., 2013). It contains data on catch weight for multiple species in the sub-Antarctic. The purpose of the surveys was primarily to estimate abundance of a particular fish species, *Macruronus novaezelandiae*, commonly known as hoki. The surveyed area has changed over time to incorporate estimation of abundance for other species (such as hake, *Merluccius australis*, and ling, *Genypterus blacodes*), but the primary focus remains on hoki.

The current set of strata is shown in Figure 2.1. All of the surveys included a core set of strata that covered depths of 300–800 m on the Southern Plateau, as well as a stratum at Puysegur that covered depths of 800–1000 m. The stratum at Puysegur was included to estimate abundance of hake. From 2000 onward, strata 3A, 3B, 5A and 5B were defined to improve the biomass estimates of hake and ling. The summer trawl surveys were suspended in 1993. This was to focus on the autumn trawl surveys that ran in March – June 1992, 1993, 1996, and 1998. However, interpretation of trends was complicated by the possibility of hoki leaving the survey area to spawn. As a result, the summer trawl surveys were re-instated in November 2000. When the summer trawl surveys were re-established in 2000, more strata were included to cover deeper waters to the east of the Southland Coast, and north and south of the Southern Plateau.

It was noted that stratum 26 (depth of 800–1000 m) south of Campbell Island was not surveyed in 2003, 2004, or 2006 due to varying reasons (weather, emergency etc.). In addition, stratum 17 at the Bounty Platform was only included in 1992 and 1993.

The core survey areas were based on depth intervals as alluded to above. Stratification was based on depth intervals of 300 – 600 m, 600 – 800 m, and later 800 – 1000 m. Furthermore, the areas were then subdivided by latitude and longitude.

2.2   Survey Design

A 2–phase stratified random design proposed by Francis (1984) was used. The motivation was to decrease the overall observed variation in a species catch weight.

The first phase of the survey uses a stratified random design to allocate stations (trawl locations) within the survey area. The number of stations allocated to a particular stratum is defined as being proportional to the area of the stratum multiplied by the standard deviation of the catch weights (up
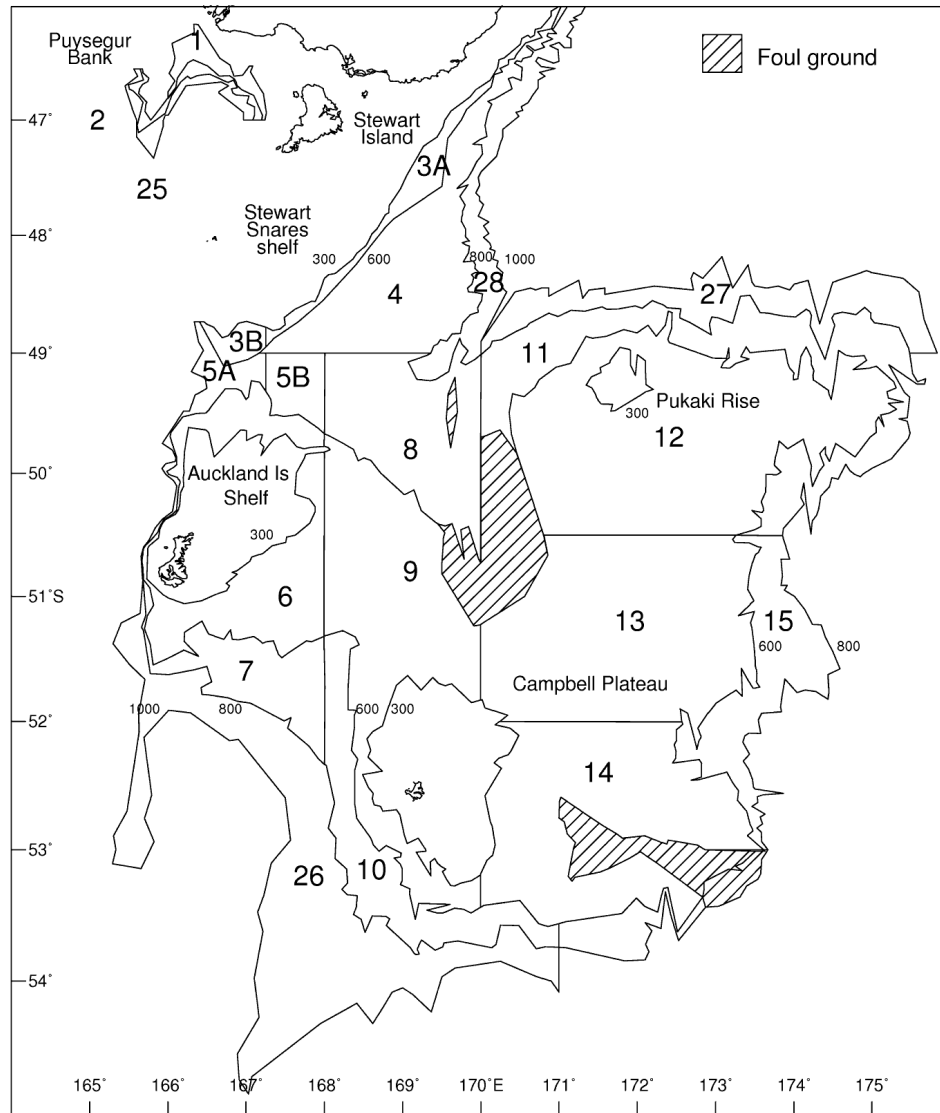
Figure 2.1: Map of the sub-Antarctic region illustrating the surveyed area and 21 strata (note that stratum 3 and 5 were split into 3A, 3B, 5A, and 5B in 2000) (Bagley et al., 2013). Stratum 17 is further east, not shown here.

to the current year) within the stratum, divided by a measure of sampling cost, that is,

$$n_j \propto \frac{A_j \text{SD}(C_j)}{\sqrt{c_j}}, \tag{2.1}$$

where n_j is the number of stations in stratum j, A_j is the area of stratum j, $\text{SD}(C_j)$ is the sample standard deviation of the catch weights, C_j, in stratum j, and c_j is the cost (effort, measured in hours aboard ship) involved with sampling stratum j.

When estimating biomass, the trawl station locations are assumed to be allocated to the stratum area randomly. However, there may be some instances where a station is too far from the ship's current location, in which case a closer point is chosen. Francis (1984) states that the surveys are carried out as if the stations were allocated randomly.

Once the first phase of sampling has been completed, the second phase begins by calculating the estimated relative gain (reduction in variance), G_j, from adding one station to stratum j. This is calculated for each stratum, with G, given by

$$G_j = \frac{A_j^2 \text{Var}(C_j)}{n_j(n_j + 1)}. \tag{2.2}$$

One station is allocated randomly within the stratum with the largest G_j value, and the relative gains are recalculated for that stratum. By adding more stations to the strata with the highest coefficients of variation, then we can reduce the catch weight variation in that strata, and hence reduce the overall catch weight variation. The second phase is therefore repeated until an overall target coefficient of variation is reached.

Sampling for the summer research trawl surveys was initially based on the catch weight of hoki, which is the species of interest in this thesis (Bagley et al., 2013). The target coefficient of variation for hoki was 15%.

2.3 Species

Hoki was chosen to be the species of focus for the analysis. Hoki is an offshore bathypelagic fish caught in bottom trawls at depths of 200-800 m throughout New Zealand waters (MPI, 2016). It plays an important role in the commercial fisheries industry in New Zealand. The decision to focus on hoki was motivated by two reasons. Firstly, the survey was designed to optimize estimation of hoki abundance to a targeted coefficient of variation. In addition, the survey reports high proportions of hoki encounters. Of all the species encountered, hoki reported the least number of zero catches. The proportion of zero catches for hoki was found to be 4.84% (64 out of 1322). This allowed for the choice between modelling encounter rate together with catch weight, or modelling catch weight alone. For the purpose of this thesis, zero catches were assumed to occur at random, and hence ignored. Figure 2.2 shows bar plots illustrating the percentage of zero catch trawls by year and by stratum. In the plot by year, the spread of zero catch trawls appears roughly even, with 2007 reporting the largest number of zero catches. The plot by stratum shows a much more uneven spread, with most of the zero catch trawls occurring in strata 17, 26, 27, and 28. These strata are not core strata, in that stratum 17 was removed after 1993, and strata 26, 27, and 28 were used in 2000 onward. We have some reservation about the randomness of zero catch trawls, however, we will continue as if they occurred randomly.

2.4 Dataset

The data were provided in two separate sets. Each set detailed different aspects of the 1322 summer research trawls in the sub-Antarctic. The first dataset describes the catch weight by species caught at each station for each trawl. It contained 28098 observations of catch weight for all species caught. Absence of a catch weight for a given record was indication that a
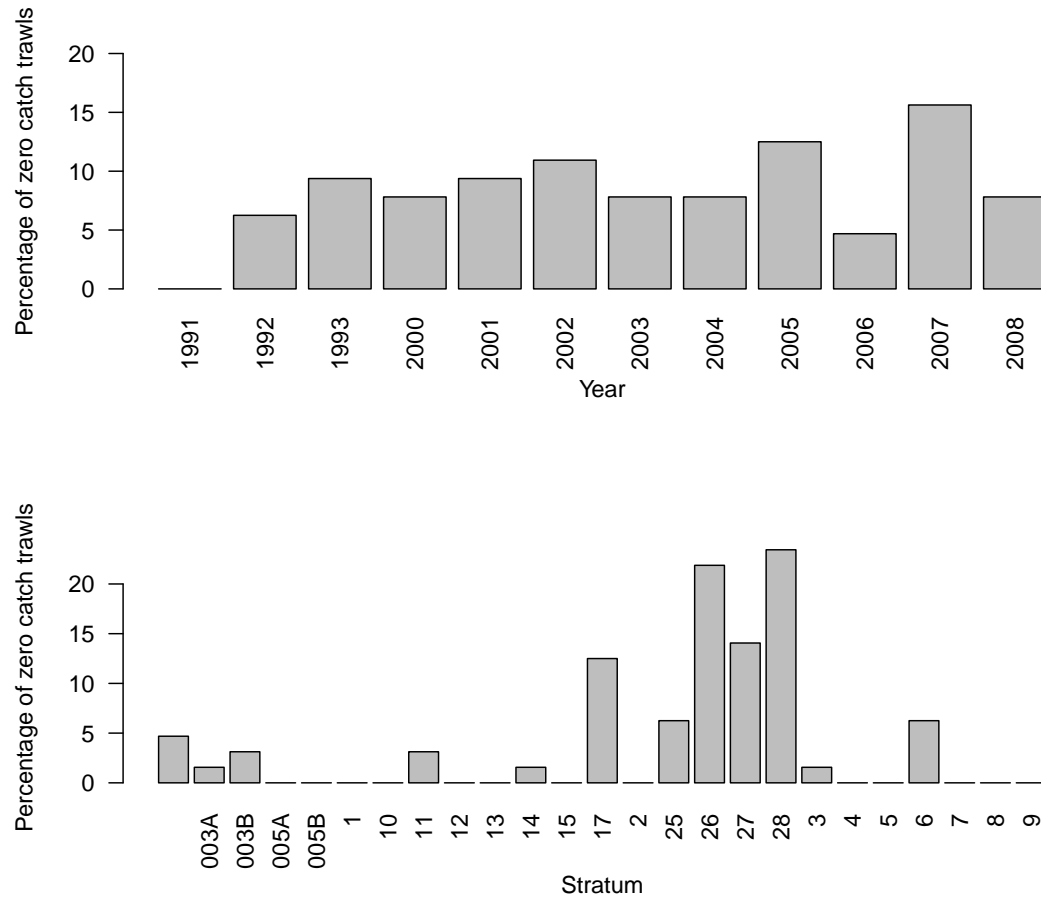
Figure 2.2: Bar plots showing the frequency of zero catch trawls by year, and by stratum.

species was not encountered during the trawl. In addition, each record had a corresponding trip code and station number, which allowed for identification of the individual trawl.

The second dataset gave information about the variables related to each of the 1322 trawls. In this dataset, trip code and station number were given, which allowed for matching of the station data to the corresponding catch weight observations. The set contained many variables, some of which permitted standardization for a range of possible effects. Tables 2.1 and 2.2 list relevant variables reported in the two datasets.

Table 2.1: List of relevant variables recorded in the catch dataset and their brief descriptions.

Variable Code	Description
species	Name of the species
weight	Catch weight of species caught in kg
index	Unique code for each individual trawl

The first step in preparing the data for analysis involved taking a subset of the catch observations corresponding to the species of interest. In this case, 1322 catch observations for hoki, one for each trawl, were extracted. The extracted records were then merged with the appropriate trawl information from the station dataset. There were 64 records with no catch weight recorded. We assumed that these records occurred randomly and hence removed them. This left us with a dataset of 1258 observations on hoki catch weight, with the associated date and time of trawl, depth, latitude, longitude, trawl distance, and stratum (Bagley et al., 2013). Furthermore, there were two observations that were not assigned to a stratum; these records were also removed. Finally, the single remaining positive catch record in stratum 17 was also removed because there was not enough information to estimate abundance in this stratum. Hence, of the 1322 original records, only 1255 were used. Figure 2.3 illustrates the process of data grooming

Table 2.2: List of relevant variables recorded in the station dataset and their brief descriptions.

Variable Code	Description
stratum	Label used to identify the area of a trawl
date_s	Trawl start date
time_s	Trawl start time in 24 hour time
dlat_s	Decimalised latitude at start of trawl
dlong_s	Decimalised longitude at start of trawl
date_f	Trawl end date
time_f	Trawl end time in 24 hour time
dlat_f	Decimalised latitude at end of trawl
dlong_f	Decimalised longitude at end of trawl
distance	Distance of trawl in km
min_gdepth	The minimum recorded depth
max_gdepth	The maximum recorded depth
index	Unique code for each individual trawl

described above.

2.5   Exploratory Data Analysis

In order to estimate fish relative abundance, it is common to obtain a standardized measure of catch per unit of effort, using covariates such as time of year, location, vessel identifiers, ecological factors and biological factors. This was discussed in detail in Section 1.2.4. Standardization accounts for the factors that influence a species' abundance over time. In the following subsections we define the dependent variable, catch per unit effort, as well as detailing the variables that potentially have an effect on it.
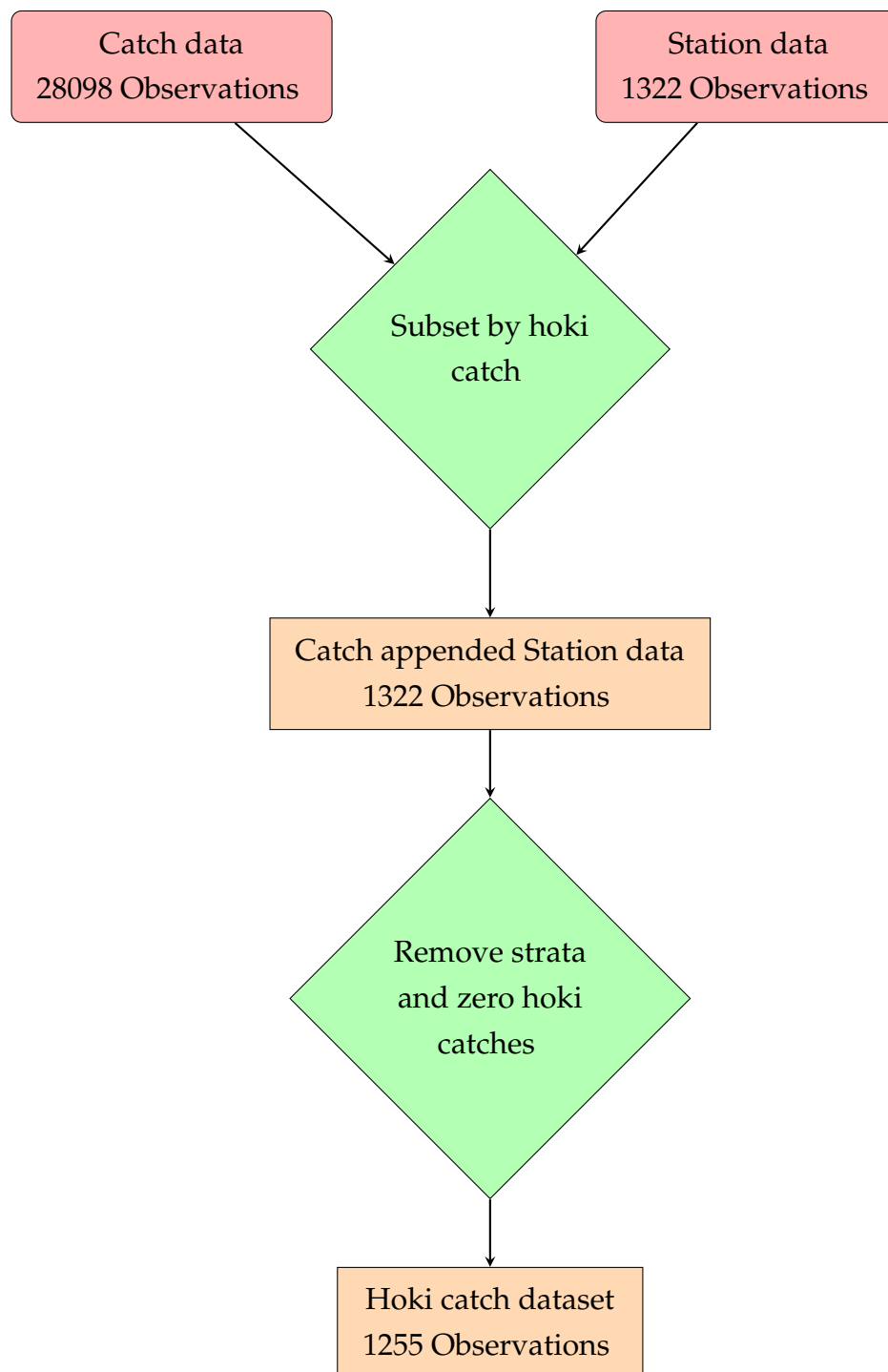
Figure 2.3: Flow diagram displaying the process of conditioning the data to an appropriate set to be used for analysis.

2.5.1   Catch per Unit Effort

When estimating relative abundance of a stock, it is necessary to use some measure of biomass. Amount of catch is assumed to be proportional to the population abundance, and hence can be used as an abundance index (see Section 1.2.3). Fisheries scientists either use the number of fish caught or the weight of the fish caught to measure amount of catch. Our dataset contains hoki catch weight, C, in kilograms.

Each trawl had a target length of 3 nautical miles. However, due to a number of reasons, most often encountering foul ground or unusually large catches (Bagley et al., 2013), trawls were sometimes cut short and accepted for biomass estimation if the length was at least 2 nautical miles. As a result, some trawls had a slightly different length, and in order to adjust for varying amounts of effort, each catch weight was divided by the respective trawl area swept. Area swept was computed by taking the distance of the trawl (l_{trawl}) and multiplying it by 100 m (the nominal width of the trawl, w_{trawl} , kept constant for each trawl). The new quantity was defined as catch per unit effort (CPUE), measured in kg/km^2, given by

$$\text{CPUE} = \frac{C}{l_{\text{trawl}} \times w_{\text{trawl}}}.\tag{2.3}$$

An initial look into the catch weight data showed the distribution of hoki CPUE to be approximately log-normal (see Figure 2.4). When CPUE was converted into log CPUE, a distribution resembling the Normal distribution was observed (see figure 2.4). As a result, CPUE was assumed to have a log-normal distribution for model fitting.

2.5.2   Factors affecting CPUE

Trawl Year

For each observation of catch, the date and time that the trawl occurred was recorded. Unfortunately, since each trawl occurred in the summer
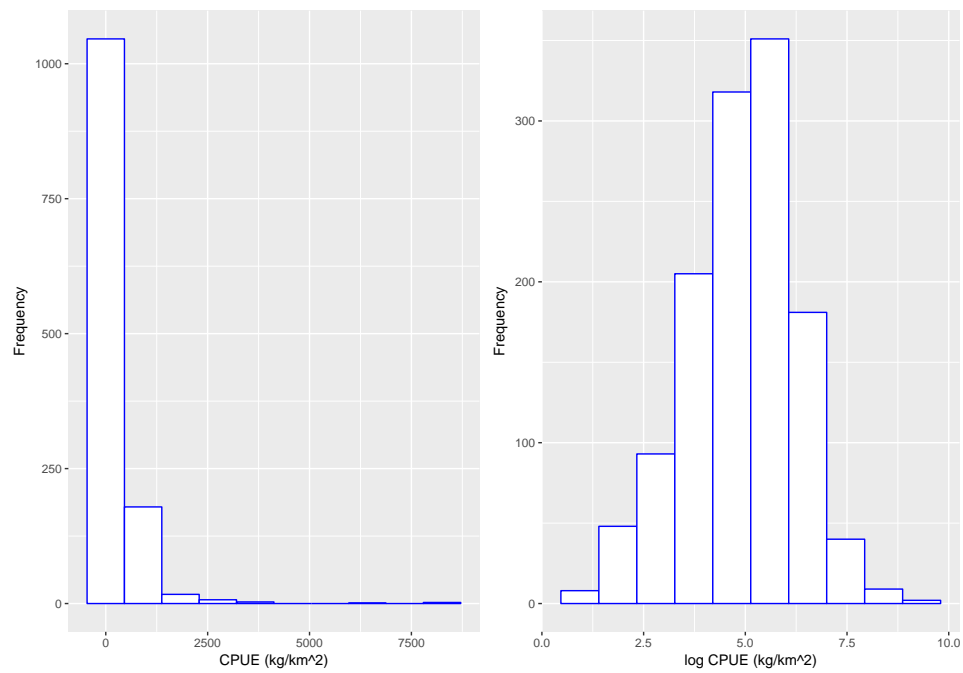
Figure 2.4: Two histograms displaying the distribution of CPUE (left) and log CPUE (right) in kg/km^2.

(November or December of each year), there was not enough information to investigate a seasonal effect. As a result, temporal effect was only explored on the year scale. It is now well known that the abundance of species changes over time (Preston, 1960), (Magurran, 2007). Hence, including a time component in the standardization of CPUE is absolutely necessary.

Table 2.3: Summary of hoki CPUE (kg/km^2) by trawl year (including zero catches).

Year	Mean	Median	Number of Trawls	Standard Deviation	Min.	Max.	Zero Catch (%)
1991	439.56	311.38	154	634.66	4.14	6337.29	0.00
1992	388.53	302.14	160	343.23	0.00	2155.49	2.50
1993	473.61	264.40	138	656.19	0.00	4094.13	4.35
2000	215.94	124.19	107	291.25	0.00	2465.20	4.67
2001	186.98	91.01	106	276.05	0.00	1948.52	5.66
2002	187.04	87.65	105	317.57	0.00	1825.45	6.67
2003	112.36	51.02	82	263.62	0.00	1907.85	6.10
2004	288.04	56.79	90	999.91	0.00	8253.96	5.56
2005	133.10	54.26	96	195.98	0.00	960.14	8.33
2006	115.58	65.66	91	158.84	0.00	800.79	3.30
2007	289.39	120.37	98	845.58	0.00	8019.35	10.20
2008	245.60	147.72	95	336.31	0.00	2043.56	5.26

Table 2.3 gives a summary of the CPUE of hoki for each year of the research trawls. Within each trawl year, the distribution of hoki CPUE is skewed to larger values. Evidence for this can be seen in Figure 2.5. Because of this asymmetry, the median provides a more robust measure of centrality.
The first three years of the survey, 1991, 1992, and 1993, have the largest number of trawls at 154, 160, and 138 respectively. In addition, these three years also have the largest median CPUE of hoki at 311 kg/km^2,
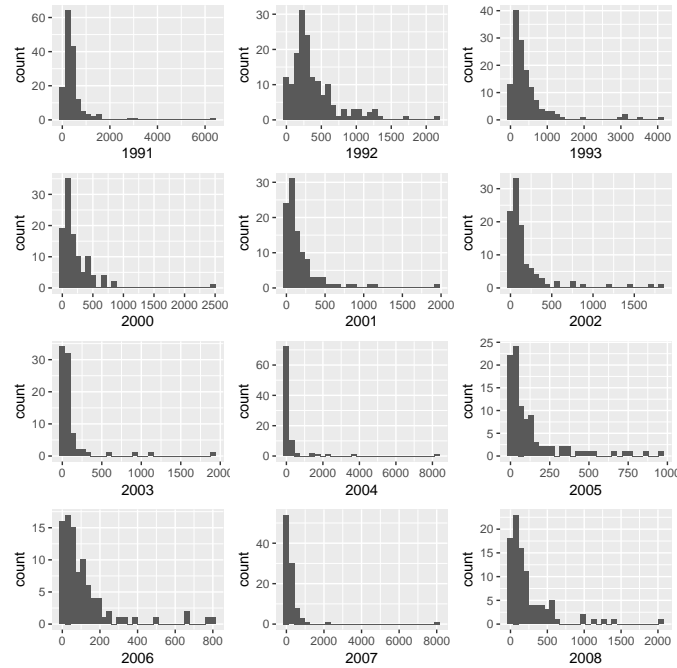
Figure 2.5: Histograms of hoki CPUE organized by trawl year.

302 kg/km^2, and 264 kg/km^2 respectively.

When the surveys began again in 2000, there were significantly fewer trawls per year. The year that had the least number of trawls was 2003, with 82 occurring that year. All of the median CPUE values were smaller for the later years (2000 − 2008) compared to the earlier ones (1991 − 1993). Of the later trawls, the year that had the largest median CPUE was recorded in 2008, with a value of 148 kg/km^2. The year that had the smallest median CPUE was 2003 with a value of 51 kg/km^2. In the years 1991, 1993, 2004, and 2007, unusually large CPUE observations were recorded with values of 6337, 4094, 8254, and 8019 kg/km^2 respectively. These large CPUE observations are potentially outliers.

Location

The dataset provided trawl location in two forms. Stratum number is included, which assigns each of the catch weights to a particular survey stratum – each of which has an estimated spatial area (km^2) as indicated in the New Zealand Fisheries Assessment Report for 2013/2014 (Bagley et al., 2013). In addition to stratum number, the latitude and longitude in degrees is included for the beginning and end of each trawl. These two types of location description allow for different procedures in modelling spatial structure. Stratification allows for an areal approach, whereas the exact locations allow for a point reference analysis. It was decided that both approaches would be conducted, and compared in this thesis, with the presumption that point referenced modelling would produce more accurate and precise estimates of abundance.

In order to obtain a single location for each trawl, the midpoint of the trawl was calculated. This assumes a straight line tow for each trawl. The CPUE could then be plotted on the map of the sub-Antarctic region (Figure 2.6). It can be seen from Figure 2.6 that the largest reported CPUE values occur near Puysegur Bank. Figure 2.7 shows that three of the largest CPUE observations were reported in stratum 1. This is reflected in Table 2.4 with the largest median CPUE of 458 kg/km^2 occurring in stratum 1.

Table 2.4 summarises the CPUE values in the hoki dataset by stratum. Just as we saw within year, within each stratum, the distribution of hoki CPUE is skewed to larger values. Evidence for this can be seen in Figure 2.7. Because of this asymmetry, the median provides a more robust measure of centrality.

The number of trawls changes more so from stratum to stratum than it does from year to year. Excluding trawls unassigned to strata, the number of trawls fluctuates from 9 to 121. After excluding zero catches, the stratum with the fewest trawls was stratum 26, with 4 trawls. Abundance index estimation may not be accurate for all strata, because of the low number of trawls in some strata.
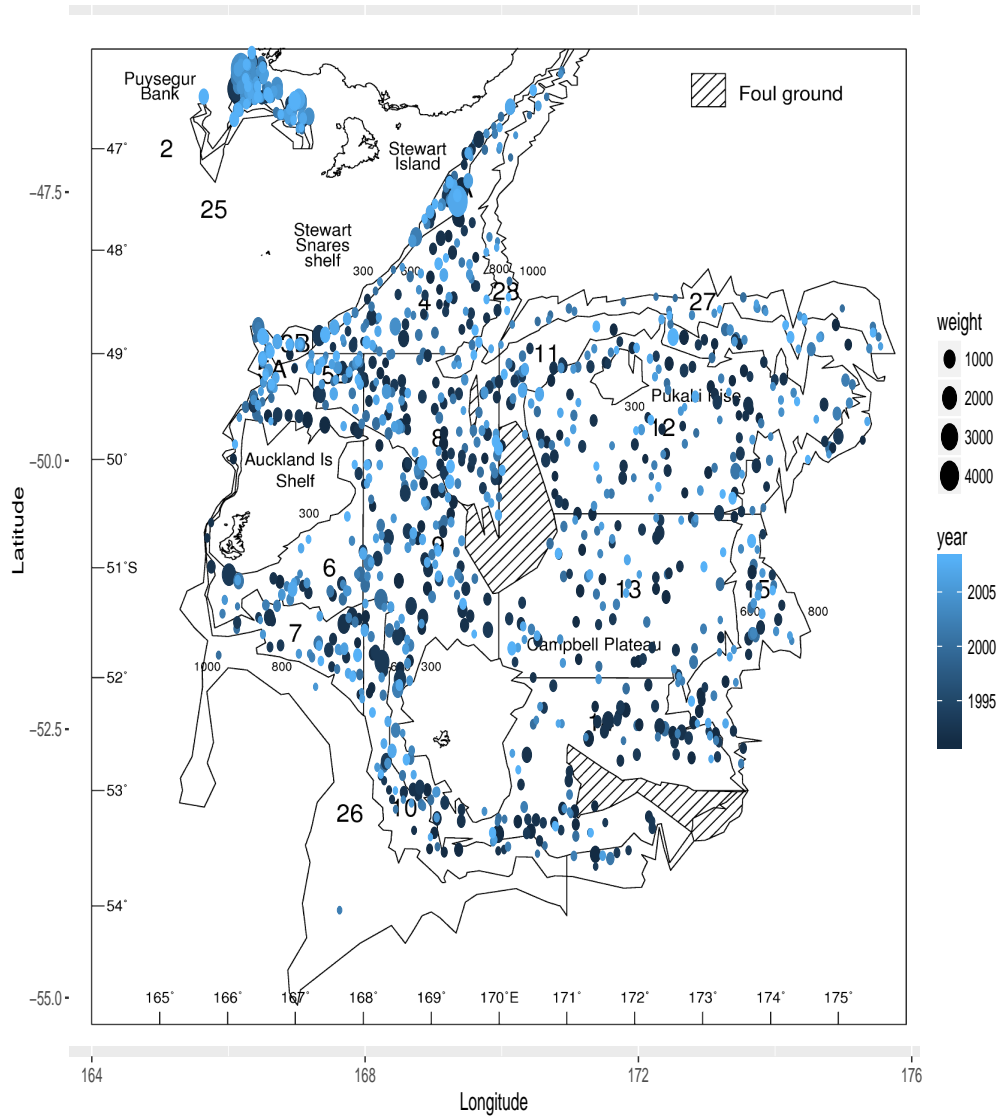
Figure 2.6: Map of 1255 trawls where hoki was caught in the sub-Antarctic region. The size of a point is representative of the CPUE in kg/km^2, whereas, the colour gradient is representative of the year.

Table 2.4: Summary of hoki CPUE by strata (including zero catches).

| Stratum	Mean	Median	Number of Trawls	Standard Deviation	Min.	Max.	Zero Catches (%)	Area (km²)
	83.43	0.00	5	170.16	0.00	386.93	60.00	
1	940.39	457.88	59	1471.90	21.76	8253.96	0.00	2150
2	248.85	206.37	48	195.93	28.61	1003.78	0.00	1318
3	709.72	305.16	14	1033.63	0.00	3151.94	7.14	6104
003A	478.26	133.82	39	1297.75	0.00	8019.35	2.56	4548
003B	365.62	205.11	34	447.58	0.00	1948.52	5.88	1556
4	212.20	133.46	74	280.94	13.54	1571.27	0.00	21018
5	398.61	295.18	16	348.51	59.04	1116.51	0.00	6262
005A	149.87	85.47	40	174.51	5.76	795.41	0.00	2981
005B	116.71	94.06	34	91.33	13.41	407.75	0.00	3281
6	308.43	102.95	63	477.32	0.00	3146.15	6.35	16682
7	288.87	156.25	45	399.43	3.60	2155.49	0.00	8497
8	253.59	205.57	93	172.46	13.27	768.53	0.00	17294
9	366.47	213.20	116	419.73	9.33	2465.20	0.00	27398
10	297.04	139.98	58	536.06	4.32	3560.48	0.00	11288
11	250.51	188.26	62	264.43	0.00	1436.90	3.23	23008
12	192.04	124.19	121	196.36	6.30	1092.15	0.00	45259
13	214.14	175.25	80	197.69	5.36	988.30	0.00	36051
14	269.26	190.33	82	308.64	0.00	1728.22	1.22	27659
15	214.74	128.98	51	225.57	4.78	1275.02	0.00	15179
17	0.40	0.00	9	1.20	0.00	3.60	88.89	11360
25	234.46	65.33	87	499.34	0.00	3151.91	4.60	1928
26	2.26	0.00	18	5.38	0.00	20.70	77.78	31778
27	25.03	16.12	39	28.08	0.00	93.59	23.08	12986
28	25.57	4.14	35	46.04	0.00	182.91	42.86	8336
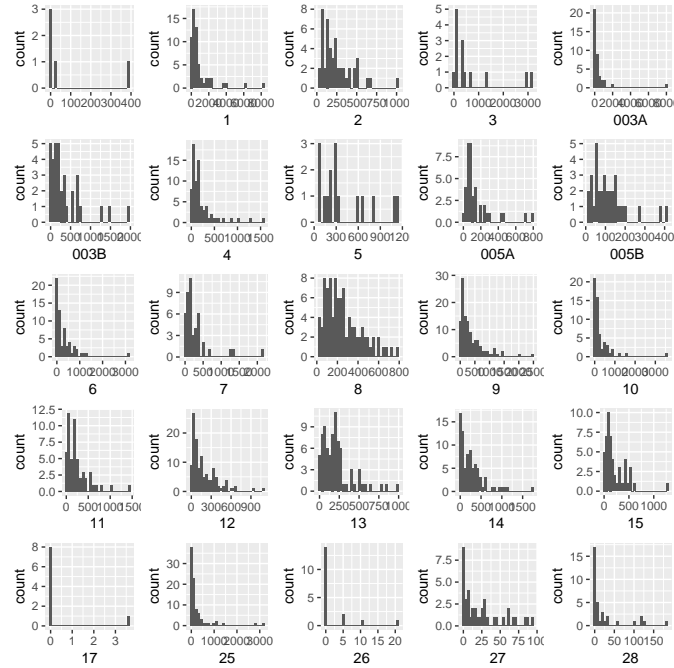
Figure 2.7: Histograms of hoki CPUE organized by Stratum.

Depth

Both minimum and maximum ground depth were included in the dataset. A single depth variable was created by taking the average of the minimum and maximum ground depth for each record. Figure 2.8 shows the depth intervals within each stratum. As expected, we can see that each stratum covers a particular interval of depths. This is of course due to the fact that strata were chosen according to depth intervals (as described in Section 2.1). We assume that depth need not be entered into the model because any variation due to depth should be accounted for by stratum.

Other factors

In addition to the factors described above, the dataset also included various other variables concerned with the fishing vessel and gear specifica-
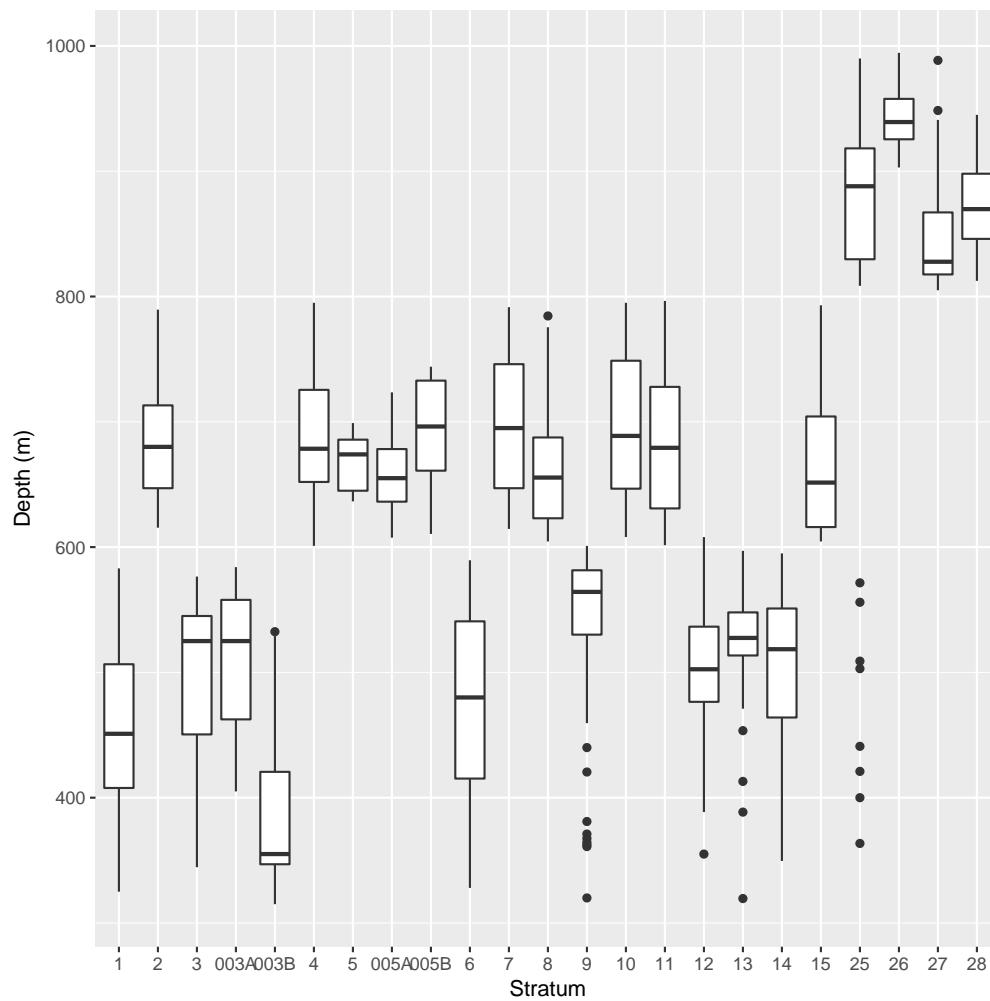
Figure 2.8: Box plot showing the trawl depth (in metres) within each stratum.

tion. The inclusion of these variables is unnecessary since only one vessel (*RV Tangaroa*) was used. Gear specification was kept constant across each trawl. Since vessel effects were controlled for, they were not explored.

It is possible that biological variables, for example body condition and sex ratios, could influence the variation in CPUE. However, such information was not collected, hence we can not include them when we model CPUE. We now look at time and location, in more detail. This is to assess whether the inclusion of temporal and spatial effects is necessary for modelling CPUE.

2.6 Autocorrelation

Autocorrelation is defined by Yule (1921) as the dependence of successive observations of a single variable. A common assumption to make when constructing models is independence between observations. When autocorrelation exists within data, care must be taken to take into account such dependencies. In the following subsections, we find evidence for the existence of spatial and temporal autocorrelation. In addition, we also show that some interaction between space and time exists within the catch weight observations for hoki.

2.6.1 Temporal Autocorrelation

The first type of autocorrelation explored in this thesis was temporal. In nature, many species distributions change over time (Preston, 1960). This change over time can happen on different time scales such as day, month, season, or year. In this dataset, the smallest time scale possible for exploration was by year. As each trawl occurred during the same time each year, we were unable to explore the possibility of seasonal or monthly effects. Before we constructed the models, investigation of possible temporal effects were explored. A box plot was constructed which displayed
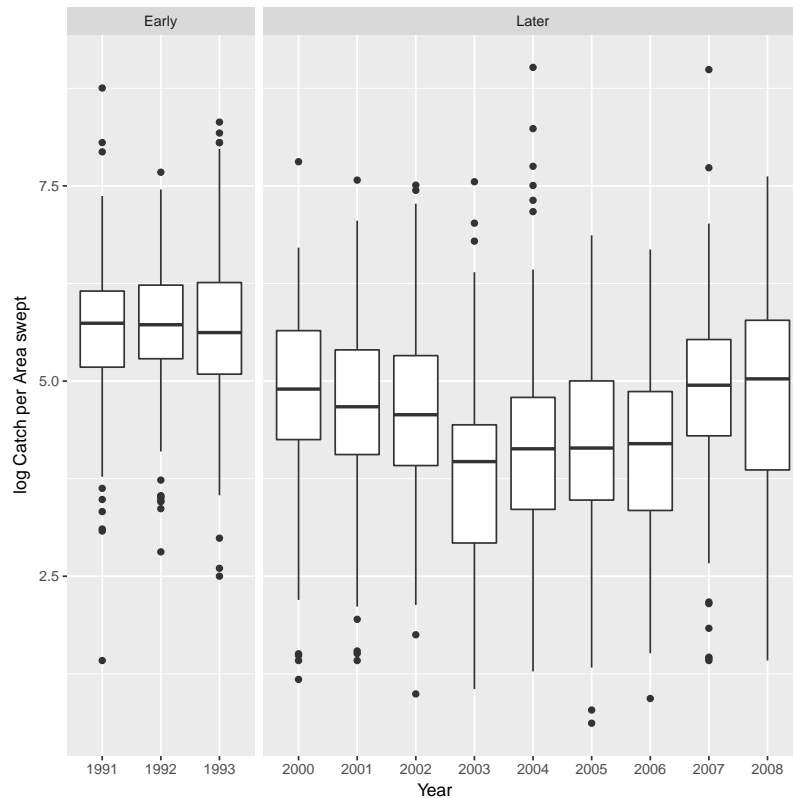
Figure 2.9: Box plots of log CPUE in kg/km^2 by year

the CPUE (in log scale) by year. Figure 2.9 shows a temporal trend, where CPUE is higher on average in the three initial years of the summer trawls, and decreases to a minimum in 2003. From 2004 onward, a general increasing trend is seen, almost returning to the initial catch weight level observed in 1991. Evidence for temporal autocorrelation is present, as successive years have similar median CPUE values. In the first three years, the median CPUE is relatively equal. The median CPUE in 2005 is closer to that of 2004 and 2006, than that of 2003 and and 2007.

In addition to the preliminary exploration through the box plots, an autocorrelation plot is constructed. The autocorrelation plot shows how the average CPUE (in log scale) is correlated with itself at different lag points. For example, when the series is lagged by 1, each observation of CPUE is

paired with the catch weight of the previous year. The correlation is calculated between the two series, and is displayed on the plot. Figure 2.10 shows noticeable correlation between observations when lagged by both one year, and four years.

The evidence described above is indicative of an autoregressive temporal effect within hoki CPUE, which needs to be accounted for in our model of catch per unit effort.
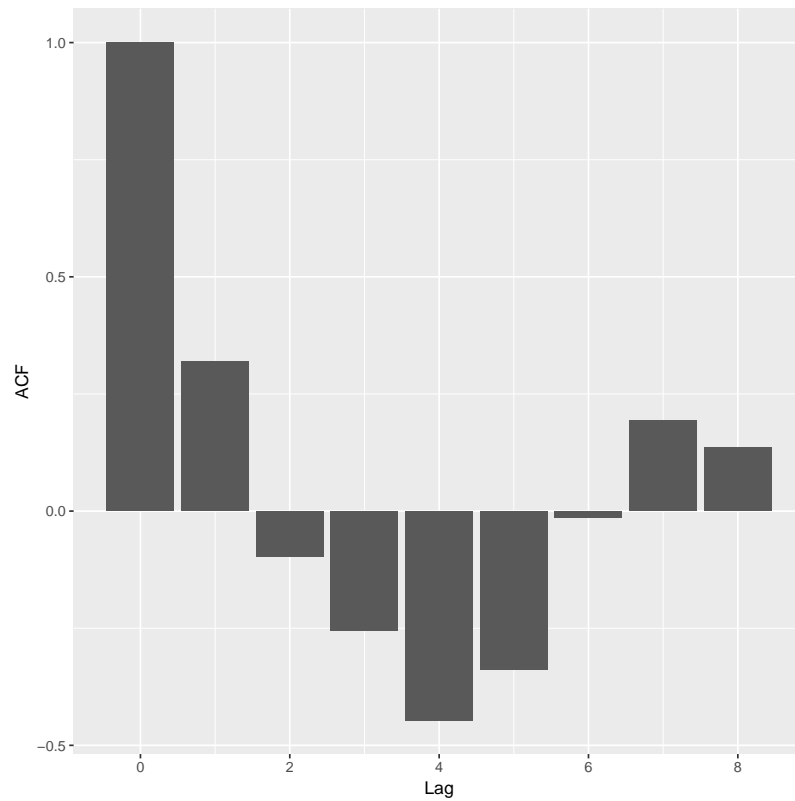
Figure 2.10: Plot of the autocorrelation of CPUE in kg/km^2 between years of increasing lags

2.6.2   Spatial Autocorrelation

In addition to temporal autocorrelation, the location of hoki within the survey area is expected to be spatially autocorrelated. In nature, the distribution of species is commonly dependent on characteristics of the area, in which the species resides. This may be due to reasons such as migration and recruitment, differences in terrain, habitability and species preferences (Abensperg-Traun & Boer, 1990; Gratwicke & Speight, 2005; Morgan et al., 2006). Visualizing spatial autocorrelation is more complicated compared to temporal autocorrelation because space extends in two dimensions. We can take an initial look at the distribution of hoki catch weight
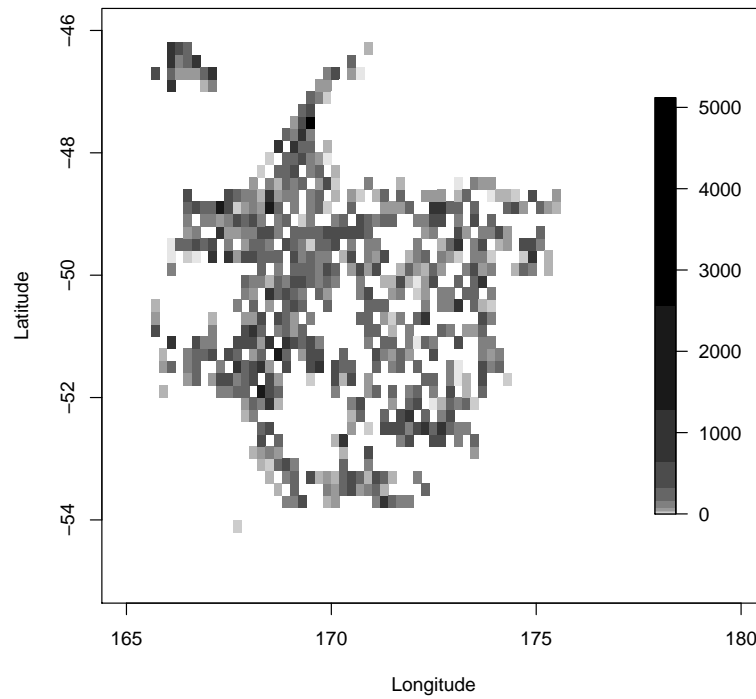
Figure 2.11: Distribution of CPUE aggregated by year.

over the sub-Antarctic area, by aggregating the observations over time. Figure 2.11 shows that CPUE clearly differs over area. At Puysegur Bank (top left), the CPUE is highest. CPUE decreases toward the bottom right. The band of no CPUE between Puysegur Bank and below Snares shelf is due to the fact that there is no survey coverage there, and can be ignored.

In addition to the plots of spatial distribution of CPUE a common way to measure spatial autocorrelation globally is to use Moran's I (Moran, 1950). Moran's I is an adaptation of the popular Pearson product moment correlation coefficient that allows us to measure spatial autocorrelation for

a univariate series (Moran, 1950). The statistic is:

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \tag{2.4}$$

where n is the number of observations, w_{ij} is the weight between observations i and j, and S_0 is the sum of all the weights. The choice of weight function between observations is important, as it allows us to specify how close two observations are in space. Observations which are closer in space are expected to have similar values of CPUE, and are given a larger weight compared to observations further apart. We used a function of the inverse of the distance between the pair of tows as the weights. To allow for the fact that some trawls can be quite close to one another, an adjustment of 1000 m was used to ensure that the value of I is not distorted (Briggs, 2010). We calculated w_{ij} using

$$w_{ij} = \frac{1000}{1000 + d_{ij}}, \tag{2.5}$$

where d_{ij} is the distance between the midpoints of tow i and tow j.
Other weight functions have been used in the literature, for example $w_{ij} = \exp(-\frac{d_{ij}}{d})$, which specifies quasi-global correlation between points derived from maximum entropy models (Chen, 2012).
Moran's I takes values between -1 and 1. If I is positive, then there is positive spatial autocorrelation within the sample, whereas, if it is negative, then there is negative spatial autocorrelation. If I is close to its expected value, then we observe no spatial autocorrelation within the sample.
We can test whether significant spatial autocorrelation exists within hoki CPUE by testing the null hypothesis that there is not spatial autocorrelation, against the hypothesis that there is.
By standardizing I by its expected value and standard error, we arrive at a convenient test statistic:

$$z_I = \frac{I - \mathrm{E}(I)}{\sqrt{\mathrm{Var}(I)}}, \tag{2.6}$$

where z_I is the test statistic with standard Normal distribution, $E(I)$ is the expected value of Moran's I, and $\sqrt{\text{Var}(I)}$ is the standard error.

The expected value of Moran's I is defined as:

$$E(I) = \frac{-1}{n-1}, \tag{2.7}$$

where, n is the number of observations (Moran, 1950). The standard error for Moran's I is $\sqrt{E(I^2) - E(I)^2}$.

For the hoki CPUE data, we obtained $I = 0.034$, and a corresponding $E(I) = 0.00076$.

We computed the standard error through bootstrapping (Efron, 1981). After computing both the observed and expected value of I for the CPUE observations, the CPUE observations were re-sampled and assigned to the weights used in calculating the observed I. We called this sample the bootstrap sample. Moran's I was then computed for the bootstrap sample. Resampling and calculating Moran's I was repeated 200 times. Finally, the standard error for the observed I was taken as the standard error of the 200 bootstrapped Moran's I. The standard error was found to be 0.00256.

We therefore obtained a test statistic of $z_I = 13.8$, with a corresponding p-value of $P(Z \geq 13.8) + P(Z \leq 13.8) < 0.0001$.

The value we obtained ($I = 0.034$) was statistically significant (p-value < 0.0001). The value is small and positive, indicating that there is some spatial autocorrelation in hoki CPUE. This result is in agreement with visual examination of the spatial distribution plot (Figure 2.11).

Measuring global spatial autocorrelation is beneficial as a starting point, however, we can also measure local autocorrelation. Local Moran's I is a measure of spatial autocorrelation between one observation and all other observations surrounding it (Anselin, 1995), and is given by

$$I_i = \frac{(x_i - \bar{x}) \sum_{j=1}^{n} w_{ij}(x_j - \bar{x})}{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}, \tag{2.8}$$

where I_i is a measure of autocorrelation for observation i, $\bar{x}$ is the mean

of the observations x, and w_{ij} is the weight between observations i and j, calculated using Equation 2.5.

Local Moran's I allows for visualization of areas where similar CPUE observations are found. By taking the sum of all I_i, we arrive at the global measure of autocorrelation described above.

Local Moran's I was then plotted against longitude and latitude. The colour gradient is now representative of the value of Local Moran's I (Figure 2.12). We can see that local Moran's I is most positive at Puysegur Bank and the northeast region, indicating positive spatial autocorrelation in this area. Trawls in these regions have similar values of hoki CPUE. Very few observations have I_i values less than zero, indicating that most spatial autocorrelation was positive.

2.6.3   Spatio-Temporal Autocorrelation

In addition to temporal and spatial autocorrelation, there is the possibility of an interaction between the two. This so-called spatio-temporal autocorrelation is the dependence of catch weight between nearby areas, within time. This is easiest visualized as a series of plots, broken down by year (Figure 2.13). Here we can see that CPUE changes over space within each year as represented by the changing colouration over space. Furthermore, the patterns change over time, which is indicative of spatio-temporal autocorrelation. For example, we see that CPUE in the west of the map starts off high in the years 1991, 1992, and 1993, but is much smaller in 2000 onward. From the year 2000, CPUE in this region appears to decrease reaching a low point in 2003, but increases from 2004 onward. This is indicative that there is some interaction between space and time within hoki CPUE.

Further evidence of spatio-temporal autocorrelation can be seen in Figure 2.14, which is a plot of Local Moran's I for each observation within each year. The plot provides evidence of spatial autocorrelation within each year, as shown by the regions of dark and light indicating changes in spa-
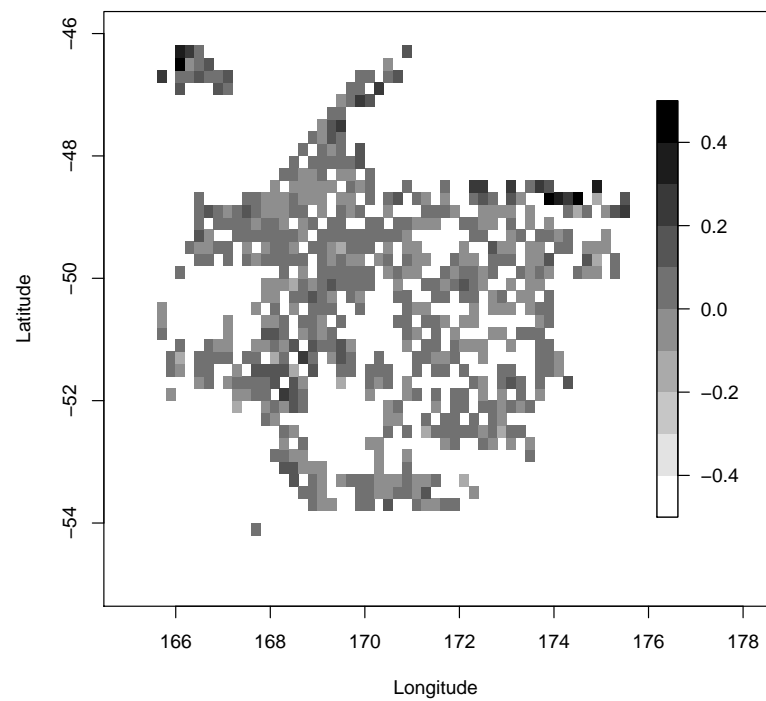
Figure 2.12: Distribution of local Moran's I for hoki CPUE observations.
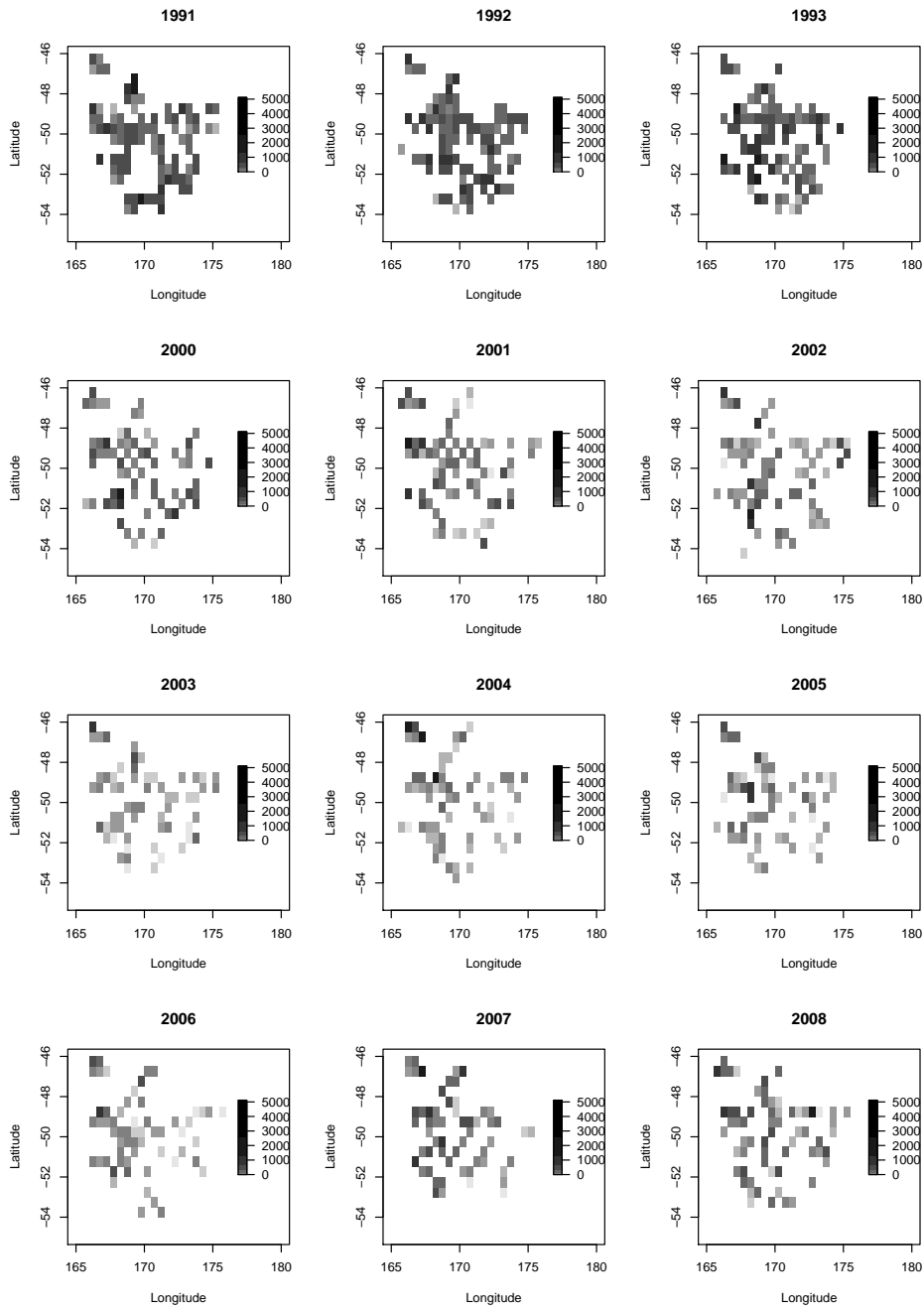
Figure 2.13: Hoki CPUE distribution over space for the year 1991 – 1993, and 2000 – 2008.

tial autocorrelation. We also see that the spatial autocorrelation patterns change from year to year. This reinforces our evidence for spatio-temporal autocorrelation within hoki CPUE.

In our data exploration, we found evidence of effects on hoki CPUE due to both the year of trawl and stratum. In addition, our presumption of temporal, spatial and spatio-temporal autocorrelation was validated by the evidence provided in Sections 2.6.1, 2.6.2, and 2.6.3. When constructing models for CPUE, autocorrelation must not be ignored. Gaussian Markov random fields provide a useful means to model the effects of trawl year, effects of stratum, and even CPUE itself. Gaussian Markov random fields are particularly useful in this case, as they allow us to impose the autocorrelation structures we see in our data. In Chapter 3, we detail the theoretical framework surrounding Gaussian Markov random fields.

Figure 2.14: Local Moran's I for hoki CPUE by year.

Chapter 3

Methodology

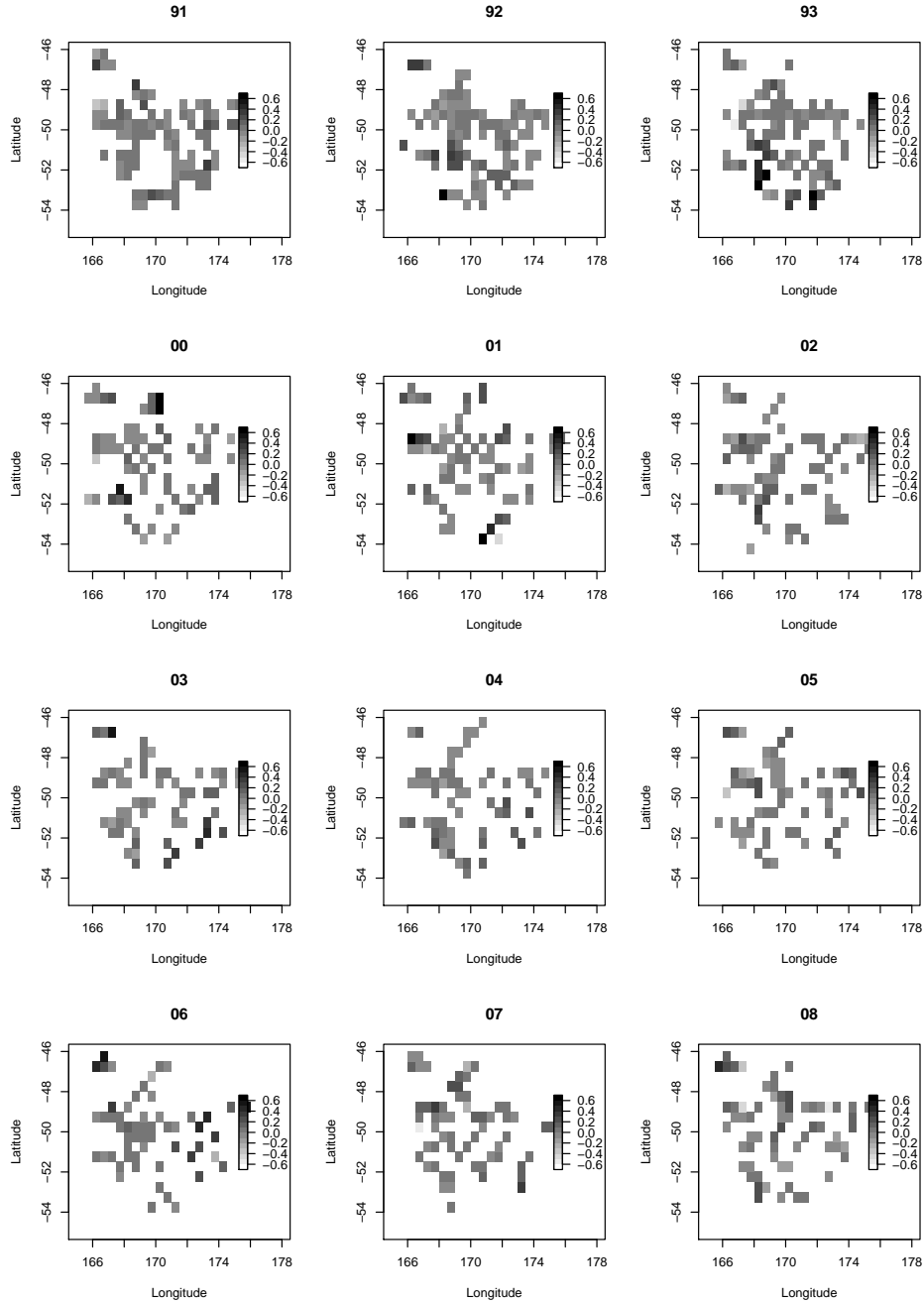This chapter outlines the theoretical framework and methodology used in this thesis. We begin with an introduction to hierarchical models and the Bayesian approach. This leads into the theory underpinning Gaussian Markov random fields including a discussion on precision matrix structures. An introduction to Hamiltonian Monte Carlo follows and an example comparing Hamiltonian Monte Carlo to Markov chain Monte Carlo is provided. A section on diagnostic analysis is also given. The chapter concludes with a section on model comparison and assessment techniques.

3.1   Bayesian Hierarchical Models

Methods for modelling a species catch per unit effort (CPUE) were reviewed in section 1.2.4. Such methods can be applied using either a frequentist approach or a Bayesian approach. Bayesian methods have been used to derive standardized abundance indices with the use of Bayesian hierarchical models (BHMs) (Bertram et al., 2015; Raghavan et al., 2016). The benefit of using a hierarchical structure to model CPUE is the ability to explicitly incorporate different variance components of the response. We are aiming to explain some variability in CPUE that is due to where the catch was obtained, and when. It is reasonable to assume that a relation-

ship exists between catch weights which are close together in distance, as well as in time. Evidence of such relationships in the data is given in section 2.5. In addition, Bayesian methods allow for easier computation of parameter estimates, compared to using maximum likelihood estimation. As such, we decided to use the Bayesian approach.

Before we build our Bayesian hierarchical model (BHM), we must first lay the groundwork for the Bayesian approach. The following subsection presents Bayes' theorem.

3.1.1   Bayesian Methods

Let $\boldsymbol{y}$ be a vector of observations from some distribution depending on fixed potential predictors $\boldsymbol{x}$ and unknown parameters $\boldsymbol{\theta}$. We first start with a joint probability model for $\boldsymbol{y}$, $\boldsymbol{x}$, and $\boldsymbol{\theta}$, given by

$$\pi(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\theta}) = \pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{x}), \tag{3.1}$$

where we refer to $\pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})$ as the data likelihood and $\pi(\boldsymbol{\theta}|\boldsymbol{x})$ as the prior distribution.

By simply conditioning the joint distribution of unknown parameters on the observed data $\boldsymbol{y}$, we arrive at an expression for the posterior density of the parameters:

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{x}) = \frac{\pi(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\theta})}{\pi(\boldsymbol{y}, \boldsymbol{x})} = \frac{\pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{x})}{\pi(\boldsymbol{y}, \boldsymbol{x})}, \tag{3.2}$$

where $\pi(\boldsymbol{y}, \boldsymbol{x}) = \int \pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{x})d\boldsymbol{\theta}$ is the marginal distribution of the data, which does not depend on any parameters. As such, we can rewrite the posterior density in its most recognizable form:

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{x}) \propto \pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{x}). \tag{3.3}$$

Equation 3.2 and 3.3 provide the theoretical framework for Bayesian statistics (Gelman et al., 2014a). From here, we can build hierarchical models.

3.1.2 Bayesian Hierarchical Model

From equation 3.3, a Bayesian hierarchical model (BHM) can be built. Let $\boldsymbol{y}$ be a response variable that is divided into groups or clusters according to space or time. Now consider a location, or a time point, j, that has data vector $\boldsymbol{y}_j$ and parameter vector $\boldsymbol{\theta}_j$. Each parameter vector $\boldsymbol{\theta}_j$ is regarded as a random draw from a population distribution governed by some parameter ϕ, such that $\pi(\boldsymbol{\theta}|\phi) = \prod_{j=1}^{J} \pi(\boldsymbol{\theta}_j|\phi)$. Therefore, our BHM has the form:

$$\pi(\boldsymbol{\theta}, \phi|\boldsymbol{y}, \boldsymbol{x}) \propto \pi(\boldsymbol{y}|\boldsymbol{\theta}, \phi, \boldsymbol{x})\pi(\boldsymbol{\theta}|\phi, \boldsymbol{x})\pi(\phi|\boldsymbol{x}) \qquad (3.4)$$

The posterior distributions for the parameters of a BHM are needed in order to perform Bayesian inference. From equation 3.4, three components are required to compute the posterior density.

At the first stage, a suitable distribution for the observed data is chosen. The choice of data distribution will depend on the type of response variable in question. For example, if the responses are counts, then a Poisson distribution might be appropriate. The second stage involves carefully choosing a prior distribution, $\pi(\boldsymbol{\theta}|\phi, \boldsymbol{x})$ for the parameters, $\boldsymbol{\theta}$, which drive the observations. At this point, we can introduce prior assumptions we have about latent processes which govern the distribution of the response variable. In this thesis, we wish to incorporate spatial and temporal autocorrelation into a model for hoki CPUE. One way of doing this is to use Gaussian Markov random fields. Finally suitable hyperprior distributions, $\pi(\phi|\boldsymbol{x})$, are selected for the parameters, ϕ, of the prior distributions.

3.2 Gaussian Markov Random Fields

A Gaussian random field (GRF) is essentially a random vector that comes from a multivariate normal distribution. By satisfying additional assumptions of conditional independence, we can focus our attention on a special

case of GRFs, the so-called Gaussian Markov random field (GMRF). GM-
RFs have many uses in the applications of time series, longitudinal and
survival data, image analyses, and spatial statistics to name a few (Rue &
Held, 2005). The need for understanding GMRFs is becoming increasingly
more important, as these techniques become more popular. Our goal is to
implement a BHM to model hoki CPUE where the temporal and spatial
effects are modelled by GMRFs. The motivation is that the abundance in-
dices obtained will be more precise than those obtained by other methods
(described in Section 1.2.4), and will be simpler and quicker to compute.
Before implementation, we must define the theory that underpins GMRFs.
We begin with conditional independence.

3.2.1   Conditional Independence

When looking at variables that evolve over space and/or time, it is im-
portant to note that points closer in space/time are assumed to have a
higher correlation than those separated by large distances. By introduc-
ing Markovian properties, we can take advantage of the assumption that
non-neighbouring points are independent.

Let $\boldsymbol{x} = (x_1, x_2, x_3)^T$ be a random vector, then x_1 and x_2 are conditionally
independent given x_3 if, for a known value of x_3, discovering x_2 gives us
no new information about the distribution of x_1. Under the assumption
of conditional independence of x_1 and x_2, the joint density of $\boldsymbol{x}$ must have
the following representation

$$\pi(\boldsymbol{x}) = \pi(x_1|x_3)\pi(x_2|x_3)\pi(x_3). \tag{3.5}$$

A common way to display the conditional independence relationships be-
tween variables is through the use of undirected graphs. We denote an
undirected graph as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ is the set of nodes in the graph,
and $\mathcal{E}$ is the set of edges $\{i, j\}$, where $i, j \in \mathcal{V}$ and $i \neq j$. An example of
an undirected graph is shown in Figure 3.1. Nodes that share an edge are
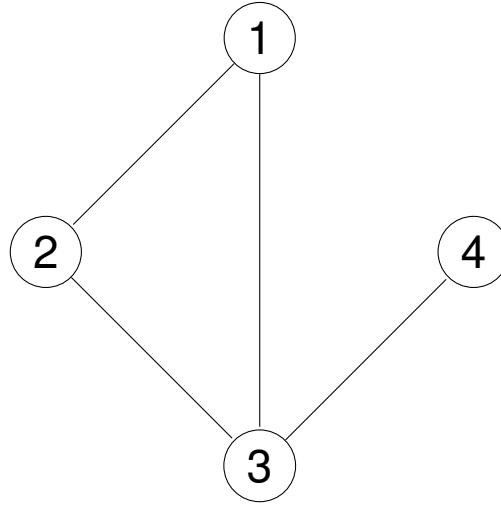considered to be conditionally dependent, given all other nodes.

Figure 3.1: Undirected graph $\mathcal{G} = (\mathcal{V} = \{1, 2, 3, 4\}, \mathcal{E} = \{\{1, 2\}, \{1, 3\}, \{2, 3\}, \{3, 4\}\})$

We can now formally define a GMRF.

3.2.2 Definition of the Gaussian Markov Random Field

A random vector $\boldsymbol{x} = (x_1, ..., x_n)^T \in \mathbb{R}^n$ is called a Gaussian Markov random field with respect to a labelled graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with mean $\boldsymbol{\mu}$ and positive definite precision matrix $\boldsymbol{Q}$, if and only if its density has the form

$$\pi(\boldsymbol{x}) = (2\pi)^{-\frac{n}{2}} |\boldsymbol{Q}|^{\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{Q}(\boldsymbol{x} - \boldsymbol{\mu}) \right) \tag{3.6}$$

and

$$Q_{ij} \neq 0 \iff \{i, j\} \in \mathcal{E} \text{ for all } i \neq j.$$

Notice that the density is defined using a precision matrix $\boldsymbol{Q}$, rather than a covariance matrix Σ. The precision matrix allows us to visualize the conditional independence between the elements of the GMRF. Formally and in general for any GMRF, if $Q_{ij} = 0$, for $i \neq j$, then x_i and x_j are conditionally independent given the other variables x_k where $k \neq i$ and $k \neq j$, and vice versa. This is illustrated in graphs such as Figure 3.1,

where two nodes not sharing an edge indicates conditional independence (and hence a zero in the precision matrix). In general, the structure of the precision matrix can be directly translated from the undirected graph.

The use of the precision matrix is beneficial because the elements of $\boldsymbol{Q}$ have nice conditional interpretations (Rue & Held, 2005). The diagonal elements of $\boldsymbol{Q}$ are the conditional precisions of x_i given every other term, $\boldsymbol{x}_{-i}$. The off diagonal elements give information about the conditional correlation between x_i and x_j, given every other term $\boldsymbol{x}_{-ij}$. These are results of the following theorem (Rue & Held, 2005).

Theorem 1 *Let $\boldsymbol{x}$ be a GMRF with respect to the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with mean $\boldsymbol{\mu}$ and positive definite precision matrix $\boldsymbol{Q}$, then*

$$E(x_i|\boldsymbol{x}_{-i}) = \mu_i - \frac{1}{Q_{ii}} \sum_{j:j\sim i} Q_{ij}(x_j - \mu_j), \tag{3.7}$$

$$Prec(x_i|\boldsymbol{x}_{-i}) = Q_{ii}, \tag{3.8}$$

and

$$Corr(x_i, x_j|\boldsymbol{x}_{-ij}) = -\frac{Q_{ij}}{\sqrt{Q_{ii}Q_{jj}}}, \quad i \neq j. \tag{3.9}$$

This is incredibly convenient, because it allows us to easily impose a particular correlation structure on a process variable (such as temporal or spatial), by treating it as a Gaussian Markov random field. We look at possible precision structures in the next section.

3.2.3   Precision Structures

Gaussian Markov random fields provide a flexible approach for modelling the dependence between latent parameters such as temporal or spatial effects, and implicitly, the dependence between observed responses, y_i. This dependence can be due to various processes such as temporal, spatial, or spatiotemporal interactions (Rue & Martino, 2007). The following subsections provide an overview of some precision structures that will be used

in Chapter 4 to reflect the various dependencies within the observed hoki CPUE data.

Temporal Process

It is well known that temporal dependence exists within data concerned with measuring population abundance over time (Preston, 1960; Magurran, 2007). The day, month, season, or year effect of an observation can represent the effect of other latent unmeasured effects that vary over time (such as climate fluctuations, or global policy changes etc.). Because of this, year of trawl was considered a useful predictor in modelling changes in CPUE over time as well as temporal dependence of abundance observations. From Section 2.6.1, we saw that temporal autocorrelation exists within the hoki CPUE observations.

It is logical to consider that observations in the future will depend on those in the past. Such a pattern is called an autoregressive structure, which is the type of structure we wish to impose on our observations (see Section 2.6.1).

Let $\boldsymbol{\tau} = (\tau_1, ..., \tau_T)$, where T is the number of time points, represent the temporal process driving some observable response variable. In this case, let τ_i represent the effect on observations recorded in year i of the survey. We can model this process as a GMRF with a mean of zero, and a precision matrix $\boldsymbol{Q}$, which has a first order autoregressive (AR(1)) structure. The model equation for an AR(1) process is shown below (Rue & Held, 2005):

$$\tau_i = \rho\tau_{i-1} + \varepsilon_i, \quad \varepsilon_i \sim \mathrm{N}(0, \phi), \quad |\rho| < 1. \tag{3.10}$$

The joint density of $\boldsymbol{\tau}$ is therefore a GMRF:

$$\boldsymbol{\tau} \sim \mathrm{MVN}(\boldsymbol{0}, \boldsymbol{Q}^{-1}), \tag{3.11}$$

with

$$\boldsymbol{Q} = \phi \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho & 0 & ... & 0 \\ -\rho & 1+\rho^2 & -\rho & ... & 0 \\ 0 & -\rho & 1+\rho^2 & ... & 0 \\ ... & ... & ... & ... & ... \\ 0 & ... & -\rho & 1+\rho^2 & -\rho \\ 0 & ... & 0 & -\rho & 1 \end{pmatrix}, \tag{3.12}$$

where $\boldsymbol{Q}$ is a $T \times T$ matrix, ϕ is the precision parameter of the temporal effect, and ρ is a measure of the temporal autocorrelation. Hence by using equation 3.8, we arrive at the full conditional precisions shown below:

$$\text{Prec}(\tau_i|\boldsymbol{\tau}_{-i}) = \frac{1}{\text{Var}(\tau_i|\boldsymbol{\tau}_{-i})} = \begin{cases} \frac{\phi}{1-\rho^2}, & \text{for } i = 1, T \\ \frac{\phi(1+\rho^2)}{1-\rho^2}, & \text{for } i = 2, ..., T-1. \end{cases} \tag{3.13}$$

We now move on to look at precision structures to model spatial dependence in the following subsections.

Spatial Process at the Areal Level

Here, we look at a possible precision structure to describe the spatial process at an areal level. A spatial process is defined at an areal level if the entire surveyed area is stratified. The value of an observation may be affected by which stratum it belongs to. As described in Chapter 2, hoki CPUE observations were recorded at an areal level.

A precision matrix with a conditional autoregressive (CAR) structure (Besag et al., 1991) is ideal to specify the spatial dependence of observations recorded at an areal level. The CAR model is derived from a GMRF specified in its full conditional form with the use of Brook's Lemma (Rue & Held, 2005; White & Ghosh, 2009).

A GMRF can be specified by its full conditional distributions, $\{\pi(x_i|\boldsymbol{x}_{-i})\}$, as opposed to using its mean and precision matrix. In order to do so, the full conditional distributions must satisfy certain conditions in order for

them to correspond to a valid GMRF (Rue & Held, 2005). Suppose the full conditional distributions are specified to be Normal distributions with:

$$\mathrm{E}(x_i|\boldsymbol{x_{-i}}) = \mu_i - \sum_{j:j\sim i} \beta_{ij}(x_j - \mu_j), \tag{3.14}$$

and

$$\mathrm{Prec}(x_i|\boldsymbol{x_{-i}}) = \zeta_i > 0, \tag{3.15}$$

for $i = 1, ..., n$, for some $\{\beta_{ij} \neq 0, i \neq j\}$, and vectors $\boldsymbol{\mu}$ and $\boldsymbol{\zeta}$. Here, $j \sim i$, means any point j which neighbours the point i. The neighbourhood structure is defined implicitly through the non-zero terms $\{\beta_{ij}\}$.

In order for a joint density $\pi(\boldsymbol{x})$ that gives rise to the full conditionals to exist, the parameters β_{ij} and ζ_i must satisfy consistency requirements. By comparing the terms above to the ones seen in 3.7, and 3.8, we see that if the entries of $\boldsymbol{Q}$ are

$$Q_{ii} = \zeta_i, \quad \text{and} \quad Q_{ij} = \zeta_i\beta_{ij}, \tag{3.16}$$

and $\boldsymbol{Q}$ is symmetric, then we have a candidate for a joint density that gives rise to the specified full conditionals, provided $\boldsymbol{Q}$ is positive definite (Rue & Held, 2005). From this, a useful theorem follows that gives way to a precision matrix that will be used to model the spatial dependence in the data used in this thesis.

The following theorem was provided in Rue & Held (2005).

Theorem 2 *Given the n Normal full conditionals with conditional mean and precision as given by expressions 3.15 and 3.16, then $\boldsymbol{x}$ is a GMRF with respect to a labelled graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with mean $\boldsymbol{\mu}$ and precision matrix $\boldsymbol{Q} = (Q_{ij})$, where*

$$Q_{ij} = \begin{cases} \zeta_i\beta_{ij} & \text{if } i \neq j \\ \zeta_i & \text{if } i = j \end{cases}, \tag{3.17}$$

provided $\zeta_i\beta_{ij} = \zeta_j\beta_{ji}$ for $i \neq j$ and $\boldsymbol{Q}$ is positive definite.

This result is proven using Brook's Lemma, which follows below.

Lemma 1 *Let $\pi(\boldsymbol{x})$ be the density for $\boldsymbol{x} \in \mathbb{R}^n$, $\pi(\boldsymbol{x}')$ be the density for $\boldsymbol{x}' \in \mathbb{R}^n$, and define $\Omega = \{\boldsymbol{x} \in \mathbb{R}^n : \pi(\boldsymbol{x}) > 0\}$. Let $\boldsymbol{x}, \boldsymbol{x}' \in \Omega$, then*

$$
\begin{aligned}
\frac{\pi(\boldsymbol{x})}{\pi(\boldsymbol{x}')} &= \prod_{i=1}^{n} \frac{\pi(x_i|x_1, ..., x_{i-1}, x'_{i+1}, ..., x'_n)}{\pi(x'_i|x_1, ..., x_{i-1}, x'_{i+1}, ..., x'_n)} \\
&= \prod_{i=1}^{n} \frac{\pi(x_i|x'_1, ..., x'_{i-1}, x_{i+1}, ..., x_n)}{\pi(x'_i|x'_1, ..., x'_{i-1}, x_{i+1}, ..., x_n)}.
\end{aligned}
\tag{3.18}
$$

We can now prove Theorem 2.

Proof. Assume that $\boldsymbol{\mu} = \boldsymbol{0}$, and fix $\boldsymbol{x}' = \boldsymbol{0}$. Then expression 3.18 simplifies to

$$
\begin{aligned}
\log \frac{\pi(\boldsymbol{x})}{\pi(\boldsymbol{0})} &= -\frac{1}{2} \sum_{i=1}^{n} \zeta_i x_i^2 - \sum_{i=2}^{n} \sum_{j=1}^{i-1} \zeta_i \beta_{ij} x_i x_j \\
&= -\frac{1}{2} \sum_{i=1}^{n} \zeta_i x_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \zeta_i \beta_{ij} x_i x_j,
\end{aligned}
\tag{3.19}
$$

which implies that $\zeta_i \beta_{ij} = \zeta_j \beta_{ji}$ for $i \neq j$. Therefore, the log density of $\boldsymbol{x}$ can be expressed as

$$
\log \pi(\boldsymbol{x}) = \text{constant} - \frac{1}{2} \sum_{i=1}^{n} \zeta_i x_i^2 - \frac{1}{2} \sum_{i \neq j} \zeta_i \beta_{ij} x_i x_j,
\tag{3.20}
$$

which corresponds to the density of a zero mean, multivariate Normal variable, $\boldsymbol{x}$, provided that $\boldsymbol{Q}$ is positive definite. The precision matrix has entries $Q_{ij} = \zeta_i \beta_{ij}$ for $i \neq j$ and $Q_{ii} = \zeta_i$. $\qquad\square$

If we define $\beta_{ii} = 0$ in the expression above, then in matrix terms, the precision matrix becomes:

$$
\boldsymbol{Q} = \text{diag}(\boldsymbol{\zeta})\big(\boldsymbol{I} + (\beta_{ij})\big),
\tag{3.21}
$$

where $\boldsymbol{Q}$ is a $J \times J$ precision matrix, diag($\boldsymbol{\zeta}$) is a $J \times J$ matrix with $\zeta_1, ..., \zeta_J$ down the diagonal, and zero elsewhere, and (β_{ij}) is a $J \times J$ matrix with zero down the diagonal, and β_{ij} elsewhere, and J is the number of areas. Then diag($\boldsymbol{\zeta}$)$\boldsymbol{I} \propto \boldsymbol{D}$ and diag($\boldsymbol{\zeta}$)$(\beta_{ij}) \propto \boldsymbol{A}$, and $\boldsymbol{D}$ is a $J \times J$ diagonal matrix of the numbers of neighbours for each area, and $\boldsymbol{A}$ is a $J \times J$ adjacency matrix with entries $a_{ij} = 1$ if area i and j are neighbours, and 0 otherwise. We rewrite equation 3.21 into a form that can be implemented in the models:

$$\boldsymbol{Q} = \eta(\boldsymbol{D} - p\boldsymbol{A}), \tag{3.22}$$

where $\boldsymbol{D}$ and $\boldsymbol{A}$ are $J \times J$ matrices defined as above, η is a precision parameter, and p is a parameter describing the relative strength of spatial dependence (White & Ghosh, 2009). In the paper by Hrafnkelsson & Cressie (2003), it is stated that p is chosen to be such that $0 \leq p < 1$, to ensure that $\boldsymbol{Q}$ is positive definite.

Spatial Process at the Point Referenced Level

When the geographic location of an observation is recorded, we call it point referenced. CPUE observations in the hoki dataset have been recorded at a point referenced level (see Chapter 2). The observations were recorded with the start and end locations of the corresponding trawls. We can incorporate point referenced spatial structure into our models by using a precision matrix that has a similar form to the conditionally autoregressive precision matrix described for areal data. This approach was explored by White & Ghosh (2009), and by Paciorek et al. (2013).

The precision matrix allows for correlations between each pair of points, rather than between groups of points. As a result, the size of the precision matrix is a lot larger ($n \times n$, where n is the number of observations). In addition, the adjacency matrix entries are defined as functions of the distances between pairs of points, as long as the resulting precision matrix is

positive definite. The precision matrix is assumed to have the following
structure (White & Ghosh, 2009),

$$\boldsymbol{Q} = \delta(\boldsymbol{G} - \kappa\boldsymbol{C}), \tag{3.23}$$

where $\boldsymbol{C}$ is an $n \times n$ distance matrix, $\boldsymbol{G}$ is an $n \times n$ matrix with diagonal en-
tries equal to the sum of the rows of $\boldsymbol{C}$ and zero elsewhere, δ is a precision
parameter, and κ is a parameter describing the spatial correlation (White
& Ghosh, 2009).

Each element of $\boldsymbol{C}$, c_{ij}, is defined as a function of distance between obser-
vations i and j,

$$\boldsymbol{C} = (c_{ij})_{n \times n}, \tag{3.24}$$

where n is the number of observations, and

$$c_{ij} = \begin{cases} 1, & \text{if } d_{ij} \leq d_l \\ f(d_{ij}) & \text{if } d_l < d_{ij} \leq d_u \\ 0 & \text{if } d_{ij} > d_u \end{cases}, \tag{3.25}$$

where d_{ij} is the euclidean distance between observations i and j. The term
d_l is a lower distance threshold such that if the distance between observa-
tions i and j is below it, then they are assumed to be neighbouring (ad-
jacency score of 1, as defined in Section 3.2.3). The term d_u is an upper
distance threshold such that if the distance between observations i and j
is above it, then they are assumed to have no correlation.

We propose a function, f, which gives an adjacency score that is propor-
tional to the inverse pairwise distance between points. It assigns each pair-
wise distance a number between 0 and 1, where 0 represents no correlation
between observations, and 1 represents neighbouring observations. Points
that are closer to each other get a higher adjacency score than those further

apart, such that

$$f(d_{ij}) = \frac{\frac{1}{d_{ij}} - \frac{1}{d_u}}{\frac{1}{d_l} - \frac{1}{d_u}} \tag{3.26}$$

and

$$c_{ii} = 0 \text{ for all } i. \tag{3.27}$$

The function is bounded by a lower and upper limit, d_l and d_u, such that 90 % of the distances between observations lies between d_l and d_u.

Other distance functions have been proposed, such as the one suggested in White & Ghosh (2009), and Paciorek et al. (2013) shown below:

$$f(d_{ij}, \alpha, d_u) = d_{ij}^{log(\alpha)/log(d_u)}, \tag{3.28}$$

where d_u is the upper distance limit, and α is a fixed term (usually set to 0.05) that specifies the amount of spatial dependence when the distance between observation i and j, d_{ij}, is equal to the upper limit.

A comparison of the two distance functions defined above was carried out for hoki data. A histogram was used to compare the spatial dependence weight given by each distance function for each pair of observations as shown in Figure 3.2.

There is a small difference between the two functions. The histogram of the function from the literature shows that weighting between points decays slower than our distance function. Observations separated by large distances are assumed to be less correlated using our function compared to the literature's function. We therefore use our proposed distance function.

In White & Ghosh (2009), a stochastic neighbourhood CAR (SNCAR) is suggested, which uses the distance function in equation 3.28, but lets the upper distance limit, d_u be a random unknown quantity to be estimated. This approach was not explored in this thesis.
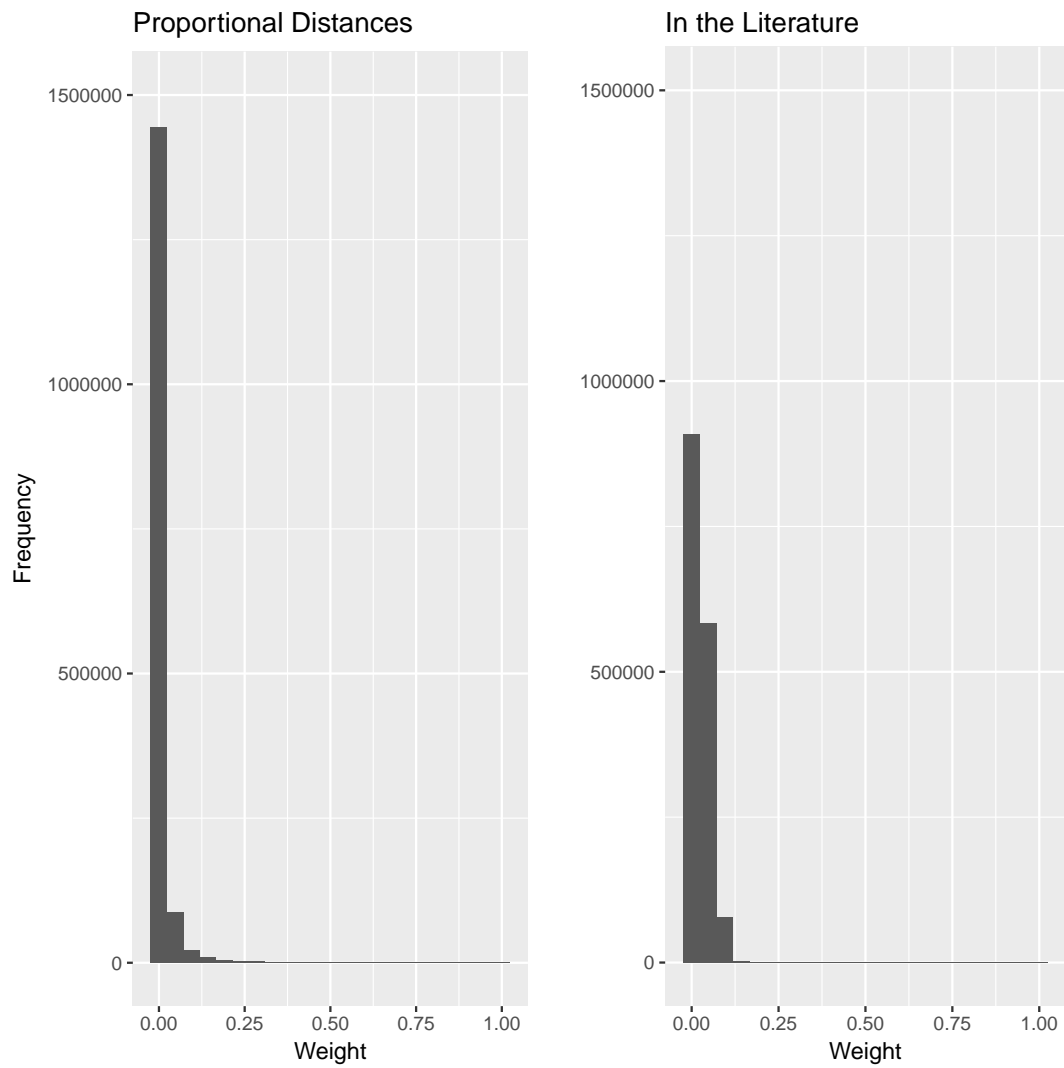
Figure 3.2: Histograms of the distance weights calculated by our function (Proportional) and the function in White & Ghosh (2009) (Literature).

Spatial Process at the Blocked Point Reference Level

When the location of an observation and the stratum it was sampled from are both reported we can impose further correlation assumptions. We assume that each pair of CPUE observations are correlated differently depending on stratum. Furthermore, points that are not from the same stratum are assumed to be independent. We can build a CAR precision matrix similar to the one in section 3.2.3. The difference here is that the precision matrix is now a block matrix, made up of precision matrices for each stratum.

$$
\boldsymbol{Q} = \begin{pmatrix} \boldsymbol{Q_1} & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q_2} & \dots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \dots & \boldsymbol{Q_J} \end{pmatrix}, \tag{3.29}
$$

where

$$
\boldsymbol{Q_j} = \delta_j(\boldsymbol{G_j} - \kappa_j \boldsymbol{C_j}), \tag{3.30}
$$

for $j = 1, ..., J$, and J is the number of strata. The matrices $\boldsymbol{G}_j$ and $\boldsymbol{C}_j$ have the same form as $\boldsymbol{G}$ and $\boldsymbol{C}$ in Equation 3.23.

We have presented four precision structures that will be used to account for spatial and temporal autocorrelation at both the areal and point referenced level. We now provide the methodology for sampling from posterior distributions.

3.3 Bayesian Posterior Sampling

3.3.1 Markov Chain Monte Carlo

Markov chain Monte Carlo (MCMC) has its origins in the field of physics. A variant of this technique, Hamiltonian Monte Carlo, involves the use of Hamiltonian dynamics to sample from a posterior distribution (Neal et al.,

2011). Unlike Metropolis random walk or Gibbs' sampler, the Hamiltonian dynamics variant does not exhibit random walk behaviour. In addition, Hamiltonian Monte Carlo (HMC) is not sensitive to correlated parameters, which is a problem in MCMC. In fisheries science, it is common to see correlated parameters. Of concern in this case, is the correlation between spatial, and temporal parameters. In order to see how Hamiltonian dynamics can be used in MCMC, we will first look at its origins.

3.3.2 Hamiltonian Dynamics

In order to describe how an object (for example, an electron or molecule) moves throughout a system we focus on two physical attributes of the object. The object's position (denoted by the d-dimensional vector $\boldsymbol{q}$), and the object's momentum (its mass multiplied by its velocity, denoted by the d-dimensional vector $\boldsymbol{p}$) are used to describe the object's motion through time. For each location, the object has an associated potential energy denoted by $U(\boldsymbol{q})$. Likewise, for each momentum, the object has an associated kinetic energy denoted by $K(\boldsymbol{p})$. Together, these two energy components are combined to give the total energy of the object, called the Hamiltonian, which is denoted by:

$$H(\boldsymbol{q}, \boldsymbol{p}) = U(\boldsymbol{q}) + K(\boldsymbol{p}) \tag{3.31}$$

Partial derivatives on $\boldsymbol{q}$ and $\boldsymbol{p}$ are used to show how the position and momentum change over time, as follows

$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i} \tag{3.32}$$

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}, \tag{3.33}$$

where $i = 1, ..., d$.
Solving these differential equations allows us to predict an object's location and momentum at any point in time t.

Four important properties are associated with Hamiltonian dynamics. Namely, reversibility, conservation, volume preservation, and symplecticness. These properties were discussed in detail in Neal et al. (2011) and are listed in the following subsection.

3.3.3 Properties of Hamiltonian Dynamics

Reversibility

Hamiltonian dynamics is reversible. We can apply a mapping T_s from the current state at time t, $(\boldsymbol{q}(t), \boldsymbol{p}(t))$ to some state at time $t + s$, $(\boldsymbol{q}(t+s), \boldsymbol{p}(t+s))$. The mapping T_s is one-to-one, which means an inverse T_{-s} exists. This inverse mapping is obtained by negating the time derivatives seen in equations 3.32 and 3.33. When $K(\boldsymbol{p}) = K(-\boldsymbol{p})$ from the Hamiltonian of the form in 3.31, then the inverse mapping can also be obtained by negating $\boldsymbol{p}$, applying T_s, and then negating $\boldsymbol{p}$ again. This property is important for showing that the Metropolis updates using these dynamics leave the distribution we wish to sample from invariant.

Conservation

The Hamiltonian is invariant, which is seen in equations 3.32 and 3.33

$$\frac{dH}{dt} = \sum_{i=1}^{d} \left[\frac{dq_i}{dt} \frac{\partial H}{\partial q_i} + \frac{dp_i}{dt} \frac{\partial H}{\partial p_i} \right] = \sum_{i=1}^{d} \left[\frac{\partial H}{\partial p_i} \frac{\partial H}{\partial q_i} - \frac{\partial H}{\partial q_i} \frac{\partial H}{\partial p_i} \right] = 0. \quad (3.34)$$

This is important, as for Metropolis updates using a proposal found by Hamiltonian dynamics, the acceptance probability (when applying MCMC) is one if H is kept invariant. An important note is that we can only make H approximately invariant.

Volume Preservation

Hamiltonian dynamics preserve volume in $(\boldsymbol{q}, \boldsymbol{p})$ space, which is a result known as Liouville's Theorem (Gibbs, 1885). If we apply the mapping T_s to the points in some region R of $(\boldsymbol{q}, \boldsymbol{p})$ space, with volume V, the image of R under T_s will also have volume V. Volume preservation is significant for MCMC, as we do not need to account for any changes in volume in the acceptance probability for Metropolis updates.

Symplecticness

Volume preservation is a consequence of Hamiltonian dynamics being symplectic. Let $z = (\boldsymbol{q}, \boldsymbol{p})$, and

$$\frac{dz}{dt} = J \nabla H(z), \tag{3.35}$$

where

$$J = \begin{bmatrix} 0_{d \times d} & I_{d \times d} \\ -I_{d \times d} & 0_{d \times d} \end{bmatrix}, \tag{3.36}$$

and $\nabla H(z)$ is the gradient of H defined in 3.31. The symplecticness condition is that the Jacobian matrix, B_s, of the mapping T_s satisfies

$$B_s^T J^{-1} B_s = J^{-1}. \tag{3.37}$$

This would imply volume preservation, since $\det(B_s^T) \det(J^{-1}) \det B_s = \det(J^{-1})$ implies that $\det(B_s)^2$ is one.

Reversibility, volume preservation and symplecticness, are all preserved exactly when Hamiltonian dynamics is approximated via the leapfrog method. Approximation is a necessary step for computation of the dynamics, explored below.

3.3.4 Discretization of Time

In order to implement Hamiltonian dynamics computationally, we must discretize the continuous time variable, using a small step size denoted by ε. There are different ways to discretize time. Three have been explored by (Neal et al., 2011).

Euler's method

A well known method for discretizing time involves the use of the approximation

$$\frac{d\boldsymbol{q}}{dt} = \frac{\Delta \boldsymbol{q}}{\varepsilon}. \tag{3.38}$$

Equation 3.38 implies the following steps to update the position and momentum,

$$p_i(t + \varepsilon) = p_i(t) + \varepsilon \frac{dp_i}{dt}(t) = p_i(t) - \varepsilon \frac{\partial U}{\partial q_i}(q(t)), \tag{3.39}$$

$$q_i(t + \varepsilon) = q_i(t) + \varepsilon \frac{dq_i}{dt}(t) = q_i(t) - \varepsilon \frac{p_i(t)}{m_i}, \tag{3.40}$$

where m_i is the variance for the momentum variable p_i as specified by the kinetic energy equation. If we start at $t = 0$ with given values for $q_i(0)$ and $p_i(0)$, we can iterate the steps above to get a trajectory of position and momentum values at times $\varepsilon, 2\varepsilon, 3\varepsilon, ...$, and hence find approximate values for $q(t^*)$ and $p(t^*)$ after t^*/ε steps.

Euler's method of discretization does not preserve volume, and as a result, trajectories of position and momentum tend to diverge with each step.

Modified Euler's method

A modified version of Euler's method exists, which allows for the preservation of volume. This gives much better results,

$$p_i(t + \varepsilon) = p_i(t) - \varepsilon \frac{\partial U}{\partial q_i}(q_i(t)), \tag{3.41}$$

$$q_i(t + \varepsilon) = q_i(t) + \varepsilon \frac{p_i(t + \varepsilon)}{m_i}. \tag{3.42}$$

We simply use the new value for the momentum variables, p_i, when computing the new value for the position variable q_i.

Leapfrog method

In Neal et al. (2011), it is shown that the leapfrog method provides the best results in the context of applying MCMC. The leapfrog method uses the following steps:

$$p_i\left(t + \frac{\varepsilon}{2}\right) = p_i(t) - \left(\frac{\varepsilon}{2}\right)\frac{\partial U}{\partial q_i}\big(q(t)\big), \tag{3.43}$$

$$q_i(t + \varepsilon) = q_i(t) + \varepsilon \frac{p_i(t + \frac{\varepsilon}{2})}{m_i}, \tag{3.44}$$

$$p_i(t + \varepsilon) = p_i\left(t + \frac{\varepsilon}{2}\right) - \left(\frac{\varepsilon}{2}\right)\frac{\partial U}{\partial q_i}\big(q_i(t + \varepsilon)\big). \tag{3.45}$$

The algorithm starts with a half step for the momentum variable seen in expression 3.43. A full step is then completed for the position variables, using the new values of the momentum variables, seen in expression 3.44. Finally, another half step is taken for the momentum variables, using the new values for the position variables (expression 3.45).

The leapfrog method preserves volume exactly, since each of the expressions in 3.43 to 3.45 are only transformations. In addition, due to its symmetry, the method is also reversible, by simply negating p, applying the same number of steps, then negating p again.

The leapfrog method is used in the probabilistic software, Stan (Gelman et al., 2015), which is used in this thesis.

3.3.5   Hamiltonian Monte Carlo Sampling

In order to relate Hamiltonian dynamics to a probability distribution, we make use of the famous canonical distribution which is derived from Maxwell-Boltzmann statistics (Tasaki, 1998).
The canonical distribution takes the following form

$$f(\theta) = \frac{1}{Z} \exp^{-E(\theta)}, \tag{3.46}$$

where f is a probability distribution on parameter vector θ, Z is a normalizing constant called the partition function, and $E(\theta)$ is an energy function that depends on parameter vector θ. In our case, we let the energy function be the Hamiltonian described by equation 3.31,

$$E(\theta) = H(\boldsymbol{q}, \boldsymbol{p}). \tag{3.47}$$

By combining, equations 3.46 and 3.47, we arrive at the following result,

$$f(\boldsymbol{q}, \boldsymbol{p}) \propto \exp^{-[U(\boldsymbol{q})+K(\boldsymbol{p})]} = \exp^{-U(\boldsymbol{q})} \exp^{-K(\boldsymbol{p})} \propto f(\boldsymbol{q})f(\boldsymbol{p}). \tag{3.48}$$

The expression in 3.48 shows us that we can write the joint probability function for position and momentum as two independent canonical distributions for position and momentum separately. Here, we will let the position vector $\boldsymbol{q}$ represent our variables of interest. The momentum vector $\boldsymbol{p}$ will therefore represent any auxiliary information that is needed to make Hamiltonian dynamics work. Because of this, we can choose any distribution to sample momentum from. Commonly, we use a zero-mean normal process with unit variance given by,

$$f(\boldsymbol{p}) \propto \frac{\boldsymbol{p}^T \boldsymbol{p}}{2}. \tag{3.49}$$

This makes the partial derivatives easy to compute in order to update the momentum vector. We now need to find a potential energy function, $U(\boldsymbol{q})$,

such that:

$$U(\boldsymbol{q}) = -\log f(\boldsymbol{q}), \tag{3.50}$$

where $f(\boldsymbol{q})$ will be the posterior distribution that we wish to sample from. If we are able to compute the partial derivative in 3.51 below,

$$-\frac{\partial \log(f(\boldsymbol{q}))}{\partial q_i}, \tag{3.51}$$

then we can simulate Hamiltonian dynamics to be used in our MCMC sampler.

Algorithm 1 below for implementing the Hamiltonian Monte Carlo (HMC) sampler has two main stages – the leapfrog stage, and the acceptance-rejection stage.

Algorithm 1 Hamiltonian Monte Carlo Algorithm

1: Generate initial position vector, $\boldsymbol{q}^0$ from some prior $\pi(0)$, and initial momentum vector, $\boldsymbol{p}^0$, from a pre-chosen canonical distribution $f(\boldsymbol{p})$.

2: Use equations 3.43, 3.44, and 3.45 to generate a momentum and position vector, $(\boldsymbol{q}^*, \boldsymbol{p}^*)$ after L leapfrog iterations.

3: Compute the acceptance probability,

$$\alpha = \min\left\{1, e^{-H(\boldsymbol{q}^*,\boldsymbol{p}^*)+H(\boldsymbol{q},\boldsymbol{p})}\right\} = \min\left\{1, e^{-U(\boldsymbol{q}^*)+U(\boldsymbol{q})-K(\boldsymbol{p}^*)+K(\boldsymbol{p})}\right\} \tag{3.52}$$

4: Generate $u \sim U(0,1)$. If $u \leq \alpha$, then the proposal state $(\boldsymbol{q}^*, \boldsymbol{p}^*)$ is accepted and set as the next state in the Markov chain. Otherwise, the current state is re-used as the next state.

At this point, we have seen the ground work for HMC. However, there are some problems related to hand-tuning the parameters ε and L. In order to implement HMC to its full potential, we must carefully choose the correct number of steps, L, and the step size, ε. If we choose the step size to be too large, then sampling is inaccurate due to low acceptance rates. If the

step size is too small, then precious computation time is wasted. If the number of steps is too large, then HMC will generate trajectories that loop back on itself. If L is too small, then successive samples will be close to one another. This will cause random walk behavior, and slow mixing. Finally, if L is chosen such that parameters jump from side-to-side, then the chain may not even be ergodic. Most commonly, the chain is slow to move between regions of low and high density. Carefully hand-tuning the HMC may take time if the distribution we wish to sample from is complex, as repeat runs of the algorithm are needed. A much more efficient way to use HMC is to implement the extension, "No U-turn Sampler" (NUTS).

3.3.6 No U-turn Sampler

The basis of the NUTS algorithm is to use a criterion that indicates that the Hamiltonian dynamics have been simulated "long enough" (Hoffman & Gelman, 2014). Hoffman & Gelman (2014) use a convenient criterion based on the dot product between the current momentum $p(t)$ and the vector from the initial position to the current position, $q(t) - q(0)$. This is just the derivative with respect to time of half the squared distance between the initial and current position,

$$
\frac{d}{dt} \frac{\big(q(t) - q(0)\big) \cdot \big(q(t) - q(0)\big)}{2} = \big(q(t) - q(0)\big) \frac{d}{dt} \big(q(t) - q(0)\big)
$$
$$
= \big(q(t) - q(0)\big) p(t) \tag{3.53}
$$

This suggests an algorithm in which leapfrog steps are run until the quantity in expression 3.53 becomes less than 0, indicating that the proposal position $q(t)$ has started to move back toward $q(0)$. A disadvantage to this algorithm is that it does not guarantee time reversibility, and hence is not guaranteed to converge to the correct distribution. However, the NUTS algorithm overcomes this issue by the means of a recursive algorithm that

uses a slice variable ν, which has conditional distribution

$$f(\nu|\boldsymbol{q}, \boldsymbol{p}) = \text{Uniform}(0, \exp(l(\boldsymbol{q}) - \frac{1}{2}\boldsymbol{p}.\boldsymbol{p})), \qquad (3.54)$$

where $l(\boldsymbol{q})$ is the likelihood for the position parameter, and $\boldsymbol{p}.\boldsymbol{p}$ denotes the inner product. Although not necessary, the slice variable simplifies the derivation and implementation of the NUTS algorithm. The algorithm can be found in Hoffman & Gelman (2014).

3.3.7   Cholesky Decomposition

In order to simulate a posterior distribution built from Gaussian Markov random field priors, we need a way to sample from a GMRF. Rue & Held (2005) provide simple algorithms for such computations, and these have been built into the software Stan (Gelman et al., 2015). One such method involves the useful result of Cholesky decomposition.

The result of a Cholesky decomposition is incredibly useful when computations on large sparse matrices are required. We can decompose a sparse matrix $\boldsymbol{V}$ into a lower triangular matrix and its transpose,

$$\boldsymbol{V} = \boldsymbol{L}\boldsymbol{L}^T, \qquad (3.55)$$

where $\boldsymbol{L}$ is a lower triangular sparse matrix.

The lower triangular matrix retains the band structure from the original matrix, which allows computations to be completed on $\boldsymbol{L}$. This has the benefit of greatly improving computational efficiency. If we choose $\boldsymbol{V}$ to be the precision matrix $\boldsymbol{Q}$ of GMRF that has been permuted such that it is a sparse band matrix, then computational cost is $\mathcal{O}(n)$, $\mathcal{O}(n^{\frac{3}{2}})$, and $\mathcal{O}(n^2)$ for GMRFs in time, space and space-time respectively.

Algorithm 2 below (Rue & Held, 2005) is used to sample from a Guassian Markov random field with a mean $\boldsymbol{\mu}$ and precision matrix $\boldsymbol{Q}$.

We review methods to diagnose convergence issues in the next section.

Algorithm 2 Sampling GMRF $x \sim N(\mu, Q^{-1})$

1: Compute the Cholesky factorization, $Q = LL^T$
2: Sample $z \sim N(0, I)$
3: Solve $L^T v = z$
4: Compute $x = \mu + v$
5: Return x

3.4 Diagnostics for Convergence Assessment

When simulating from a posterior distribution, using iterative approaches such as MCMC and HMC, it is necessary to perform checks to ensure that the resulting distribution of posterior draws are representative of the target posterior distribution. Many methods exist that help monitor the convergence and mixing of a posterior sample. In order to make inferences, we use the collection of simulated draws after discarding warm-up samples from the posterior distribution to summarize the density. The collection of draws come from multiple chains, each initiated with over dispersed starting values.

The number of iterations is an important factor when assessing the quality of the posterior sample. If an insufficient amount of iterations are used, then convergence to the target distribution may not be satisfied. This would cause the posterior draws to be unrepresentative of the target posterior density. This is still an issue even if convergence is achieved, as the earlier iterations would represent the starting approximation, rather than the target density (Gelman et al., 2014a). A solution is to discard some of the earlier iterations.

Another issue occurs due to the correlation of successive draws. This is not a large problem, as the order of the draws does not matter when it comes to inference, however, this can lead to slow convergence. If successive draws are highly correlated, then there is slow movement about the support of the target distribution.

We monitor these problems visually with the use of trace plots, density plots, autocorrelation plots and Gelman-Rubin plots (which use the potential scale reduction factor) to assess convergence.

3.4.1   Trace Plots

Trace plots display the parameter value at each iteration of sampling and are useful for visually assessing the convergence of the simulated draws to a posterior distribution. They allow us to check whether any values are rejected repeatedly causing the chain to become stuck on a single value (poor mixing). When poor mixing occurs, a particular value may be over-represented in the posterior sample.

In addition to monitoring chain mixing, trace plots allow us to check if any patterns are present. Clear patterns in a trace plot indicate that the algorithm may not have converged. In addition, if we use multiple chains and observe that they traverse different parts of the parameter space, then this is also indicative of non-convergence. To rectify these problems, we can increase the number of iterations.

3.4.2   Posterior Density Plots

Density plots of the posterior parameters are also useful for visualizing the quality of the posterior draws. The shapes of the densities are dependent on the distributions involved in the construction of the posteriors. However, if large and erratic peaks are observed in the density plots, then this indicates a lack of convergence to a single target distribution.

3.4.3   Autocorrelation Plots

When consecutive values in a Markov chain are highly correlated, then traversing the sample space will be slow, leading to issues such as poor mixing. This is because the proposal parameters are more likely to be

close to the current state. A plot of the autocorrelation for each parameter chain can be used to assess whether the correlation between successive draws will cause issues. If high correlation is detected, then thinning the sample (taking every k^{th} draw) will make the values less dependent (Link & Eaton, 2012).

3.4.4 Potential Scale Reduction Factor

The potential scale reduction factor, $\hat{R}$, is used to monitor convergence of a posterior distribution for a parameter, ψ, to a stationary distribution. It is an estimate of the factor, by which the scale of the current distribution for a parameter might be reduced if simulations were continued in the limit $n \to \infty$.

To compute $\hat{R}$ for a posterior simulation of ψ, we must first estimate the marginal posterior variance, $\text{Var}(\psi|y)$. This can be done by a weighted average of within-chain variance, and between chain variance.

Let ψ_{ij} represent the posterior draw for iteration $i = 1, ..., n$ from chain $j = 1, ..., m$. Then the between chain variance is defined as

$$B = \frac{n}{m-1} \sum_{j=1}^{m} (\bar{\psi}_{.j} - \bar{\psi}_{..})^2, \tag{3.56}$$

where

$$\bar{\psi}_{.j} = \frac{1}{n} \sum_{i=1}^{n} \psi_{ij}, \quad \text{and} \quad \bar{\psi}_{..} = \frac{1}{m} \sum_{j=1}^{m} \bar{\psi}_{.j}, \tag{3.57}$$

and the within chain variance is defined as

$$W = \frac{1}{m} \sum_{j=1}^{m} s_j^2, \tag{3.58}$$

where

$$s_j^2 = \frac{1}{n-1} \sum (\psi_{ij} - \bar{\psi}_{.j})^2. \tag{3.59}$$

The marginal posterior variance is given by

$$\hat{\text{Var}}(\psi|y) = \frac{n-1}{n}W + \frac{1}{n}B. \tag{3.60}$$

Thus, the potential scale reduction factor is given by

$$\hat{R} = \sqrt{\frac{\hat{\text{Var}}(\psi|y)}{W}}. \tag{3.61}$$

This expression converges toward 1 as n tends to ∞. As a result, if a parameter chain has a potential scale reduction factor of 1 (or close to 1), then it can be assumed that the chain has converged to the target distribution. The potential scale reduction factor can be calculated and cumulatively plotted at each iteration, for each chain. Such plots are referred to as Gelman-Rubin plots (Gelman et al., 2014a) and allow for visually assessing the convergence of posterior parameter distributions.

We now provide an example that compares HMC to MCMC, using a simple target distribution.

3.4.5   Example of Hybrid Monte Carlo Sampler

In this example, we compare the simulation of a Normal(0,1) distribution using Metropolis random walk, and Hamiltonian Monte Carlo (without the NUTS algorithm).

We wish to sample y from $N(0,1)$. In the MCMC random-walk, we need to use a candidate distribution which allows for good mixing. We choose a uniform distribution, $x \sim \text{Uniform}(-\delta, \delta)$. It was found that the value of δ which gave the optimum random-walk was 3. The total number of iterations was 1000.

For the HMC, we used a momentum function which was equivalent to a $N(0,1)$ distribution. We defined the stepsize, ε, to be 0.1, and L to be 20. The total number of iterations was 1000.

A trace plot and histogram for the chains produced by each method can be seen in Figure 3.3. Clearly, there appears to be much more scatter and exploration of the support using the HMC sampler. As a result, we obtain a histogram of observations that look much more Normal, compared with the MCMC random-walk. Figure 3.4 shows plots of the cumulative means for each sampling method. The convergence to the true mean of zero is much quicker using the HMC method, compared to the MCMC random-walk. In addition, once convergence is achieved, there is less deviation from the true mean when using the HMC sampler which is ideal.

3.5   Model Comparison & Assessment

Following on after constructing the joint probability model and computing the posterior distributions for each parameter it is necessary to assess the fit of the model to the observed data. In the Section 1.2.6, we reviewed common information criteria used to assess Bayesian models in the literature. In this section, we focus on constructing the Watanabe-Akaike information criterion (Watanabe, 2010), a fairly new model assessment criterion for Bayesian models.

Watanabe-Akaike Information Criterion

The Watanabe-Akaike information criterion (also known as the widely applicable information criterion, WAIC) is more Bayesian than DIC (described in Section 1.2.6). It was conceived by Watanabe (2010), and uses the computed log pointwise posterior predictive distribution 1.17, as opposed to the log posterior density evaluated at a point estimate.

$$\text{WAIC} = -2\widehat{\text{elppd}}_{\text{WAIC}} = -2\text{lppd} + 2p_{\text{WAIC}}. \qquad (3.62)$$

Here, the effective number of parameters that is used to adjust for over
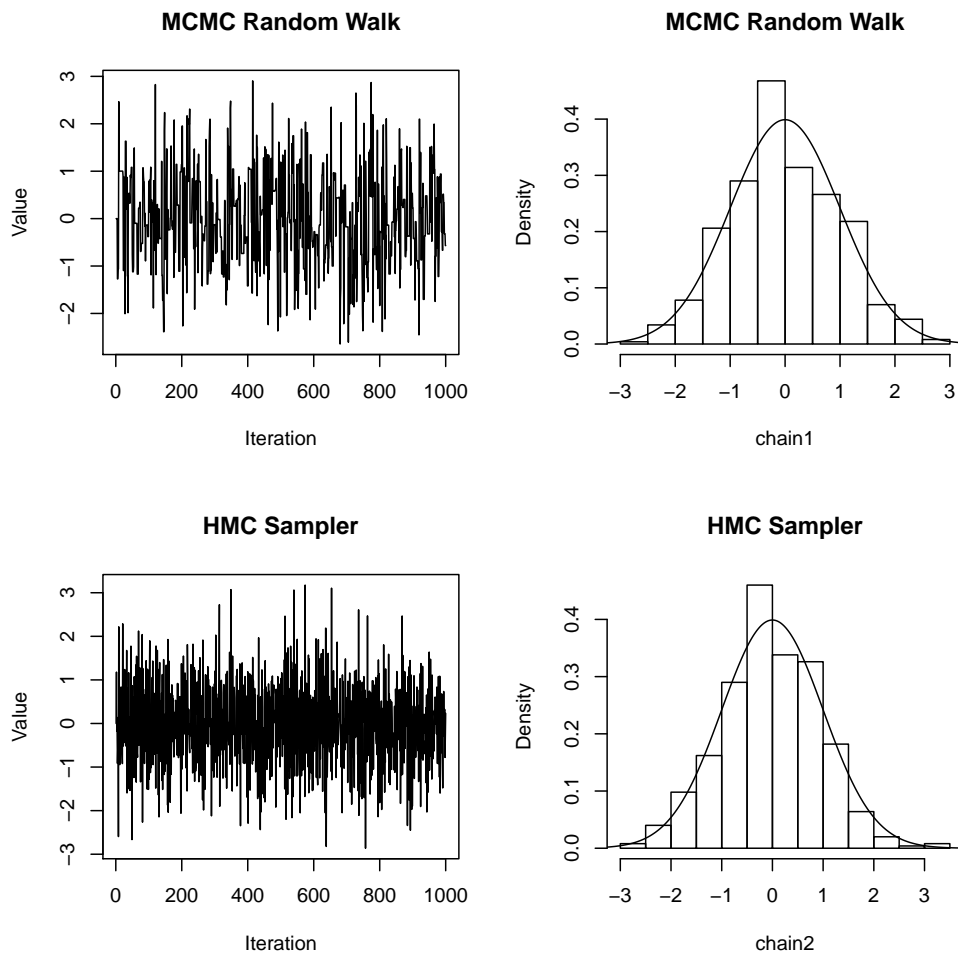
Figure 3.3: Trace plots and histograms for the posterior mean of a N(0,1) variable using both MCMC random walk and HMC.
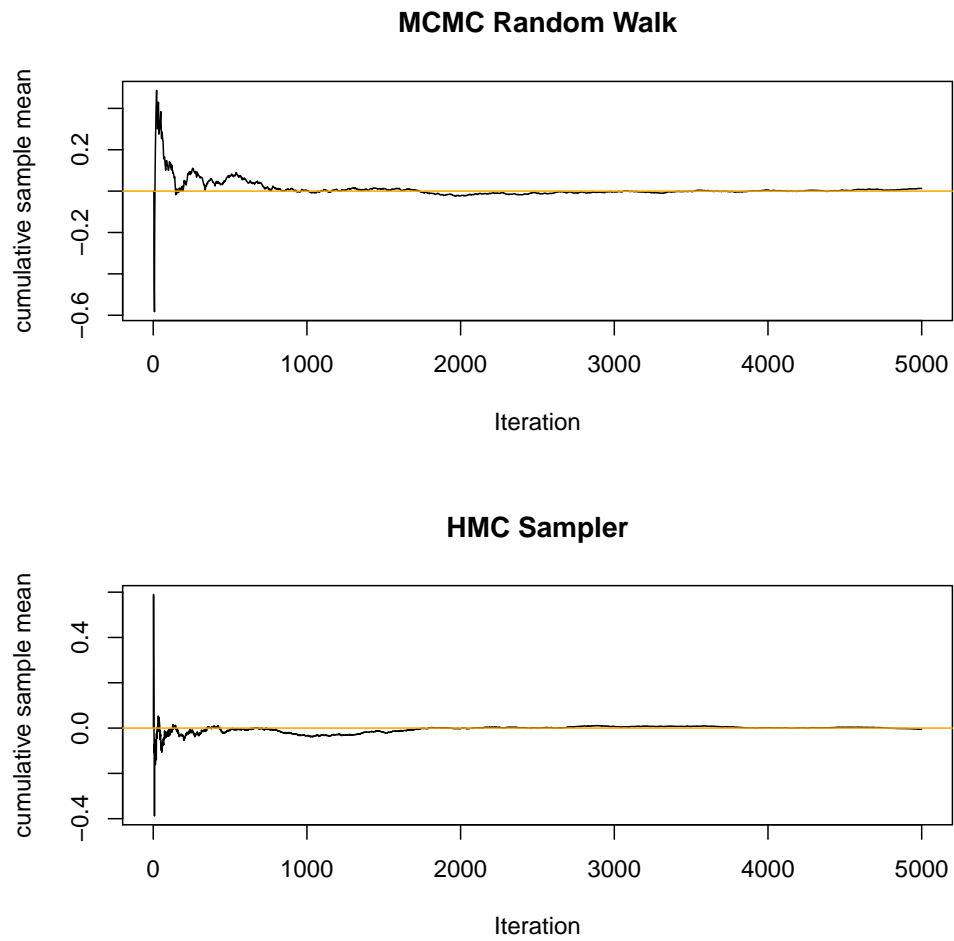
MCMC Random Walk

HMC Sampler

Figure 3.4: Mean plot for the posterior mean of a N(0,1) variable using both MCMC random walk and HMC.

fitting is similar to Equation 1.21 for DIC. It is given by

$$p_{\text{WAIC,1}} = 2 \sum_{i=1}^{n} \left(\log(\text{E}_{\text{post}} p(y_i | \theta)) - \text{E}_{\text{post}}(\log p(y_i | \theta)) \right). \tag{3.63}$$

WAIC and the information criteria described in Section 1.2.6 all have their downfalls, and so there is no general selection criterion we can use in this thesis. AIC does not work in settings where strong prior information exists (Gelman et al., 2014a), such as exhibited in the models to come (see Section 4.1). DIC is a more likely candidate, however, it fails when the posterior distribution is not well summarized by its mean (Gelman et al., 2014a). This leaves WAIC, which we can use for some (but not all) of the models to be fitted. WAIC cannot be used where data is structured, such as when observations are point referenced. This is because WAIC depends on partitions of the data, which would subsequently destroy the inherent dependence between point referenced observations.

As a result, we compute WAIC where it is appropriate, as this is better than having no formal comparison at all.

3.5.1   Residuals

It is common practice to compute the residuals between the observed data, and the values predicted by a model. By plotting the residuals on an axis that represents an increase in some level of a covariate (usually just the predicted values themselves), we can assess whether the unexplained variation is due to natural changes in the response variable, or whether there is some underlying latent variable that is not considered by the model. We can apply similar ideas in a Bayesian sense, by computing posterior residuals for each observed value. In order to do so, we first compute the posterior predicted values.

Let $\boldsymbol{y}$ be a vector of observations of size n that depends on some parameter vector $\boldsymbol{\theta}$, and observed covariates $\boldsymbol{X}$. Let the likelihood function of

the observations be Normal with $\boldsymbol{y}|\boldsymbol{\theta} \sim \mathrm{N}(\boldsymbol{X}\boldsymbol{\theta}, \sigma^2)$. We sample from the posterior distribution of $\boldsymbol{\theta}$, $p(\boldsymbol{\theta}|\boldsymbol{y})$, which gives posterior draws, $\boldsymbol{\theta}^{(S)}$. The posterior predicted values for iteration s are given by

$$\hat{\boldsymbol{y}}^{(s)} = \boldsymbol{X}\boldsymbol{\theta}^{(s)}. \tag{3.64}$$

We end up with a distribution of predicted values for each observation. From here, we calculate the posterior residuals by taking the difference between the observed values and the expected values under the model.

$$\boldsymbol{r}^{(s)} = \boldsymbol{y} - \hat{\boldsymbol{y}}^{(s)} \tag{3.65}$$

Now that we have a distribution of posterior residuals, we plot the posterior mean against an index which is ordered by a covariate (in our models, ordered by trawl year). We can visually inspect the plot to quickly ascertain whether the model excludes important features in the data.

Chapter 4

Models

In this chapter, we introduce five models used to analyse hoki catch per unit effort data. Descriptions of the models and their specifications are given, followed by a section on implementation and convergence diagnostics. The main results are given in Chapter 5.

We would like to fit models to determine whether accounting for spatial and temporal variation and autocorrelation is important when constructing CPUE statistics. We attempt this by modelling hoki CPUE using both trawl year and stratum as factors. The survey data are collected in such a way that other factors, such as vessel and gear differences, are controlled for. Therefore no other factors need to be included in the models.

4.1 Overview of Models

Five models were constructed that included variables that incorporated temporal, spatial and both temporal and spatial patterns within the hoki CPUE data. Table 4.1 provides an overview of each model. Model 1 attempts to account for temporal autocorrelation in hoki CPUE through a GMRF prior with a first order autoregressive (AR(1)) precision matrix structure (Equation 3.12). Model 2 attempts to account for spatial auto-correlation in hoki CPUE through a GMRF prior with a conditionally au-

Table 4.1: Description of each of the models fitted to the data. Labels T, AS, P, and BP in the Description column stand for "temporal", "areal spatial", "point referenced", and "block point referenced" respectively

Model	Description	Data Likelihood
1	T	$\log \boldsymbol{y}|\mu, \boldsymbol{\tau}, \sigma, \rho, \phi \sim \mathrm{MVN}(\mathbf{1}\mu + \boldsymbol{X}\boldsymbol{\tau}, \sigma^2\boldsymbol{I})$
2	AS	$\log \boldsymbol{y}|\mu, \boldsymbol{\omega}, \sigma, p, \eta \sim \mathrm{MVN}(\mathbf{1}\mu + \boldsymbol{W}\boldsymbol{\omega}, \sigma^2\boldsymbol{I})$
3	T, AS	$\log \boldsymbol{y}|\mu, \boldsymbol{\tau}, \boldsymbol{\omega}, \sigma, \rho, \phi, p, \eta \sim \mathrm{MVN}(\mathbf{1}\mu + \boldsymbol{X}\boldsymbol{\tau} + \boldsymbol{W}\boldsymbol{\omega}, \sigma^2\boldsymbol{I})$
4	T, AS, P	$\log \boldsymbol{y}|\mu, \boldsymbol{\tau}, \boldsymbol{\omega}, \rho, \phi, \eta, p, \delta, \kappa \sim \mathrm{MVN}(\mathbf{1}\mu + \boldsymbol{X}\boldsymbol{\tau} + \boldsymbol{W}\boldsymbol{\omega}, \boldsymbol{Q}^{-1})$
5	T, AS, BP	$\log \boldsymbol{y}|\mu, \boldsymbol{\tau}, \boldsymbol{\omega}, \rho, \phi, \eta, p, \boldsymbol{\delta}, \boldsymbol{\kappa} \sim \mathrm{MVN}(\mathbf{1}\mu + \boldsymbol{X}\boldsymbol{\tau} + \boldsymbol{W}\boldsymbol{\omega}, \boldsymbol{Q}^{-1})$

toregressive (CAR) precision matrix structure (Equation 3.22). Model 3 attempts to account for both spatial and temporal autocorrelation by including GMRF priors for both effects. Models 4 and 5 explore different ways of accounting for a spatial effect. Model 4 accounts for spatial dependence in a point referenced way. It contains a temporal effect, a stratum effect, and models CPUE dependence using a point referenced CAR precision matrix (Equation 3.23). Model 5 accounts for spatial dependence in a blocked point referenced way. It contains a temporal effect, a stratum effect, and models CPUE dependence using a blocked point reference CAR precision matrix (Equation 3.29).

Table 4.2 provides a list of descriptions of the terms used in the five models shown in Table 4.1.

4.1.1   Data Likelihood

In Section 2.5.1, we found evidence that suggests hoki CPUE follows a log-normal distribution. This was illustrated in Figure 2.4. As such, the models in Table 4.1 all assume log-normal data likelihoods.

Table 4.2: Description of the notation used in each model

Term	Size	Description
$\log \boldsymbol{y}$	1255×1	Natural log of hoki CPUE in kg km^{-2}
$\mathbf{1}\mu$	1255×1	Intercept
$\boldsymbol{\tau}$	12×1	Temporal effects
$\boldsymbol{\omega}$	23×1	Areal spatial effects
σ	1×1	Observation standard deviation
$\boldsymbol{R}$	12×12	Autoregressive precision matrix for temporal effect
$\boldsymbol{S}$	23×23	CAR precision matrix for areal spatial effect
$\boldsymbol{Q}$	1255×1255	CAR Precision matrix for CPUE
$\boldsymbol{Q_j}$	$n_j \times n_j$	CAR Precision matrix for CPUE in stratum j
ρ	1×1	Measure of temporal correlation
ϕ	1×1	Precision parameter for the temporal effect
p	1×1	Measure of the strength of areal spatial association
η	1×1	Precision parameter for the areal spatial effect
κ	1×1	Measure of the strength of CPUE correlation
δ	1×1	Precision parameter for CPUE
$\boldsymbol{\kappa}$	23×1	Correlation strengths of CPUE within each stratum
$\boldsymbol{\delta}$	23×1	Precision parameters for CPUE within each stratum
$\boldsymbol{X}$	1255×12	Temporal design matrix
$\boldsymbol{W}$	1255×23	Areal spatial design matrix
$\boldsymbol{D}$	23×23	Stratum neighbourhood matrix
$\boldsymbol{A}$	23×23	Stratum adjacency matrix
$\boldsymbol{G}$	1255×1255	CPUE neighbourhood matrix
$\boldsymbol{C}$	1255×1255	CPUE adjacency matrix
$\boldsymbol{G_j}$	$n_j \times n_j$	Neighbourhood matrix for CPUE in stratum j
$\boldsymbol{C_j}$	$n_j \times n_j$	Adjacency matrix for CPUE in stratum j
n_j	1×1	Number of CPUE observations in stratum j

4.1.2   Temporal Effect

Models 1, 3, 4, and 5 all include trawl year as an explanatory variable to account for temporal autocorrelation in hoki CPUE. Trawl year is a factor that has 12 levels ranging from 1991 to 1993, and then again from 2000 to 2008. The year that had the fewest catch weight observations was 2003 with 77 records, and the year that had the most was 1992 with 156 records.

The vector of temporal effects, $\boldsymbol{\tau} = (\tau_2, ..., \tau_{12})^T$, is modelled as a zero-mean Gaussian Markov random field that enters the hierarchical model as a prior (Rue & Martino, 2007). The precision matrix used in the GMRF allows for specification of the temporal structure within the data, using

$$\boldsymbol{\tau}|\phi, \rho \sim \text{MVN}(\mathbf{0}, \boldsymbol{R}^{-1}). \tag{4.1}$$

Here, $\boldsymbol{R}$ is the precision matrix that has been assumed to have an AR(1) structure (see Section 3.12).

In nature, temporal dependence tends to decrease as the time interval between successive observations widens. It was observed in the autocorrelation plot (Figure 2.10) that the dependence within CPUE changes as the lag between years increases. We use an AR(1) structure to model first order autocorrelation within hoki CPUE. There is also evidence to suggest fourth order autocorrelation exists, however, we do not model it in this thesis.

Not all years will be correlated with one another. There is a significant gap in data between the years 1993 and 2000, and this needs to be represented in the precision matrix. As such, we prevent any correlations between the earlier years (1991-1993) and the later years (2000-2008). Hence, $\boldsymbol{R}$ is

assumed to have the form below:

$$R = \phi \frac{1}{1 - \rho^2} \begin{pmatrix} 1 & -\rho & 0 & 0 & \ldots & 0 \\ -\rho & 1 + \rho^2 & -\rho & 0 & \ldots & 0 \\ 0 & -\rho & 1 + \rho^2 & 0 & \ldots & 0 \\ 0 & 0 & 0 & 1 + \rho^2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & 0 & -\rho & 1 + \rho^2 & -\rho \\ 0 & \ldots & 0 & 0 & -\rho & 1 \end{pmatrix}, \qquad (4.2)$$

where $\boldsymbol{R}$, ρ, and ϕ are defined in Table 4.2. The diagonal of the precision matrix gives the conditional precisions for each year effect, given all the rest, while the off-diagonals are the conditional correlations (see Equation 3.13).

4.1.3 Areal spatial Effect

Models 2, 3, 4, and 5 all include stratum as an explanatory variable to account for the spatial effect in hoki CPUE. To include an areal spatial effect, the stratum variable is used. Stratum has 24 levels, corresponding to stratification of the sub-Antarctic region (see Figure 2.1). Two catch weight observations were not assigned to a stratum, and were hence removed. Stratum 17 was excluded because there was only one observation for that stratum. There is not enough information to estimate an effect for that stratum. As a result, a stratum effect with 23 levels is used. The stratum with the smallest number observations was stratum 26 with 4 observations. The stratum with the largest number of observations was stratum 9 with 116.

The dataset was organized by chronological year. The vector of spatial effects, $\boldsymbol{\omega} = (\omega_2, ..., \omega_{23})^T$, is modelled by a zero-mean Gaussian Markov random field that enters the hierarchical model as a prior. The precision matrix used in the GMRF allows for specification of the spatial autocorre-

lation structure within the data.

$$\boldsymbol{\omega}|\eta, p \sim \text{MVN}(\boldsymbol{0}, \boldsymbol{S}^{-1}) \tag{4.3}$$

Here, $\boldsymbol{S}$ is the precision matrix. The structure of $\boldsymbol{S}$ is representative of our assumption that neighbouring strata (strata which share a boundary) are likely to be correlated with one another. The precision matrix is assumed to have a CAR structure (see Section 3.22). It is given by

$$\boldsymbol{S} = \eta(\boldsymbol{D} - p\boldsymbol{A}), \tag{4.4}$$

where $\boldsymbol{S}$, $\boldsymbol{D}$, $\boldsymbol{A}$, p, and η are described in Table 4.2.

The adjacency matrix $\boldsymbol{A}$ is defined below:

$$\boldsymbol{A} = \begin{pmatrix} 0 & a_{1,2} & \dots & a_{1,23} \\ a_{2,1} & 0 & \dots & a_{2,23} \\ \dots & \dots & \dots & \dots \\ a_{23,1} & a_{23,2} & \dots & 0 \end{pmatrix}, \tag{4.5}$$

where $a_{i,j} = 1$ if stratum i and stratum j are neighbours, and $a_{i,j} = 0$ otherwise.

The neighbourhood matrix $\boldsymbol{D}$ is defined below:

$$\boldsymbol{D} = \begin{pmatrix} d_{1,1} & 0 & \dots & 0 \\ 0 & d_{2,2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & d_{23,23} \end{pmatrix}, \tag{4.6}$$

where $d_{i,i} = \sum_{j=1}^{23} a_{i,j}$ is the number of neighbours surrounding stratum i. The diagonal of the precision matrix gives the conditional precisions for each stratum effect, given all the rest, while the off-diagonals are the conditional correlations (see Equations 3.15 and 3.16).

4.1.4 Spatial Dependence with Point Referenced CPUE

Model 4 makes use of the point referenced CPUE records of hoki. In this model, pairs of CPUE observations are assumed to be correlated such that the correlation depends on the distance between them. The point referenced observations ignore strata. To specify this in the model, a CAR precision matrix Q enters the model through the data likelihood (see Section 3.23).

The precision matrix Q was assumed to have the following form.

$$Q = \delta(G - \kappa C), \tag{4.7}$$

where G, C, κ, and δ are defined in Table 4.2.

The adjacency matrix C takes the same form as the adjacency matrix in Equation 3.23. Here,

$$C = (c_{ij})_{1255 \times 1255}, \tag{4.8}$$

where

$$c_{ij} = \begin{cases} 1, & \text{if } d_{ij} \leq d_l \\ f(d_{ij}) & \text{if } d_l < d_{ij} \leq d_u \\ 0 & \text{if } d_{ij} > d_u \end{cases}, \tag{4.9}$$

and

$$c_{ii} = 0 \text{ for all } i, \tag{4.10}$$

and d_{ij} is the distance between observations i and j, d_l is a lower distance threshold, and d_u is an upper distance threshold. The distance function in Equation 4.9, $f(d_{ij})$, is given by

$$f(d_{ij}) = \frac{\frac{1}{d_{ij}} - \frac{1}{d_u}}{\frac{1}{d_l} - \frac{1}{d_u}}. \tag{4.11}$$

We formulated Equation 4.11 by adapting the function proposed by White & Ghosh (2009). This was discussed in Section 3.23.

We chose d_l and d_u, such that 90 % of the distances between CPUE observations lie between d_l and d_u. In this case, $d_l =$ and $d_u =$.

The neighbourhood matrix G takes the same form as the neighbourhood matrix in Equation 3.23, in that G is a 1255×1255 matrix with diagonal entries equal to the sum of the rows of C and zero elsewhere (White & Ghosh, 2009).

4.1.5 Spatial Dependence with Blocked Point Referenced CPUE

Model 5 also makes use of the point referenced CPUE records of hoki. In this model, pairs of CPUE observations are assumed to be correlated with correlation dependent on distance, but only within strata. This allows for a different correlation structure for each stratum, and assumes observations in different strata are independent. To specify this in the model, a blocked CAR precision matrix Q enters the model through the data likelihood.

The precision matrix Q was assumed to have the following form:

$$Q = \begin{pmatrix} Q_1 & 0 & ... & 0 \\ 0 & Q_2 & ... & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & ... & Q_{23} \end{pmatrix}, \tag{4.12}$$

where

$$Q_j = \delta_j(G_j - \kappa_j C_j), \tag{4.13}$$

where the matrices G_j and C_j have the same structures as the matrices G and C in Section 4.1.4 above, (for all stratum $j = 1, ..., 23$). The parameter vectors $\delta = (\delta_1, ..., \delta_{23})^T$ and $\kappa = (\kappa_1, ..., \kappa_{23})^T$ describe the precision of

CPUE within each stratum, and strength of dependence of CPUE within each stratum respectively.

4.1.6 Prior Parameter Distributions

Models 1, 3, 4, and 5 contain a temporal effect. As described in section 4.1.2, the vector of temporal effects $\boldsymbol{\tau}$ is a GMRF with a mean of zero and an autoregressive precision matrix $\boldsymbol{R}$. The parameters ρ and ϕ that $\boldsymbol{R}$ depends on are assigned prior distributions.

The parameter ϕ takes positive values. The Uniform distribution is used with support $[0, +\infty)$. The parameter ρ takes values between -1 and 1, hence a Uniform distribution on support $[-1, 1]$ is used, and we have

$$\phi \sim \text{Uniform}(0, +\infty),$$
$$\rho \sim \text{Uniform}(-1, 1).$$

Models 2, 3, 4, and 5 contain an areal spatial effect. As described in Section 4.1.3, the vector of spatial effects $\boldsymbol{\omega}$ is a GMRF with a mean of zero and a conditionally autoregressive precision matrix, $\boldsymbol{S}$. The parameters p and η that $\boldsymbol{S}$ depends on are assigned prior distributions.

The parameter η takes positive values. In Models 2 and 3, a Uniform distribution is used with support $[0, +\infty)$. However, in Models 4 and 5, it was found that the posterior distribution for η would not converge to a sensible value. As a result, Models 4 and 5 assign a Normal(10,10) prior distribution to η and this resulted in improved convergence. The parameter p takes values between 0 and 1 to ensure that the precision matrix $\boldsymbol{S}$ is positive definite (Rue & Held, 2005). Hence a Uniform distribution on support [0,1] is used as a prior:

$$\eta \sim \text{Uniform}(0, +\infty) \quad \text{for Models 2 and 3,}$$
$$\eta \sim \text{Normal}(10, 10) \quad \text{for Models 4 and 5,}$$
$$p \sim \text{Uniform}(0, 1).$$

Model 4 contained the precision matrix $\boldsymbol{Q}$ which was dependent on parameters κ and δ. The precision parameter takes positive values, and was assigned a Uniform(0,+∞), prior. The parameter κ takes values between -1 and 1. It was assigned a Uniform(-1,1) distribution:

$$\delta \sim \text{Uniform}(0, +\infty),$$
$$\kappa \sim \text{Uniform}(-1, 1).$$

Model 5 contained parameter vectors $\boldsymbol{\delta}$ and $\boldsymbol{\kappa}$. Each element of $\boldsymbol{\delta}$ takes positive values, and was assigned Uniform(0, +∞) priors. Each element of $\boldsymbol{\kappa}$ takes values between -1 and 1. It was assigned a Uniform(-1,1) priors:

$$\delta_j \sim \text{Uniform}(0, +\infty),$$
$$\kappa_j \sim \text{Uniform}(-1, 1).$$

All models include the parameter μ. The parameter μ is a real valued intercept term that can theoretically take any value from $-\infty$ to $+\infty$. As a result, a non-informative Uniform distribution with support $(-\infty, +\infty)$ was used as a prior.

$$\mu \sim \text{Uniform}(-\infty, +\infty).$$

Models 1, 2, and 3 include the observation variance σ term. This parameter takes positive values, so a Uniform distribution on $[0, +\infty)$ was used:

$$\sigma \sim \text{Uniform}(0, +\infty).$$

We use non-informative Uniform priors for many of the parameters. This is to reflect our lack of knowledge on those parameters. Improper uniform priors are implementable in Stan (Gelman et al., 2015).

4.1.7   Model Posterior Distributions

The posterior distribution for Model 1 is:

$$
\begin{aligned}
\pi(\mu, \sigma, &\boldsymbol{\tau}, \rho, \phi | \boldsymbol{y}, \boldsymbol{X}) \\
&\propto \pi(\log \boldsymbol{y} | \mu, \sigma, \boldsymbol{\tau}, \rho, \phi, \boldsymbol{X}) \times \pi(\boldsymbol{\tau} | \rho, \phi) \times \pi(\mu) \times \pi(\sigma) \times \pi(\rho) \times \pi(\phi) \\
&\propto \frac{1}{(2\pi)^{\frac{1255}{2}} |\sigma^2 \boldsymbol{I}|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} \big(\log \boldsymbol{y} - (\boldsymbol{1}\mu + \boldsymbol{X}\boldsymbol{\tau})\big)^T \sigma^2 \boldsymbol{I}^{-1} \right. \\
&\quad \left. \times \big(\log \boldsymbol{y} - (\boldsymbol{1}\mu + \boldsymbol{X}\boldsymbol{\tau})\big) \right\} \frac{1}{(2\pi)^{\frac{11}{2}} |\boldsymbol{R}^{-1}|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} \boldsymbol{\tau}^T \boldsymbol{R} \boldsymbol{\tau} \right\}.
\end{aligned}
$$

$$(4.14)$$

The posterior distribution for Model 2 is:

$$
\begin{aligned}
\pi(\mu, \sigma, &\boldsymbol{\omega}, p, \eta | \boldsymbol{y}, \boldsymbol{W}) \\
&\propto \pi(\log \boldsymbol{y} | \mu, \sigma, \boldsymbol{\omega}, p, \eta, \boldsymbol{W}) \times \pi(\boldsymbol{\omega} | p, \eta) \times \pi(\mu) \times \pi(\sigma) \times \pi(p) \times \pi(\eta) \\
&\propto \frac{1}{(2\pi)^{\frac{1255}{2}} |\sigma^2 \boldsymbol{I}|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} \big(\log \boldsymbol{y} - (\boldsymbol{1}\mu + \boldsymbol{W}\boldsymbol{\omega})\big)^T \sigma^2 \boldsymbol{I}^{-1} \right. \\
&\quad \left. \times \big(\log \boldsymbol{y} - (\boldsymbol{1}\mu + \boldsymbol{W}\boldsymbol{\omega})\big) \right\} \frac{1}{(2\pi)^{\frac{22}{2}} |\boldsymbol{S}^{-1}|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} \boldsymbol{\omega}^T \boldsymbol{S} \boldsymbol{\omega} \right\}.
\end{aligned}
$$

$$(4.15)$$

The posterior distribution for Model 3 is:

$$
\begin{aligned}
\pi(\mu, \sigma, &\boldsymbol{\tau}, \boldsymbol{\omega}, \rho, \phi, p, \eta | \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{W}) \\
&\propto \pi(\log \boldsymbol{y} | \mu, \sigma, \boldsymbol{\tau}, \boldsymbol{\omega}, \rho, \phi, p, \eta, \boldsymbol{X}, \boldsymbol{W}) \times \pi(\boldsymbol{\tau} | \rho, \phi) \times \pi(\boldsymbol{\omega} | p, \eta) \\
&\times \pi(\mu) \times \pi(\sigma) \times \pi(\rho) \times \pi(\phi) \times \pi(p) \times \pi(\eta) \\
&\propto \frac{1}{(2\pi)^{\frac{1255}{2}} |\sigma^2 \boldsymbol{I}|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} \big(\log \boldsymbol{y} - (\mathbf{1}\mu + \boldsymbol{X}\boldsymbol{\tau} + \boldsymbol{W}\boldsymbol{\omega}) \big)^T \sigma^2 \boldsymbol{I}^{-1} \right. \\
&\times \left. \big(\log \boldsymbol{y} - (\mathbf{1}\mu + \boldsymbol{X}\boldsymbol{\tau} + \boldsymbol{W}\boldsymbol{\omega}) \big) \right\} \frac{1}{(2\pi)^{\frac{11}{2}} |\boldsymbol{R}^{-1}|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2}\boldsymbol{\tau}^T \boldsymbol{R}\boldsymbol{\tau} \right\} \\
&\times \frac{1}{(2\pi)^{\frac{22}{2}} |\boldsymbol{S}^{-1}|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2}\boldsymbol{\omega}^T \boldsymbol{S}\boldsymbol{\omega} \right\}.
\end{aligned}
$$

$$(4.16)$$

The posterior distribution for Model 4 is:

$$
\begin{aligned}
\pi(\mu, &\boldsymbol{\tau}, \boldsymbol{\omega}, \rho, \phi, p, \eta, \kappa, \delta | \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{W}) \\
&\propto \pi(\log \boldsymbol{y} | \mu, \boldsymbol{\tau}, \boldsymbol{\omega}, \rho, \phi, p, \eta, \kappa, \delta, \boldsymbol{X}, \boldsymbol{W}) \times \pi(\boldsymbol{\tau} | \rho, \phi) \times \pi(\boldsymbol{\omega} | p, \eta) \\
&\times \pi(\mu) \times \times \pi(\rho) \times \pi(\phi) \times \pi(p) \times \pi(\eta) \times \pi(\kappa) \times \pi(\delta) \\
&\propto \frac{1}{(2\pi)^{\frac{1255}{2}} |\boldsymbol{Q}^{-1}|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} \big(\log \boldsymbol{y} - (\mathbf{1}\mu + \boldsymbol{X}\boldsymbol{\tau} + \boldsymbol{W}\boldsymbol{\omega}) \big)^T \boldsymbol{Q} \right. \\
&\times \left. \big(\log \boldsymbol{y} - (\mathbf{1}\mu + \boldsymbol{X}\boldsymbol{\tau} + \boldsymbol{W}\boldsymbol{\omega}) \big) \right\} \frac{1}{(2\pi)^{\frac{11}{2}} |\boldsymbol{R}^{-1}|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2}\boldsymbol{\tau}^T \boldsymbol{R}\boldsymbol{\tau} \right\} \\
&\times \frac{1}{(2\pi)^{\frac{22}{2}} |\boldsymbol{S}^{-1}|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2}\boldsymbol{\omega}^T \boldsymbol{S}\boldsymbol{\omega} \right\} \\
&\times \frac{1}{10 \times (2\pi)^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2}\frac{(\eta - 10)^2}{10^2} \right\}.
\end{aligned}
$$

$$(4.17)$$

The posterior distribution for Model 5 is:

$$
\begin{aligned}
\pi(\mu, &\boldsymbol{\tau}, \boldsymbol{\omega}, \rho, \phi, p, \eta, \boldsymbol{\kappa}, \boldsymbol{\delta} | \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{W}) \\
&\propto \pi(\log \boldsymbol{y} | \mu, \boldsymbol{\tau}, \boldsymbol{\omega}, \rho, \phi, p, \eta, \boldsymbol{\kappa}, \boldsymbol{\delta}, \boldsymbol{X}, \boldsymbol{W}) \times \pi(\boldsymbol{\tau} | \rho, \phi) \times \pi(\boldsymbol{\omega} | p, \eta) \\
&\times \pi(\mu) \times \times \pi(\rho) \times \pi(\phi) \times \pi(p) \times \pi(\eta) \times \prod_{j=1}^{23} \pi(\kappa_j) \times \prod_{j=1}^{23} \pi(\delta_j) \\
&\propto \frac{1}{(2\pi)^{\frac{1255}{2}} |\boldsymbol{Q}^{-1}|^{\frac{1}{2}}} \\
&\exp \left\{ -\frac{1}{2} \big(\log \boldsymbol{y} - (\boldsymbol{1}\mu + \boldsymbol{X}\boldsymbol{\tau} + \boldsymbol{W}\boldsymbol{\omega}) \big)^T \boldsymbol{Q} \big(\log \boldsymbol{y} - (\boldsymbol{1}\mu + \boldsymbol{X}\boldsymbol{\tau} + \boldsymbol{W}\boldsymbol{\omega}) \big) \right\} \\
&\times \frac{1}{(2\pi)^{\frac{11}{2}} |\boldsymbol{R}^{-1}|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} \boldsymbol{\tau}^T \boldsymbol{R} \boldsymbol{\tau} \right\} \\
&\times \frac{1}{(2\pi)^{\frac{22}{2}} |\boldsymbol{S}^{-1}|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} \boldsymbol{\omega}^T \boldsymbol{S} \boldsymbol{\omega} \right\} \\
&\times \frac{1}{10 \times (2\pi)^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} \frac{(\eta - 10)^2}{10^2} \right\}.
\end{aligned}
$$

(4.18)

Models 1 and 2 attempt to explain some of the variation in hoki CPUE using trawl year and strata respectively. By comparing these models to each other and to Model 3 (which contains both variables), we can assess whether accounting for temporal or accounting for spatial effects (or both) is more important. Models 4 and 5 make use of CPUE at the point referenced level, allowing for a different way to account for spatial effects. By comparing Model 4 to Model 3, we will be able to see whether an areal approach explains more variation in hoki CPUE compared to the point referenced approach. Our hypothesis is that the point referenced data would provide more information to the model, and should hence be better in explaining CPUE variability. Models 4 and 5 can be compared to assess whether modelling CPUE correlations separately for each stratum would give a better model.

Model	Iterations	Chains	Warm-up	Thinning	Post. Size
1	50000	4	100000	2	50000
2	50000	4	100000	2	50000
3	50000	4	100000	2	50000
4	10000	2	10000	1	10000
5	10000	2	10000	1	10000

Table 4.3: Summary of each model's fitting procedure

4.2   Implementation

These models were compiled using the software Stan (Gelman et al., 2015), implemented through RStudio (Team, 2015). A summary of the fitting procedure is provided in Table 4.3. For Models 1 – 3, each parameter had four chains of 50000 iterations with over dispersed starting values. This gave posterior distributions with 200000 iterations for each parameter, of which half were discarded as warm-up. A thinning interval of 2 was then applied resulting in a sample size of 50000 for each posterior parameter distribution.

For Models 4 – 5, each parameter had two chains of 10000 iterations with over dispersed starting values. This gave posterior distributions with 20000 iterations for each parameter, of which half were discarded as warm-up. Thinning was not applied to these samples. The resulting size of each posterior parameter distribution was therefore 10000.

We used a smaller number of iterations in Models 4 and 5 because of computational time constraints. Due to the large precision matrices in these models, significantly more time was needed to compute posterior draws.

We assess convergence in Subsection 4.2.1.

4.2.1 Convergence Diagnostics

To assess whether the posterior parameter distributions in our models have converged to stationary distributions, it is necessary to perform convergence diagnostics. We used trace plots (Appendix A), posterior density plots (Appendix B), autocorrelation plots (Appendix C) and Gelman-Rubin plots (Appendix D) to visually monitor the convergence of the parameter distributions. In addition, the software Stan computes the potential scale reduction factor, $\hat{R}$, for each parameter.

Figures A.1 to A.15 display the trace plots for the parameters in Models 1 – 5. For the first four models, there appears to be good mixing for all of the parameters. In addition, there are no obvious patterns suggesting that the chains for each parameter have converged to a stationary distribution. Larger fluctuations in the trace plots for the Gaussian Markov random field parameters ρ, ϕ, p, η, κ, and δ can be seen in all of the models, however, these are of no concern due to their rarity. We have some reservations about whether the parameters of Model 5 have converged to the target posterior distributions. The majority of trace plots for the parameters τ, ω, μ, ρ, ϕ, p, and η display random walk behaviour, even after discarding warm-up samples. This indicates that the HMC chains for Model 5 need to be run for longer.

Figures B.1 – B.15 display the density plots for the posterior parameter distributions of each model. In all plots, we see smooth and symmetric curves, with the exception of ρ, ϕ, p, η, κ, and δ, which are skewed due to being bounded. There is no evidence of over representation, indicating the distributions are sufficiently mixed for all models.

Figures C.1 – C.15 display the autocorrelation plots for the parameters in Models 1 – 5. For all parameters in Models 1 – 4, the correlation between values of increasing lag decreases rapidly. Autocorrelation is virtually zero for all parameters at lag 10 and over. Candidate values are being computed far enough away from the current value, which has allowed for sufficient exploration of the posterior distribution. For Models 2 – 4,

the autocorrelation plots for μ show that autocorrelation decreases slower, however, since the chains are permuted before use, this is not an issue. The autocorrelation plots for Model 5 show that there is still significant autocorrelation at lag 40 for the parameters ω, μ, and p.

Figures D.1 – D.21 display the Gelman-Rubin plots for the parameters in Models 1 – 5. In the plots for Models 1 – 4, we see that the potential scale reduction factor (called "shrink factor" on the plots) reduces to one at the end of each chain. This indicates convergence is satisfied for the parameters in each of these models. In general for Model 4, the potential scale reduction factor takes longer to converge to 1. For Model 5, the Gelman-Rubin plots show that the potential scale reduction factor takes much longer to converge to 1, and in some cases (for example, ω) does not converge to 1 at all.

We are satisfied with the convergence of parameter distributions for Models 1 – 4. However, for Model 5, it is clear that convergence to the target densities has not been achieved. In order to rectify this, the HMC chains for Model 5 should be run for more iterations. Unfortunately, due to time restraints, we must continue with what we currently have for Model 5.

Chapter 5

Results

This chapter presents the main results from fitting the five models described in Chapter 4. In each section, we interpret the posterior density plots (see Appendix B) for the parameter estimates of the fitted models. In addition, we provide tables for each model summarising the posterior parameter distributions by their median and 95% highest posterior density (HPD) interval. We also compute model predicted catch per unit effort values using the posterior distributions and the data likelihoods for each model. These are presented visually, as well as summarised in tables. Throughout this chapter, comparisons are made to the observed data. At the end of this chapter, a section is dedicated to the comparison of the models.

5.1   Model 1: Temporal Effects

Posterior distributions for the parameters $\mu, \tau_2, ..., \tau_{12}, \sigma, \rho$, and ϕ in Model 1 (defined in Table 4.1) are presented in Figures B.1 and B.2. Table 5.1 summarises these posterior parameter distributions.

Figure B.1 gives the posterior densities of $\tau_2, ..., \tau_{12}$, which are the temporal effects on hoki catch per unit effort for the years $1992 - 2008$ with respect to the year 1991 (the reference level for the temporal effects). First of all,

Table 5.1: Summary of the posterior distributions for the parameters in Model 1.

	Median	95% HPD
μ	5.58	(5.40, 5.76)
τ_2	0.09	(-0.16, 0.35)
τ_3	0.11	(-0.15, 0.36)
τ_4	-0.72	(-1.00, -0.44)
τ_5	-0.96	(-1.24, -0.67)
τ_6	-1.07	(-1.35, -0.79)
τ_7	-1.63	(-1.95, -1.32)
τ_8	-1.36	(-1.66, -1.07)
τ_9	-1.41	(-1.70, -1.12)
τ_{10}	-1.42	(-1.72, -1.12)
τ_{11}	-0.78	(-1.07, -0.49)
τ_{12}	-0.70	(-0.99, -0.41)
σ	1.19	(1.14, 1.23)
ρ	0.85	(0.66, 0.99)
ϕ	1.81	(0.12, 4.19)

we observe a noticeable difference between the effects of each trawl year. The posterior distributions of the effects for 1992, and 1993 have most of their density above zero suggesting that average CPUE was higher in these years than it was in 1991. Furthermore, the effects for the later trawls in years 2000 – 2008 have posterior distributions centred below zero suggesting that CPUE was lower in these years compared to in 1991.

The posterior densities for the temporal effect parameters are symmetric. We summarise the densities in Table 5.1 by their corresponding medians and 95% HPD intervals.

Model predicted CPUE values were calculated for each trawl year, using the posterior distributions and the data likelihood for Model 1. For each

value of the HMC chain we calculated:

$$
\widehat{\text{CPUE}}_t = \begin{cases} \exp(\hat{\mu}), & \text{for } t = 1991 \\ \exp(\hat{\mu} + \hat{\tau}_t), & \text{for } t = 1992, ..., 1993, 2000, ..., 2008, \end{cases} \tag{5.1}
$$

where $\widehat{\text{CPUE}}_t$ is the model predicted CPUE value for trawl year t, $\hat{\mu}$ is the estimate for μ, and $\hat{\tau}_t$ is the estimate for τ_t, the effect of trawl year t.

The model predicted CPUE distributions and the observed median CPUE for each trawl year are shown in Figure 5.1 and summarised in Table 5.2. We can see that each density overlaps with the observed median CPUE for each trawl year. In general, the model predicts median CPUE better in the middle trawl years (2000 – 2006), than in the earlier trawl years (1991 – 1993) and the later trawl years (2007 – 2008). This is indicated by the fact that the observed CPUE values are closer to centre of the predicted densities in years 2000 – 2006, and are in the tails of the densities in 1991 – 1993 and 2007 – 2008. The fact that not all of the observed CPUE values line up with the centres of the model predicted CPUE densities suggests there is some unexplained variation in hoki CPUE.

Figure B.2 displays the posterior densities of the rest of the parameters in Model 1. The posterior density for μ represents the model predicted log CPUE for the reference trawl year, 1991.

The posterior distribution for ρ estimates the correlation in hoki log CPUE between successive trawl years. The 95 % highest posterior density interval for the posterior density of ρ is $(0.66, 0.99)$, and the posterior median is 0.85, which suggests there is strong and positive year-to-year autocorrelation within hoki CPUE that is captured by the model.

The parameter ϕ gives a measure of the total variation of mean CPUE between trawl years. The 95% HPD interval for ϕ is (0.12, 4.19).

Residuals were calculated by taking the median of $(\log \text{CPUE} - \log \widehat{\text{CPUE}}_t)$, where $\widehat{\text{CPUE}}_t$ is the model predicted CPUE for all iterations in the HMC chain for year t.

Figure 5.2 displays a plot of the residuals. The first thing to note is the few
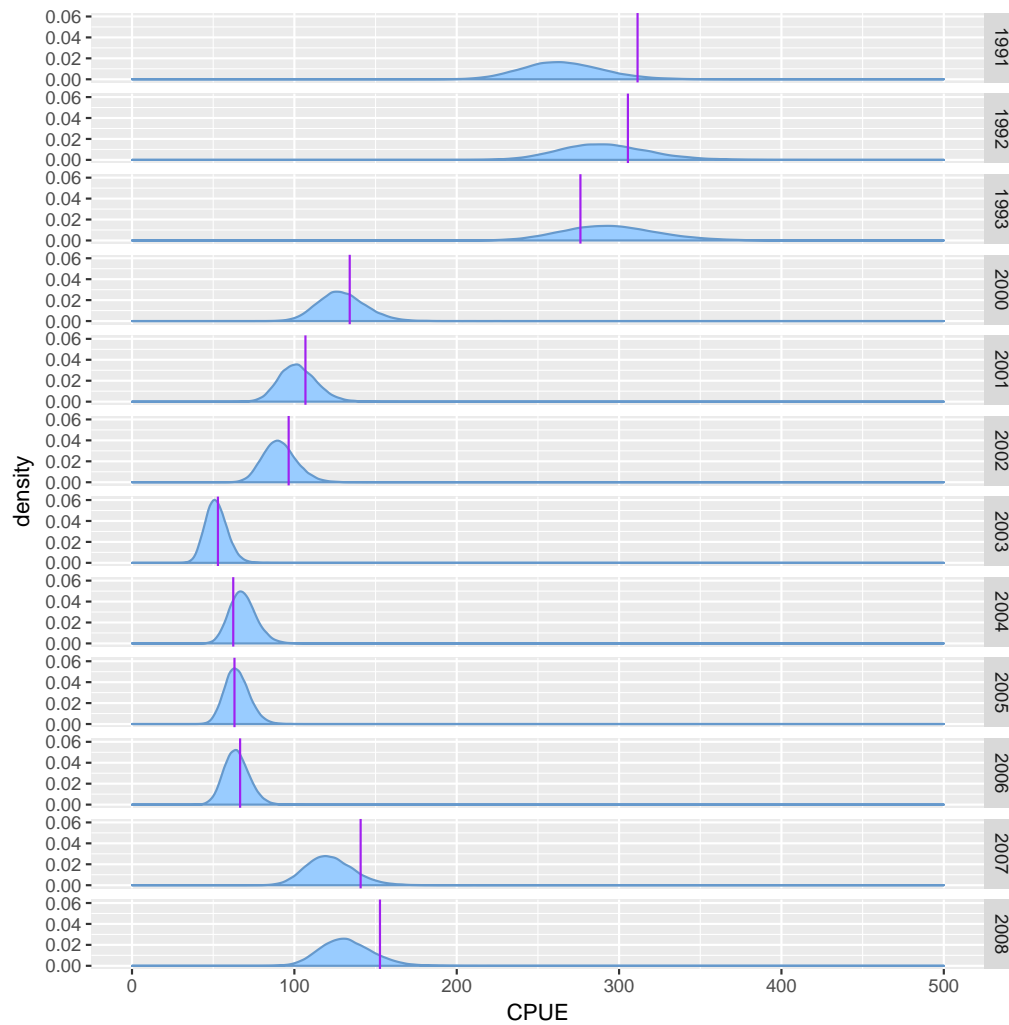
Figure 5.1: The posterior distributions of CPUE by trawl year, as predicted by Model 1. Vertical lines represent the observed CPUE for each year.
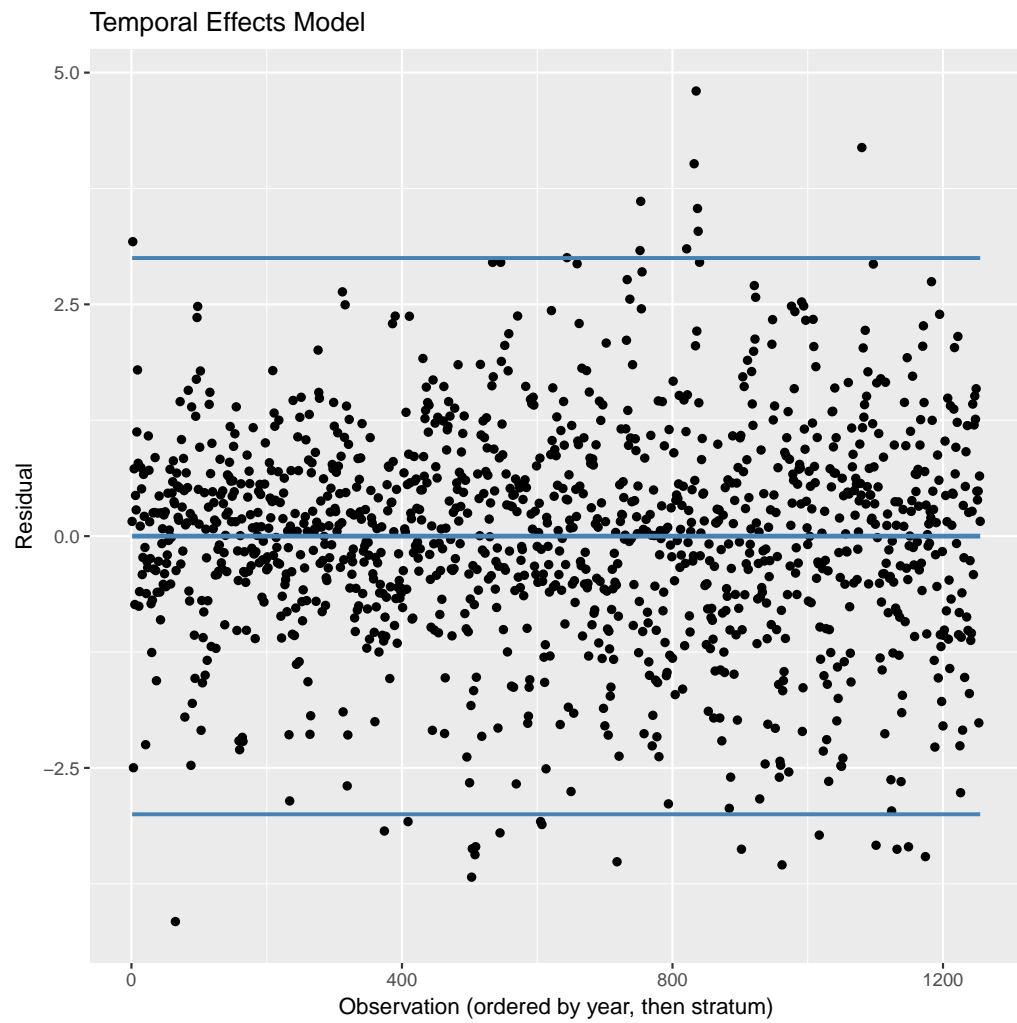
Figure 5.2: Residuals from Model 1, ordered by trawl year.

Table 5.2: The median CPUE for each trawl year, as predicted by Model 1 and observed in the dataset.

Year	Predicted	(95% HPD)	Observed
1991	264.41	(218.91, 314.25)	311.38
1992	289.74	(240.23, 344.61)	305.44
1993	293.60	(241.22, 353.05)	276.19
2000	128.64	(102.31, 158.43)	134.09
2001	101.59	(81.23, 125.21)	106.82
2002	90.48	(72.16, 111.93)	96.48
2003	51.57	(39.35, 65.16)	52.92
2004	67.81	(52.92, 84.56)	62.35
2005	64.29	(50.32, 79.74)	63.10
2006	64.09	(49.78, 79.77)	66.57
2007	121.16	(94.93, 151.16)	140.81
2008	131.26	(102.60, 163.66)	152.72

large outliers associated with the year 2004. There were a few trawls in 2004 which obtained higher than usual catch weights for hoki. If we ignore the upper two outliers in 2004, we observe an even, but wide, spread of residuals. This suggests that most of the variation in CPUE is explained by trawl year, but there is some variation left unexplained.

In Section 5.2, we look at the results of the areal spatial effects models, which will illustrate the importance of including trawl year to account for temporal autocorrelation.

5.2   Model 2: Areal Spatial Effects

Posterior densities for the parameters $\mu, \omega_2, ..., \omega_{23}, \sigma, p,$ and η in Model 2 (defined in Table 4.1) are presented in Figures B.3 and B.4. Table 5.3 summarises these posterior parameter distributions.

Figure B.3 gives the posterior densities for $\omega_2, ..., \omega_{23}$, the effects due to stratum on hoki catch per unit effort in 22 of the strata with respect to stratum 1 (the reference level). We observe differences in the stratum effects, most noticeably between strata 26, 27, 28, and all others. In addition to this, we can see that the strata form clusters with relatively similar effects on hoki CPUE. One such cluster would be strata 26, 27, and 28. Another would be strata 10 to 15. Initially it was thought that the groups of strata would correspond to the depth intervals they represent. Figure 5.3 shows separation between the depth interval 800m-1000m and the others, however, there is no clear difference in hoki CPUE for depth intervals 300m – 600m and 600m – 800m. Therefore clustering may be based on some other variable (for example, sea temperature).

All of the stratum effects have the majority of their densities below zero, indicating that CPUE was lower in strata 2 to 28, compared to stratum 1. The model suggests that strata 26, 27, and 28 had much lower CPUE compared to all others. Strata 27 and 28 are neighbouring, suggesting that hoki catches in the north eastern area of the sub-Antarctic region is lower than near Puysegur bank.

Model predicted CPUE values were calculated for each stratum, using the posterior distributions and the data likelihood for Model 2. This is given by

$$
\widehat{\text{CPUE}}_w = \begin{cases} \exp(\hat{\mu}), & \text{for } w = 1 \\ \exp(\hat{\mu} + \hat{\omega}_w), & \text{for } w = 2, ..., 23 \end{cases} \tag{5.2}
$$

where $\widehat{\text{CPUE}}_w$ is the model predicted CPUE value for stratum w, $\hat{\mu}$ is the estimate for μ, and $\hat{\omega}_t$ is the estimate for ω_w, the effect of stratum w.

The model predicted log CPUE distributions and log observed CPUE for each stratum are shown in Figure 5.3 and summarised in Table 5.4. The model predicted log CPUE densities overlap with the log observed median CPUE for most strata. In some cases, such as for stratum 3, and 5, and 26, we can see that the density of predicted log CPUE has more spread

compared with the rest of the strata.

In general, the difference in observed and predicted median CPUE is larger by stratum than it is by year. This suggests that trawl year accounts for more variation in hoki CPUE than stratum.

Figure B.4 displays the posterior densities of the rest of the parameters in Model 2. The posterior density for μ represents the model predicted log CPUE for the reference stratum, 1.

The posterior distribution of p has 95% of its highest density between 0.87 and 1. This suggests that there is a strong association between the mean CPUEs of each stratum. The value of η gives a measure of the total variation of mean CPUE between strata. The 95% HPD interval for η is (0.30, 1.40).

Residuals were calculated by taking the median of $(\log \text{CPUE} - \log \widehat{\textbf{CPUE}}_w)$, where $\widehat{\textbf{CPUE}}_w$ is the model predicted CPUE for all iterations in the HMC chain for stratum w.

Figure 5.4 displays a plot of the residuals between the model predicted log CPUE and the observed log CPUE ordered by time. The first thing that strikes the viewer is the obvious trend over time. A slow decline from positive to negative residuals can be seen in the first two-thirds of the plot, followed by a sharper increase from negative to positive residuals in the final third. This matches the trend observed in CPUE over time, as reflected by the box plot displayed in Figure 2.9. This indicates that Model 2 has not accounted for the obviously apparent year effect within CPUE, and highlights the importance of including trawl year as an explanatory variable.

In the following section, we see the main results from Model 3, which contains both a year component, and stratum component to explain variation within hoki CPUE.
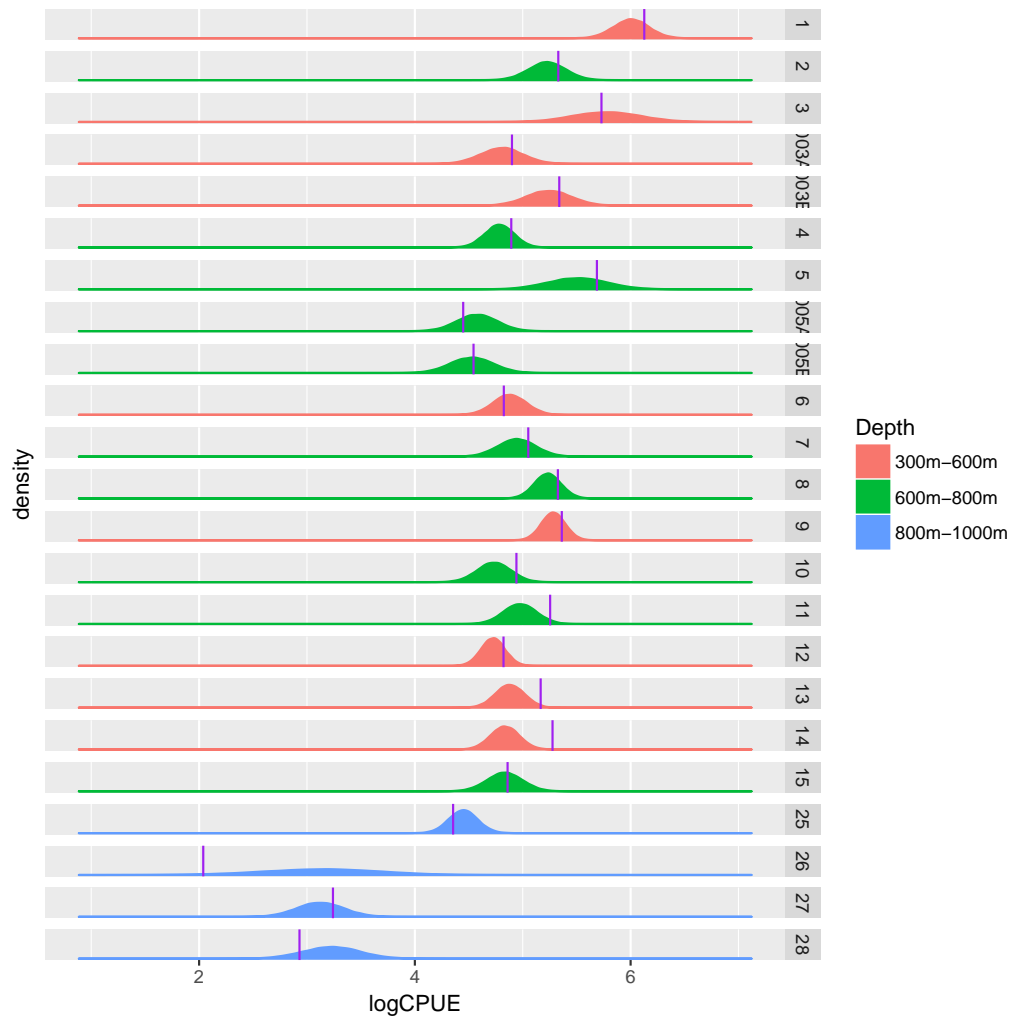
Figure 5.3: The posterior distributions of log CPUE by stratum, as predicted by Model 2. Vertical lines represent the observed median log CPUE.

Table 5.3: Summary of the posterior distributions for the parameters in Model 2.

	Median	95% HPD
μ	6.01	(5.68, 6.32)
ω_2	-0.78	(-1.24, -0.32)
ω_3	-0.23	(-0.93, 0.46)
ω_4	-1.20	(-1.70, -0.70)
ω_5	-0.76	(-1.28, -0.25)
ω_6	-1.22	(-1.63, -0.79)
ω_7	-0.50	(-1.12, 0.13)
ω_8	-1.43	(-1.93, -0.95)
ω_9	-1.48	(-1.97, -0.95)
ω_{10}	-1.13	(-1.57, -0.69)
ω_{11}	-1.06	(-1.54, -0.60)
ω_{12}	-0.77	(-1.18, -0.38)
ω_{13}	-0.72	(-1.11, -0.33)
ω_{14}	-1.27	(-1.72, -0.83)
ω_{15}	-1.03	(-1.46, -0.58)
ω_{16}	-1.28	(-1.65, -0.88)
ω_{17}	-1.13	(-1.54, -0.71)
ω_{18}	-1.17	(-1.58, -0.76)
ω_{19}	-1.17	(-1.62, -0.71)
ω_{20}	-1.56	(-1.97, -1.14)
ω_{21}	-2.87	(-4.01, -1.76)
ω_{22}	-2.88	(-3.43, -2.36)
ω_{23}	-2.78	(-3.41, -2.18)
σ	1.23	(1.18, 1.28)
p	0.97	(0.87, 1.00)
η	0.75	(0.30, 1.40)

Table 5.4: The median CPUE for each stratum, as predicted by Model 2 and observed in the dataset.

	Predicted	95% HPD	Observed	Depth
1	405.74	(282.25, 542.53)	457.88	300m–600m
2	186.10	(128.56, 255.12)	206.37	600m–800m
3	323.57	(152.36, 559.80)	307.90	300m–600m
003A	122.31	(80.60, 174.22)	134.32	300m–600m
003B	189.06	(118.62, 272.57)	208.25	300m–600m
4	119.50	(89.84, 153.64)	133.46	600m–800m
5	246.89	(128.91, 396.80)	295.18	600m–800m
005A	97.23	(65.09, 136.55)	85.47	600m–800m
005B	92.60	(60.39, 132.96)	94.06	600m–800m
6	131.47	(95.30, 174.37)	124.55	300m–600m
7	140.47	(95.32, 193.02)	156.25	600m–800m
8	186.99	(143.79, 235.24)	205.57	600m–800m
9	197.04	(156.79, 242.59)	213.20	300m–600m
10	113.77	(81.44, 151.65)	139.98	600m–800m
11	144.72	(104.41, 191.80)	191.34	600m–800m
12	112.73	(89.01, 138.08)	124.19	300m–600m
13	131.71	(99.68, 168.90)	175.25	300m–600m
14	125.83	(95.11, 160.67)	195.64	300m–600m
15	125.65	(87.94, 169.86)	128.98	600m–800m
25	85.51	(64.74, 110.02)	77.83	800m–1000m
26	22.99	(5.18, 55.26)	7.68	800m–1000m
27	22.72	(14.18, 33.61)	25.56	800m–1000m
28	25.22	(13.78, 39.95)	18.74	800m–1000m
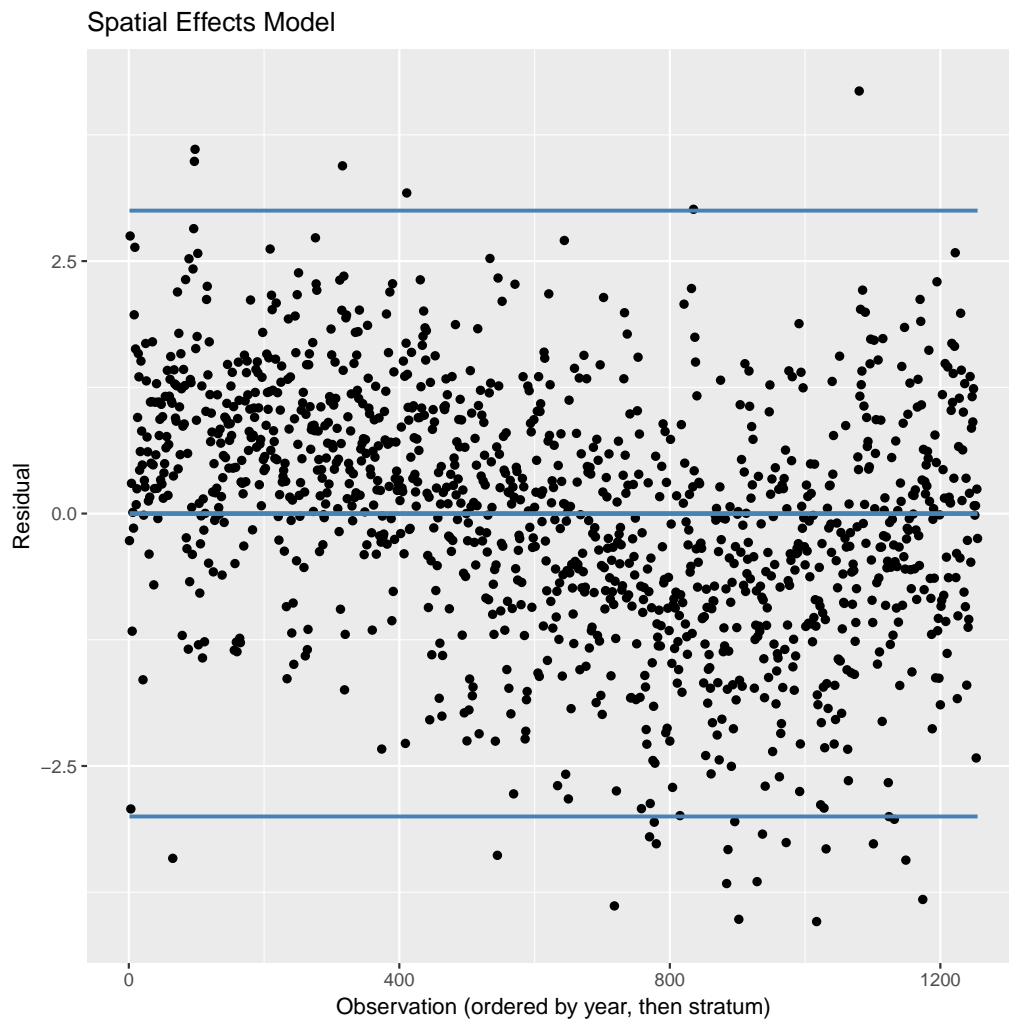
Figure 5.4: Residuals from Model 2, ordered by trawl year.

5.3 Model 3: Temporal and Areal Spatial Effects

Posterior densities for the parameters in Model 3 (defined in Table 4.1) are presented in Figures B.5 to B.7. Table 5.5 summarises these posterior parameter distributions.

Figure B.5 gives the posterior densities of $\tau_2, ..., \tau_{12}$, which are the temporal effects on hoki catch per unit effort for the years $1992 - 2008$ with respect to the year 1991 (the reference level for the temporal effects). The distributions of the effects due to trawl year for Model 3 are nearly identical to the ones in Model 1. We observe a noticeable difference between the effects of each trawl year. The posterior distributions of the effects for 1992, and 1993 have most of their density above zero suggesting that CPUE was higher in these years than it was in 1991. Furthermore, the effects for the later trawls in years $2000 - 2008$ have posterior distributions centred below zero suggesting that CPUE was lower in these years compared to in 1991.

Figure B.6 gives the posterior densities for $\omega_2, ..., \omega_{23}$, the effects due to stratum on hoki catch per unit effort in 22 of the strata with respect to stratum 1 (the reference level). Again, we see a striking resemblance to the densities observed for the stratum effects in Model 2. As we saw for Model 2, we observe differences in the stratum effects, most noticeable between strata 26, 27, 28, and all others. In addition, the same stratum clustering observed in Model 2 is evident here.

All of the stratum effects have the majority of their densities below zero, indicating that CPUE was lower in strata 2 to 28, compared to stratum 1. The model suggests that strata 26, 27, and 28 had much lower CPUE compared to all others. Strata 27 and 28 are neighbouring, suggesting that hoki catches in the north eastern area of the sub-Antarctic region is lower than near Puysegur bank.

Model predicted CPUE for each year ($\widehat{\text{CPUE}}_t$) and each stratum ($\widehat{\text{CPUE}}_w$) for each iteration in the HMC chains were calculated by aggregating over

strata and year respectively:

$$\widehat{\mathrm{CPUE}}_t = \sum_{w=1}^{23} \widehat{\mathrm{CPUE}}_{t,w}, \tag{5.3}$$

$$\widehat{\mathrm{CPUE}}_w = \sum_{t=1}^{12} \widehat{\mathrm{CPUE}}_{t,w}, \tag{5.4}$$

where

$$\widehat{\mathrm{CPUE}}_{t,w} = \begin{cases} \exp(\hat{\mu}) & \text{for } t = 1, w = 1 \\ \exp(\hat{\mu} + \hat{\omega}_w) & \text{for } t = 1, w = 2, ..., 23 \\ \exp(\hat{\mu} + \hat{\tau}_t) & \text{for } t = 2, ..., 12, w = 1 \\ \exp(\hat{\mu} + \hat{\tau}_t + \hat{\omega}_w) & \text{for } t = 2, ..., 12, w = 2, ..., 23 \end{cases} \tag{5.5}$$

The model predicted CPUE distributions and the observed median CPUE for each trawl year are shown in Figure 5.5 and summarised in Table 5.6. We can see that most densities overlap with the observed median CPUE for each trawl year. In general, the model predicts median CPUE better in the years 1991 and 2000 – 2008 than the years 1992 and 1993. This is indicated by the fact that the observed CPUE values are closer to centre of the densities in years 1991, and 2000 – 2008, and are in the tails of the densities in 1992 and 1993. Compared to Model 1, the predictions for the years 1991, 2007 and 2008 have improved. However, the fact that not all of the observed CPUE values line up with the centres of the model predicted CPUE densities suggests there is some unexplained variation in hoki CPUE.

The model predicted log CPUE distributions and log observed CPUE for each stratum are shown in Figure 5.5 and summarised in Table 5.7. The model predicted log CPUE densities overlap with the log observed median CPUE for some strata. Prediction of CPUE by strata appears to be less accurate in Model 3 compared to Model 2. This may be due to the aggregation over year.

Table 5.5: Summary of the posterior distributions for the parameters in Model 3.

	Median	95% HPD		Median	95% HPD
μ	6.98	(6.65, 7.29)	ω_2	-0.84	(-1.24, -0.45)
τ_2	0.09	(-0.14, 0.32)	ω_3	-1.23	(-1.85, -0.62)
τ_3	0.11	(-0.13, 0.34)	ω_4	-1.09	(-1.51, -0.65)
τ_4	-0.72	(-0.98, -0.47)	ω_5	-0.60	(-1.05, -0.16)
τ_5	-0.92	(-1.17, -0.66)	ω_6	-1.58	(-1.94, -1.22)
τ_6	-1.11	(-1.36, -0.84)	ω_7	-1.42	(-1.96, -0.87)
τ_7	-1.69	(-1.99, -1.41)	ω_8	-1.24	(-1.66, -0.81)
τ_8	-1.52	(-1.79, -1.25)	ω_9	-1.34	(-1.79, -0.91)
τ_9	-1.48	(-1.75, -1.21)	ω_{10}	-1.42	(-1.80, -1.04)
τ_{10}	-1.52	(-1.80, -1.25)	ω_{11}	-1.33	(-1.73, -0.92)
τ_{11}	-0.86	(-1.14, -0.60)	ω_{12}	-1.15	(-1.49, -0.80)
τ_{12}	-0.71	(-0.98, -0.44)	ω_{13}	-0.95	(-1.28, -0.62)
σ	1.06	(1.02, 1.11)	ω_{14}	-1.68	(-2.05, -1.29)
ρ	0.86	(0.68, 0.99)	ω_{15}	-1.37	(-1.75, -0.99)
ϕ	1.71	(0.16, 3.99)	ω_{16}	-1.63	(-1.96, -1.29)
p	0.98	(0.93, 1.00)	ω_{17}	-1.50	(-1.86, -1.15)
η	0.96	(0.36, 1.76)	ω_{18}	-1.71	(-2.07, -1.35)
			ω_{19}	-1.54	(-1.93, -1.14)
			ω_{20}	-1.71	(-2.06, -1.35)
			ω_{21}	-2.96	(-3.94, -2.00)
			ω_{22}	-2.75	(-3.21, -2.28)
			ω_{23}	-2.74	(-3.26, -2.21)

In general, the difference in observed and predicted median CPUE is larger by stratum than it is by year. This suggests that trawl year accounts for more variation in hoki CPUE than stratum.

Figure B.7 displays the posterior densities of the rest of the parameters in Model 2. The posterior density for μ represents the model predicted log CPUE for the reference year, 1991, and reference stratum, 1.

The posterior distribution for ρ estimates the correlation in hoki log CPUE between successive trawl years. The 95% highest posterior density interval for the posterior density of ρ is $(0.68, 0.99)$, and the posterior median is 0.86 which, suggests there is strong and positive year-to-year autocorrelation within hoki CPUE that is captured by the model.

The parameter ϕ gives a measure of the total variation of mean CPUE between trawl years. The 95% HPD interval for ϕ is (0.16, 3.99).

The posterior distribution of p has 95% of its highest density between 0.93 and 1. This suggests that there is a strong association between the mean CPUEs of each stratum. The value of η gives a measure of the total variation of mean CPUE between strata. The 95% HPD interval for η is (0.36, 1.76).

Residuals were calculated by taking the median of ($\log \text{CPUE} - \log \widehat{\text{CPUE}}_{t,w}$), where $\widehat{\text{CPUE}}_{t,w}$ is the model predicted CPUE for all iterations in the HMC chain for for year t, and stratum w.

Figure 5.7 displays a plot of the residuals between the model predicted log CPUE and the observed log CPUE. We see an even, but wide spread of residuals, with no obvious curvature. This suggests that most of the variation in CPUE is explained by trawl year and stratum, but there is some variation left unexplained. There are fewer potential outliers compared to the residuals of Model 1, suggesting that more variation in CPUE is explained by Model 3.
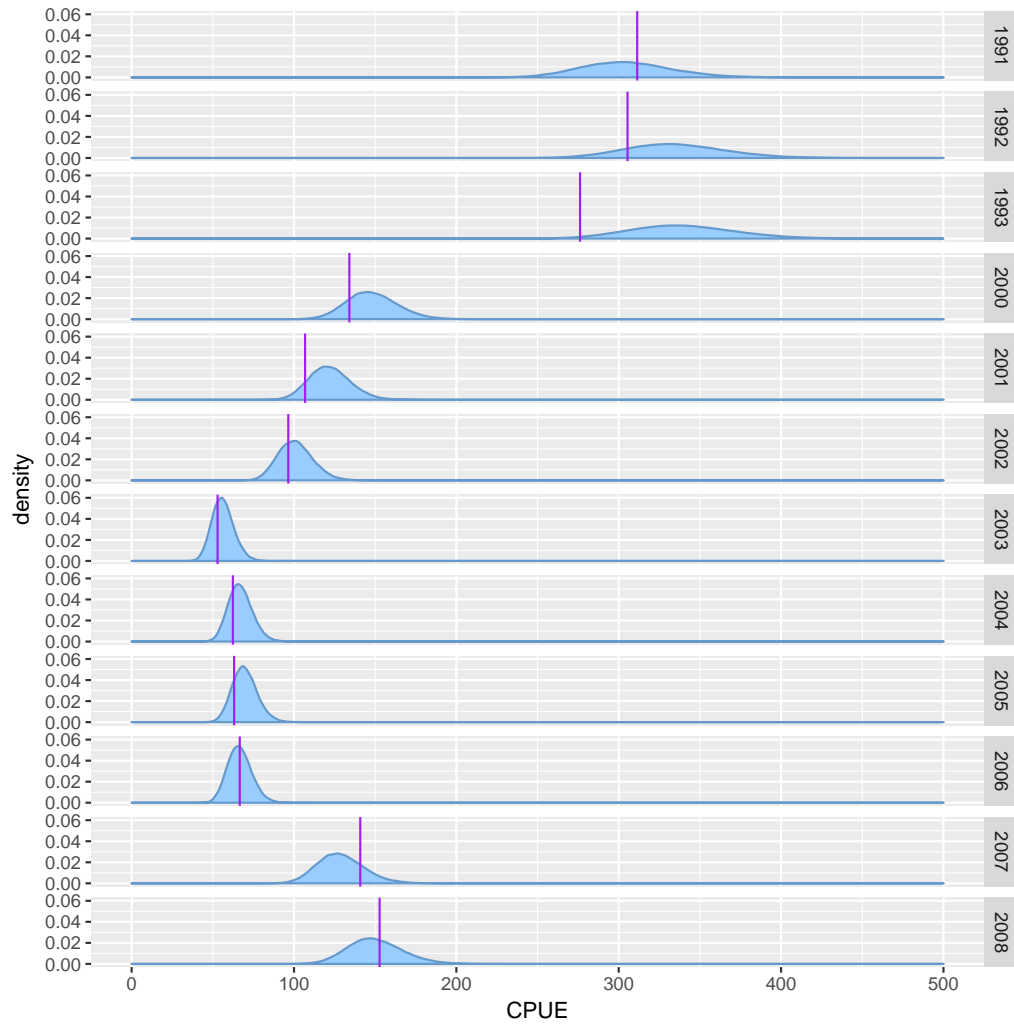
Figure 5.5: The posterior distributions of CPUE by trawl year, as predicted by Model 3. Here, the predictions were aggregated over strata. Vertical lines represent the observed CPUE.
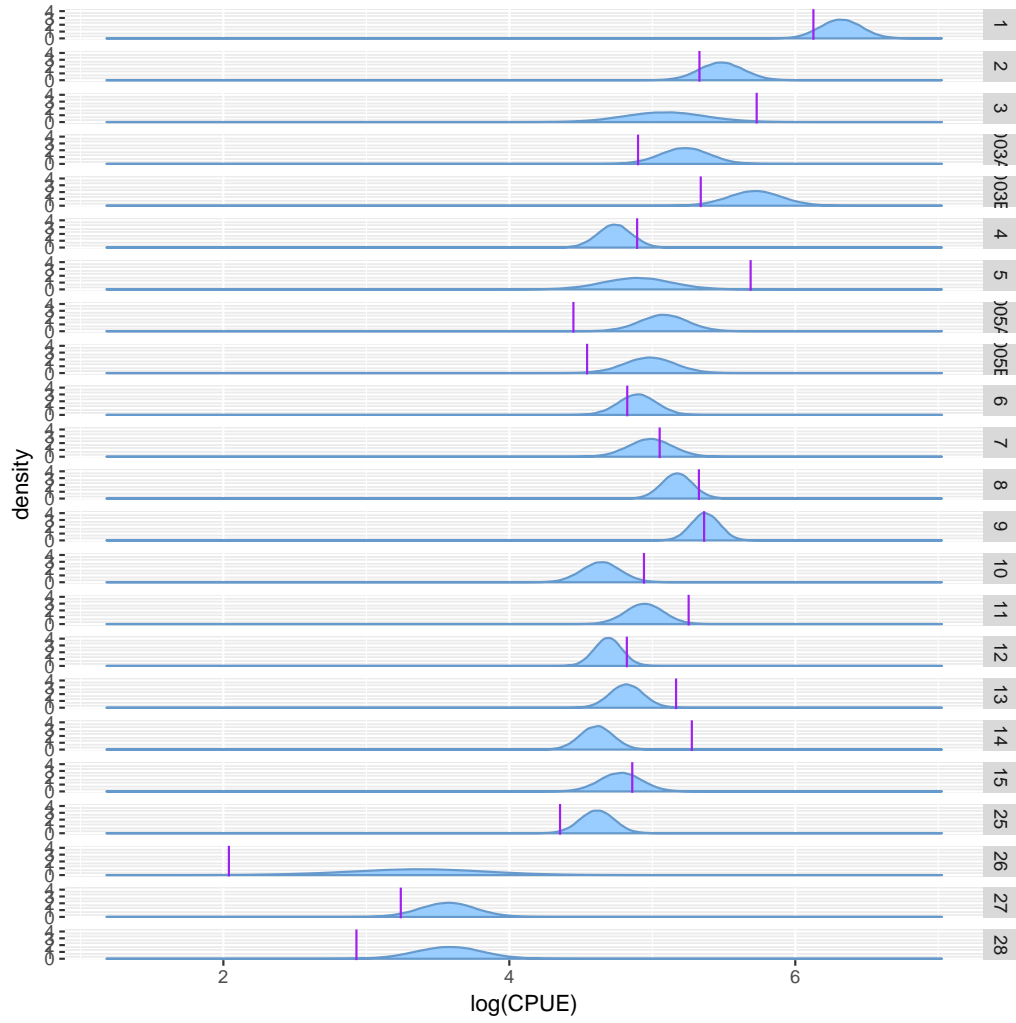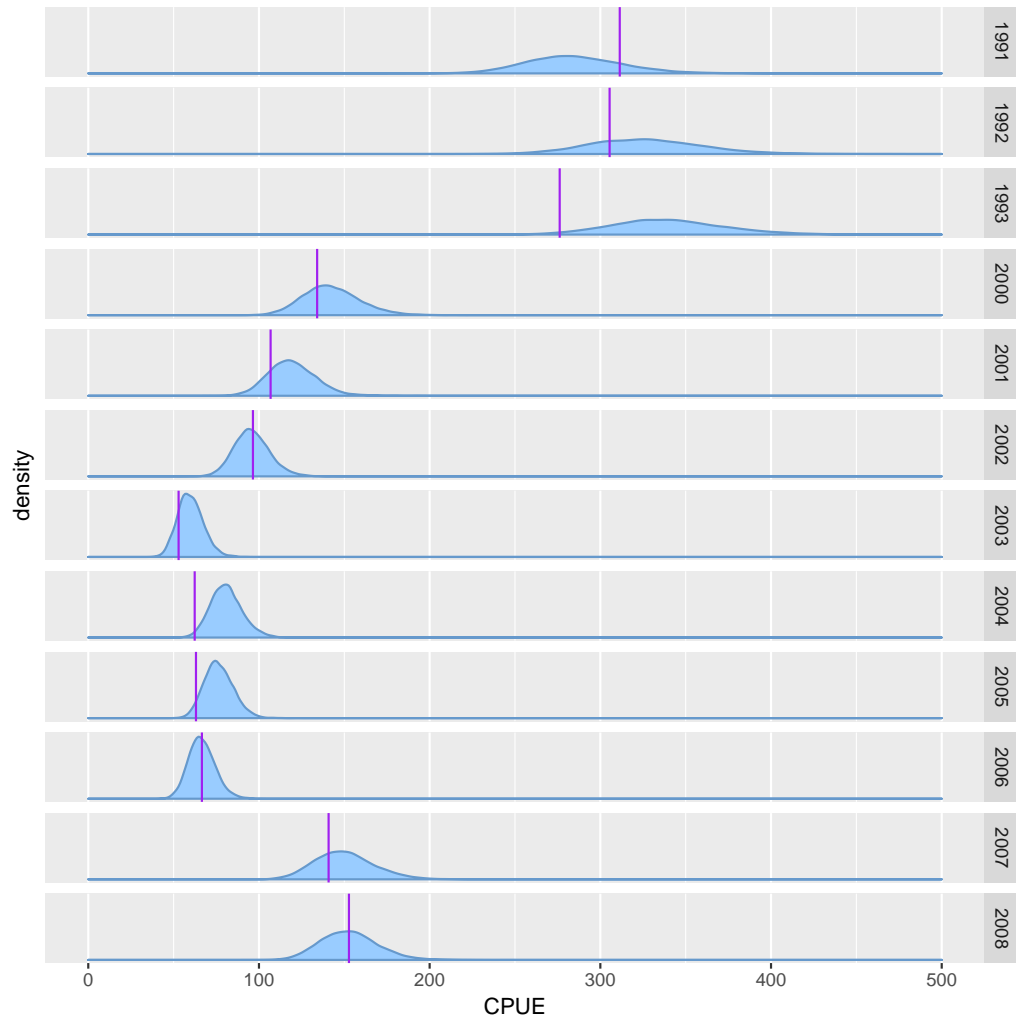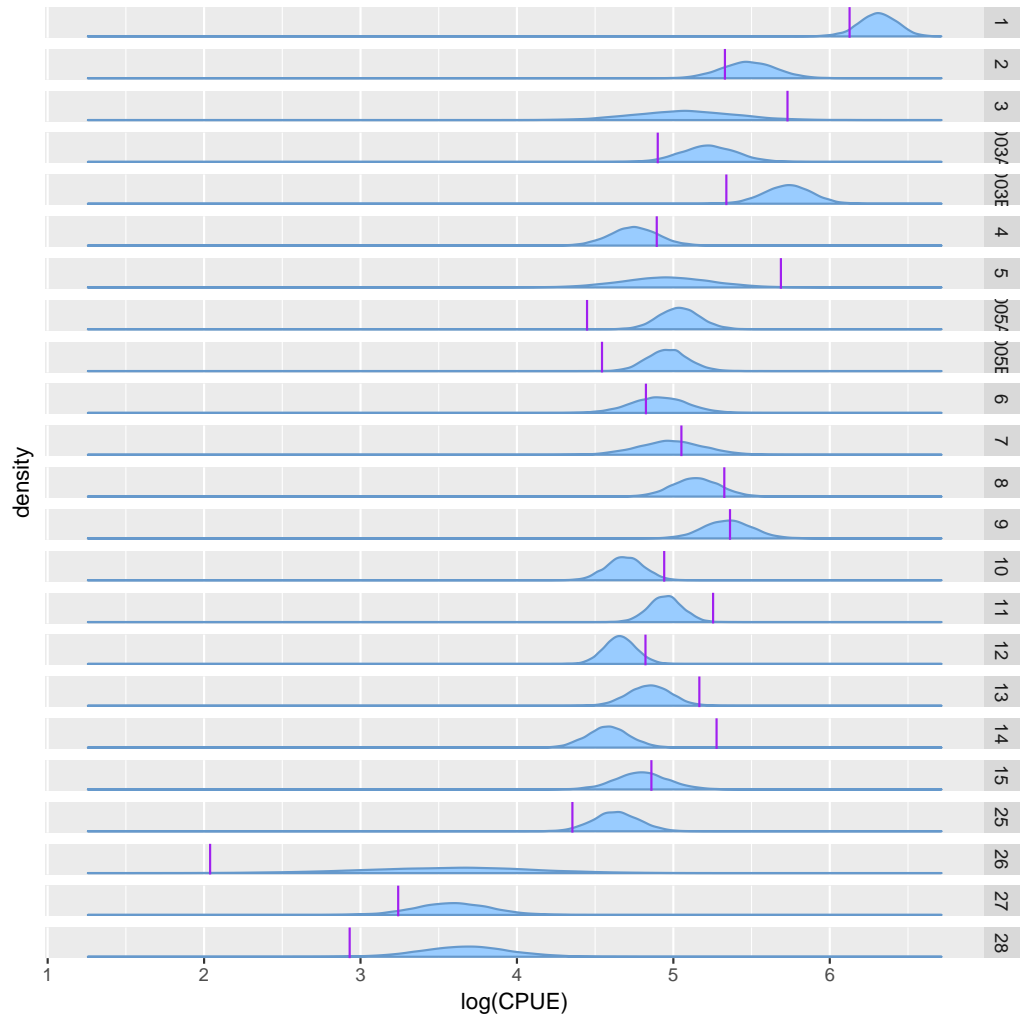
Figure 5.6: The posterior distributions of log CPUE by stratum, as predicted by Model 3. Here, the predictions were aggregated over trawl year. Vertical lines represent the observed log CPUE.
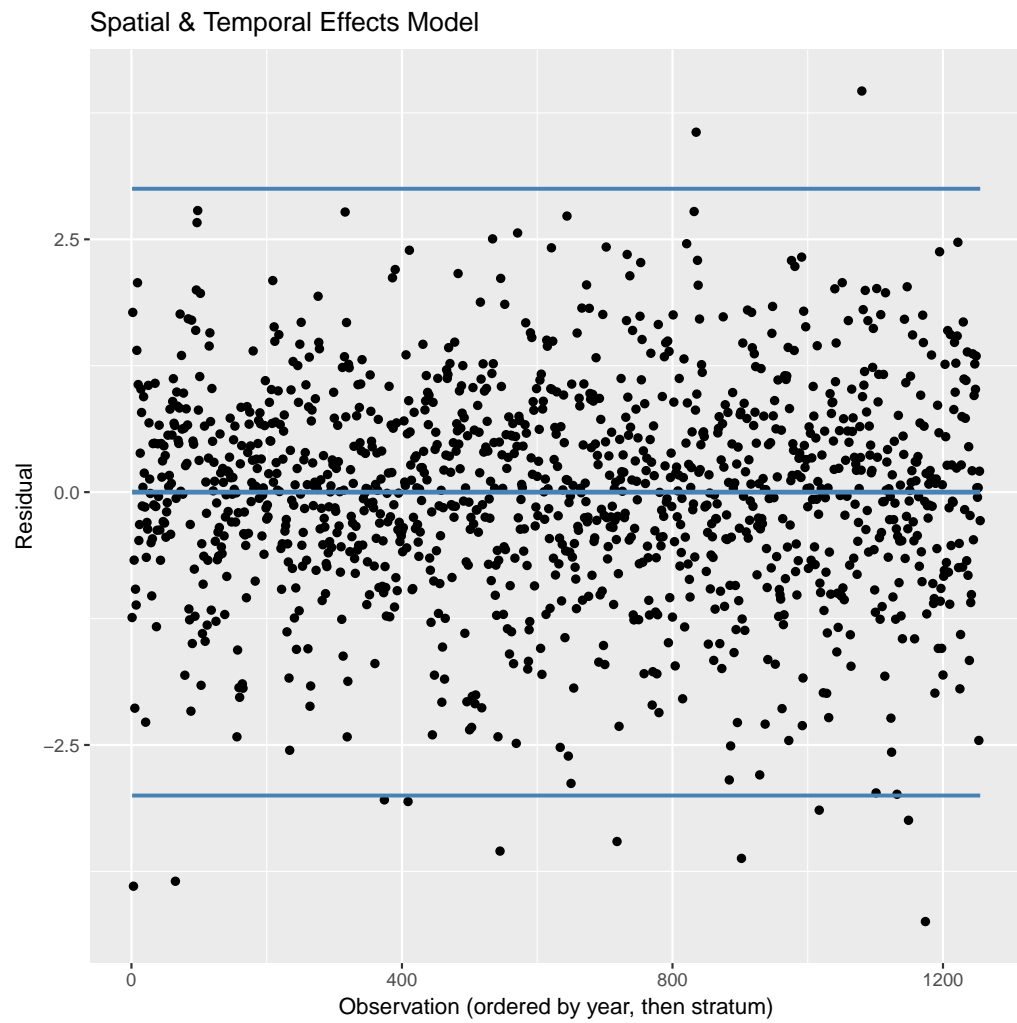
Figure 5.7: Residuals from Model 3, ordered by trawl year.

Table 5.6: The median CPUE for each trawl year, as predicted by Model 3 and observed in the dataset.

Year	Predicted	95% HPD	Observed
1991	303.61	(411.67, 720.40)	311.38
1992	333.49	(173.81, 317.32)	305.44
1993	338.03	(85.52, 263.67)	276.19
2000	147.21	(128.47, 252.77)	134.09
2001	121.47	(202.27, 424.94)	106.82
2002	100.39	(89.49, 142.15)	96.48
2003	55.85	(77.75, 203.75)	52.92
2004	66.52	(111.83, 216.92)	62.35
2005	69.25	(99.37, 199.73)	63.10
2006	66.13	(101.22, 170.78)	66.57
2007	127.96	(104.61, 193.23)	140.81
2008	149.06	(141.16, 215.58)	152.72

5.4   Model 4: Point Referenced CPUE Model

Posterior densities for the parameters in Model 4 (defined in Table 4.1) are presented in Figures B.8 to B.10. Table 5.8 summarises these posterior parameter distributions.

Figure B.8 gives the posterior densities of $\tau_2, ..., \tau_{12}$, which are the temporal effects on hoki catch per unit effort for the years 1992 – 2008 with respect to the year 1991 (the reference level for the temporal effects). The distributions of the effects due to trawl year for Model 4 are similar to the ones in Model 1 and Model 3. We observe a noticeable difference between the effects of each trawl year. The posterior distributions of the effects for 1992, and 1993 have most of their density above zero suggesting that CPUE was higher in these years than it was in 1991. Furthermore, the effects for the later trawls in years 2000 – 2008 have posterior distributions centred below zero suggesting that CPUE was lower in these years compared to in

Table 5.7: The median CPUE for each stratum, as predicted by Model 3 and observed in the dataset.

Stratum	Predicted	95% HPD	Observed
1	554.84	(411.67, 720.40)	457.88
2	240.48	(173.81, 317.32)	206.37
3	162.27	(85.52, 263.67)	307.90
003A	186.55	(128.47, 252.77)	134.32
003B	304.65	(202.27, 424.94)	208.25
4	113.96	(89.49, 142.15)	133.46
5	133.68	(77.75, 203.75)	295.18
005A	160.81	(111.83, 216.92)	85.47
005B	145.76	(99.37, 199.73)	94.06
6	133.84	(101.22, 170.78)	124.55
7	146.49	(104.61, 193.23)	156.25
8	176.57	(141.16, 215.58)	205.57
9	214.97	(175.70, 258.21)	213.20
10	103.79	(78.35, 133.38)	139.98
11	140.50	(105.70, 179.37)	191.34
12	108.97	(89.14, 130.20)	124.19
13	123.94	(97.87, 154.08)	175.25
14	100.57	(79.70, 125.05)	195.64
15	119.03	(86.43, 154.02)	128.98
25	100.84	(79.04, 125.47)	77.83
26	28.66	(8.41, 62.08)	7.68
27	35.57	(23.12, 50.06)	25.56
28	35.85	(21.76, 54.15)	18.74

1991.

Figure B.9 gives the posterior densities for $\omega_2, ..., \omega_{23}$, the effects due to stratum on hoki catch per unit effort in 22 of the strata with respect to stratum 1 (the reference level). Again, we see resemblance to the densities observed for the stratum effects in Models 2 and 3. As we saw for Models 2 and 3, we observe differences in the stratum effects, most noticeable between strata 26, 27, 28, and all others. In addition, the same stratum clustering observed in Model 2 is evident here.

All of the stratum effects have the majority of their densities below zero, indicating that CPUE was lower in strata 2 to 28, compared to stratum 1. The model suggests that strata 26, 27, and 28 had much lower CPUE compared to all others. Strata 27 and 28 are neighbouring, suggesting that hoki catches in the north eastern area of the sub-Antarctic region is lower than near Puysegur bank.

Model predicted CPUE for each year ($\widehat{\text{CPUE}}_t$) and each stratum ($\widehat{\text{CPUE}}_w$) for each iteration in the HMC chains were calculated by aggregating over strata and year respectively:

$$\widehat{\text{CPUE}}_t = \sum_{w=1}^{23} \widehat{\text{CPUE}}_{t,w}, \tag{5.6}$$

$$\widehat{\text{CPUE}}_w = \sum_{t=1}^{12} \widehat{\text{CPUE}}_{t,w}, \tag{5.7}$$

where

$$\widehat{\text{CPUE}}_{t,w} = \begin{cases} \exp(\hat{\mu}) & \text{for } t = 1, w = 1 \\ \exp(\hat{\mu} + \hat{\omega}_w) & \text{for } t = 1, w = 2, ..., 23 \\ \exp(\hat{\mu} + \hat{\tau}_t) & \text{for } t = 2, ..., 12, w = 1 \\ \exp(\hat{\mu} + \hat{\tau}_t + \hat{\omega}_w) & \text{for } t = 2, ..., 12, w = 2, ..., 23 \end{cases} \tag{5.8}$$

The model predicted CPUE distributions and the observed median CPUE for each trawl year are shown in Figure 5.8 and summarised in Table 5.9.

Table 5.8: Summary of the posterior distributions for the parameters in Model 4.

	Median	95% HPD		Median	95% HPD
μ	6.90	(6.60, 7.17)	ω_2	-0.83	(-1.22, -0.42)
τ_2	0.14	(-0.11, 0.39)	ω_3	-1.24	(-1.90, -0.57)
τ_3	0.18	(-0.07, 0.42)	ω_4	-1.08	(-1.49, -0.65)
τ_4	-0.70	(-0.96, -0.41)	ω_5	-0.58	(-0.95, -0.24)
τ_5	-0.87	(-1.15, -0.59)	ω_6	-1.56	(-1.94, -1.18)
τ_6	-1.09	(-1.36, -0.81)	ω_7	-1.35	(-1.94, -0.71)
τ_7	-1.56	(-1.86, -1.26)	ω_8	-1.27	(-1.60, -0.93)
τ_8	-1.26	(-1.54, -0.97)	ω_9	-1.35	(-1.69, -0.99)
τ_9	-1.32	(-1.60, -1.04)	ω_{10}	-1.40	(-1.81, -0.97)
τ_{10}	-1.46	(-1.75, -1.17)	ω_{11}	-1.32	(-1.79, -0.87)
τ_{11}	-0.64	(-0.92, -0.37)	ω_{12}	-1.16	(-1.54, -0.79)
τ_{12}	-0.62	(-0.90, -0.35)	ω_{13}	-0.96	(-1.33, -0.58)
ρ	0.83	(0.62, 0.99)	ω_{14}	-1.62	(-1.94, -1.30)
ϕ	1.88	(0.19, 4.21)	ω_{15}	-1.36	(-1.66, -1.04)
η	1.05	(0.39, 1.99)	ω_{16}	-1.65	(-1.94, -1.33)
p	0.99	(0.94, 1.00)	ω_{17}	-1.46	(-1.83, -1.13)
κ	0.21	(-0.37, 0.77)	ω_{18}	-1.73	(-2.07, -1.35)
δ	9.02	(8.34, 9.78)	ω_{19}	-1.51	(-1.91, -1.10)
			ω_{20}	-1.68	(-2.06, -1.31)
			ω_{21}	-2.73	(-3.86, -1.69)
			ω_{22}	-2.72	(-3.23, -2.18)
			ω_{23}	-2.64	(-3.23, -2.03)

Figure 5.8: The posterior distributions of CPUE by trawl year, as predicted by Model 4. Here, the predictions were aggregated over strata. Vertical lines represent the observed CPUE.

Figure 5.9: The posterior distributions of log CPUE by stratum, as predicted by Model 4. Here, the predictions were aggregated over trawl year. Vertical lines represent the observed log CPUE.

Table 5.9: The median CPUE for each year, as predicted by Model 4 and observed in the dataset.

Year	Predicted	95% HPD	Observed
1991	283.33	(231.24, 339.99)	311.38
1992	325.24	(262.15, 389.66)	305.44
1993	338.09	(276.63, 405.88)	276.19
2000	141.34	(110.43, 174.47)	134.09
2001	118.50	(93.39, 146.76)	106.82
2002	95.32	(76.12, 116.90)	96.48
2003	59.46	(46.01, 74.96)	52.92
2004	80.46	(63.37, 99.64)	62.35
2005	75.90	(60.51, 93.33)	63.10
2006	66.07	(51.02, 81.75)	66.57
2007	149.02	(118.87, 185.32)	140.81
2008	152.18	(121.41, 185.14)	152.72

We can see that densities overlap with the observed median CPUE for some trawl years. The model appears to predict median CPUE for trawl years 2000, 2002, 2006, 2007, and 2008. This is indicated by the fact that the observed CPUE values are closer to centre of the densities in those years. Compared to Models 1 and 3, the predictions have become worse in general. This suggests that the model has not improved the fit to the data.

The model predicted log CPUE distributions and log observed CPUE for each stratum are shown in Figure 5.9 and summarised in Table 5.10. We see that stratum does not predict the observed log CPUE very well. For most strata, there is no overlap between the log CPUE densities and observed log CPUE medians. This suggests that stratum does not explain much variation in hoki CPUE.

In general, the difference in observed and predicted median CPUE is larger

Table 5.10: The median CPUE for each stratum, as predicted by Model 4 and observed in the dataset.

Stratum	Predicted	95% HPD	Observed
1	549.45	(419.83, 681.81)	457.88
2	240.25	(169.82, 330.73)	206.37
3	159.56	(75.28, 276.03)	307.90
003A	186.08	(128.86, 257.64)	134.32
003B	308.83	(228.21, 410.86)	208.25
4	115.27	(82.85, 152.56)	133.46
5	141.89	(73.63, 232.47)	295.18
005A	154.05	(116.70, 195.74)	85.47
005B	143.10	(108.52, 181.92)	94.06
6	135.50	(90.27, 184.83)	124.55
7	146.13	(93.60, 211.57)	156.25
8	171.33	(124.87, 226.58)	205.57
9	210.37	(150.85, 278.90)	213.20
10	108.83	(84.85, 134.92)	139.98
11	141.08	(113.95, 172.91)	191.34
12	105.47	(86.32, 128.38)	124.19
13	127.64	(94.88, 162.27)	175.25
14	97.89	(74.03, 125.74)	195.64
15	121.31	(84.25, 165.10)	128.98
25	102.77	(76.78, 135.48)	77.83
26	35.78	(7.93, 82.97)	7.68
27	36.23	(20.55, 54.74)	25.56
28	39.44	(21.10, 64.82)	18.74

by stratum than it is by year. This suggests that trawl year accounts for a little more variation in hoki CPUE than stratum.

Figure B.7 displays the posterior densities of the rest of the parameters in Model 2. The posterior density for μ represents the model predicted log CPUE for the reference year, 1991, and reference stratum, 1.

The posterior distribution for ρ estimates the correlation in hoki log CPUE between successive trawl years. The 95% highest posterior density interval for the posterior density of ρ is $(0.62, 0.99)$, and the posterior median is 0.83 which, suggests there is strong and positive year-to-year autocorrelation within hoki CPUE and is captured by the model.

The parameter ϕ gives a measure of the total variation of mean CPUE between trawl years. The 95% HPD interval for ϕ is $(0.19, 4.21)$.

The posterior distribution of p has 95% of its highest density between 0.94 and 1. This suggests that there is a strong association between the mean CPUEs of each stratum. The value of η gives a measure of the total variation of mean CPUE between strata. The 95% HPD interval for η is $(0.39, 1.99)$.

The posterior distribution of κ has 95% of its highest density between -0.37 and 0.77, with a median of 0.21. This suggests that there is a weak association between each pair of CPUE observations. The value of δ gives a measure of the precision in hoki CPUE. The 95% HPD interval for δ is $(8.34, 9.78)$.

Residuals were calculated by taking the median of $(\log \text{CPUE} - \log \widehat{\text{CPUE}}_{t,w})$, where $\widehat{\text{CPUE}}_{t,w}$ is the model predicted CPUE for all iterations in the HMC chain for for year t, and stratum w.

Figure 5.10 displays a plot of the residuals between the model predicted log CPUE and the observed log CPUE. We see a wide spread of residuals, with some very slight curvature. This suggests that some of the variation in CPUE is explained by trawl year and stratum, but there is some variation left unexplained. This model appears to explain less variation in hoki CPUE compared to Models 1 and 3.
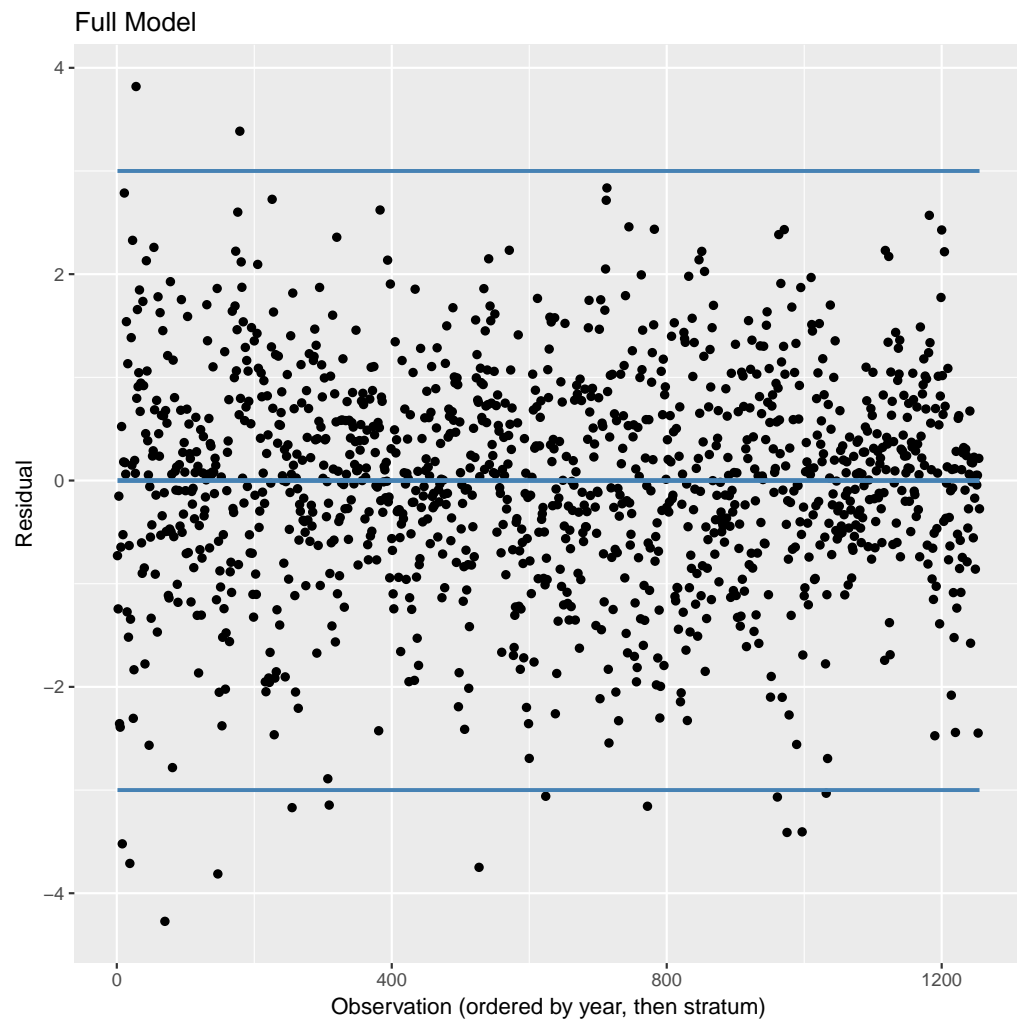
Figure 5.10: Residuals from Model 4, ordered by trawl year.

5.5   Model 5: Blocked Point Referenced CPUE Model

Posterior densities for the parameters in Model 5 (defined in Table 4.1) are presented in Figures B.11 to B.15. Tables 5.11 and 5.12 summarise these posterior parameter distributions.

Figure B.11 gives the posterior densities of $\tau_2, ..., \tau_{12}$, which are the temporal effects on hoki catch per unit effort for the years $1992 - 2008$ with respect to the year 1991 (the reference level for the temporal effects). The distributions of the effects due to trawl year for Model 5 are similar to the ones in Model 1, Model 3, and Model 4. We observe a noticeable difference between the effects of each trawl year. The posterior distribution of the effect for 1993 has most of its density above zero suggesting that CPUE was higher in this years than it was in 1991. Furthermore, the effects for the later trawls in years $2000 - 2008$ have posterior distributions centred below zero suggesting that CPUE was lower in these years compared to in 1991. The effect for the year 1992 appears to be centred around zero. This indicates that 1992 had a similar effect on hoki CPUE to the year 1991.

Figure B.12 gives the posterior densities for $\omega_2, ..., \omega_{23}$, the effects due to stratum on hoki catch per unit effort in 22 of the strata with respect to stratum 1 (the reference level). Again, we see resemblance to the densities observed for the stratum effects in Models 2, 3, and 4. As we saw for Models 2, 3, and 4, we observe differences in the stratum effects, most noticeable between strata 26, 27, 28, and all others. However, these differences are not as extreme compared to the previous models.

Most of the stratum effects have their densities centred on zero, indicating that stratum has little effect on hoki CPUE in model 5. The model suggests that strata 26, 27, and 28 had much lower CPUE compared to all others. Strata 27 and 28 are neighbouring, suggesting that hoki catches in the north eastern area of the sub-Antarctic region is lower than near Puysegur bank.

Model predicted CPUE for each year ($\widehat{\text{CPUE}}_t$) and each stratum ($\widehat{\text{CPUE}}_w$)

for each iteration in the HMC chains were calculated by aggregating over strata and year respectively:

$$\widehat{\mathrm{CPUE}}_t = \sum_{w=1}^{23} \widehat{\mathrm{CPUE}}_{t,w}, \tag{5.9}$$

$$\widehat{\mathrm{CPUE}}_w = \sum_{t=1}^{12} \widehat{\mathrm{CPUE}}_{t,w}, \tag{5.10}$$

where

$$\widehat{\mathrm{CPUE}}_{t,w} = \begin{cases} \exp(\hat{\mu}) & \text{for } t = 1, w = 1 \\ \exp(\hat{\mu} + \hat{\omega}_w) & \text{for } t = 1, w = 2, ..., 23 \\ \exp(\hat{\mu} + \hat{\tau}_t) & \text{for } t = 2, ..., 12, w = 1 \\ \exp(\hat{\mu} + \hat{\tau}_t + \hat{\omega}_w) & \text{for } t = 2, ..., 12, w = 2, ..., 23 \end{cases} \tag{5.11}$$

The model predicted CPUE distributions and the observed median CPUE for each trawl year are shown in Figure 5.11 and summarised in Table 5.13. We can see that densities overlap with the observed median CPUE for some trawl years. The model appears to predict median CPUE for trawl years 1991, 2002, 2003, 2005, 2006, and 2008. This is indicated by the fact that the observed CPUE values are closer to centre of the densities in those years. Compared to Models 1 and 3, the predictions have become worse in general. However, there appears to be an improvement compared to Model 4.

The model predicted log CPUE distributions and log observed CPUE for each stratum are shown in Figure 5.11 and summarised in Table 5.14. We see that stratum does not predict the observed log CPUE very well. For most strata, there is no overlap between the log CPUE densities and observed log CPUE medians. This suggests that stratum does not explain much variation in hoki CPUE in Model 5.

In general, the difference in observed and predicted median CPUE is larger by stratum than it is by year. This suggests that trawl year accounts for a little more variation in hoki CPUE than stratum in Model 5.

Table 5.11: Summary of the posterior distributions for the parameters in Model 5.

	Median	HPD		Median	95% HPD
τ_2	0.02	(-0.17, 0.20)	ω_2	0.54	(-0.20, 1.04)
τ_3	0.21	(0.01, 0.42)	ω_3	0.13	(-0.86, 0.89)
τ_4	-0.63	(-0.85, -0.43)	ω_4	-0.11	(-1.01, 0.63)
τ_5	-0.95	(-1.18, -0.71)	ω_5	0.51	(-0.33, 1.18)
τ_6	-1.09	(-1.33, -0.85)	ω_6	-0.23	(-1.10, 0.36)
τ_7	-1.79	(-2.04, -1.53)	ω_7	-0.22	(-1.08, 0.43)
τ_8	-1.77	(-2.02, -1.53)	ω_8	-0.12	(-0.88, 0.37)
τ_9	-1.58	(-1.83, -1.31)	ω_9	-0.14	(-0.71, 0.37)
τ_{10}	-1.53	(-1.82, -1.28)	ω_{10}	-0.25	(-0.98, 0.24)
τ_{11}	-0.92	(-1.18, -0.67)	ω_{11}	-0.07	(-0.88, 0.45)
τ_{12}	-0.72	(-1.00, -0.46)	ω_{12}	-0.06	(-0.79, 0.34)
μ	5.82	(5.40, 6.53)	ω_{13}	0.24	(-0.45, 0.70)
ρ	0.86	(0.69, 0.99)	ω_{14}	-0.46	(-1.21, 0.08)
ϕ	1.57	(0.10, 3.56)	ω_{15}	-0.03	(-0.81, 0.43)
η	1.29	(0.34, 3.07)	ω_{16}	-0.40	(-1.12, 0.02)
p	0.63	(0.09, 1.00)	ω_{17}	-0.33	(-1.07, 0.10)
			ω_{18}	-0.03	(-0.78, 0.45)
			ω_{19}	0.41	(-0.39, 1.01)
			ω_{20}	-0.41	(-1.16, 0.15)
			ω_{21}	-1.55	(-2.83, -0.35)
			ω_{22}	-1.32	(-2.34, -0.49)
			ω_{23}	-1.26	(-2.32, -0.38)

Table 5.12: Summary of the posterior distributions for the parameters in Model 5.

	Median	95% HPD		Median	95% HPD
κ_1	0.44	(-0.31, 0.99)	δ_1	0.45	(0.30, 0.63)
κ_2	0.00	(-0.81, 0.75)	δ_2	1.29	(0.84, 1.88)
κ_3	0.66	(-0.32, 1.00)	δ_3	2.22	(0.71, 4.16)
κ_4	0.89	(0.63, 1.00)	δ_4	13.45	(7.75, 20.58)
κ_5	0.85	(0.41, 1.00)	δ_5	1.20	(0.62, 1.87)
κ_6	0.07	(-0.85, 0.95)	δ_6	0.71	(0.48, 0.96)
κ_7	0.34	(-0.48, 0.97)	δ_7	5.03	(2.04, 9.01)
κ_8	-0.00	(-0.82, 0.71)	δ_8	2.42	(1.46, 3.57)
κ_9	-0.01	(-0.75, 0.75)	δ_9	18.25	(10.56, 27.20)
κ_{10}	-0.22	(-0.97, 0.53)	δ_{10}	4.31	(2.86, 5.96)
κ_{11}	0.67	(0.06, 1.00)	δ_{11}	2.08	(1.26, 3.00)
κ_{12}	-0.08	(-1.00, 0.74)	δ_{12}	0.66	(0.48, 0.86)
κ_{13}	0.98	(0.92, 1.00)	δ_{13}	2.03	(1.49, 2.57)
κ_{14}	-0.78	(-1.00, -0.19)	δ_{14}	2.16	(1.40, 3.04)
κ_{15}	0.20	(-0.76, 0.97)	δ_{15}	0.84	(0.55, 1.16)
κ_{16}	0.77	(-0.40, 1.00)	δ_{16}	0.43	(0.32, 0.55)
κ_{17}	0.01	(-0.91, 0.80)	δ_{17}	1.11	(0.80, 1.47)
κ_{18}	-0.01	(-0.90, 0.82)	δ_{18}	1.16	(0.82, 1.54)
κ_{19}	-0.40	(-1.00, 0.55)	δ_{19}	3.09	(2.01, 4.47)
κ_{20}	0.09	(-0.78, 0.90)	δ_{20}	4.56	(3.25, 6.15)
κ_{21}	0.09	(-0.84, 0.97)	δ_{21}	1.53	(0.19, 4.03)
κ_{22}	0.05	(-0.91, 0.84)	δ_{22}	2.12	(1.13, 3.25)
κ_{23}	0.36	(-0.62, 0.99)	δ_{23}	1.49	(0.66, 2.51)

Figure 5.11: The posterior distributions of CPUE by trawl year, as predicted by Model 5. Here, the predictions were aggregated over strata. Vertical lines represent the observed CPUE.
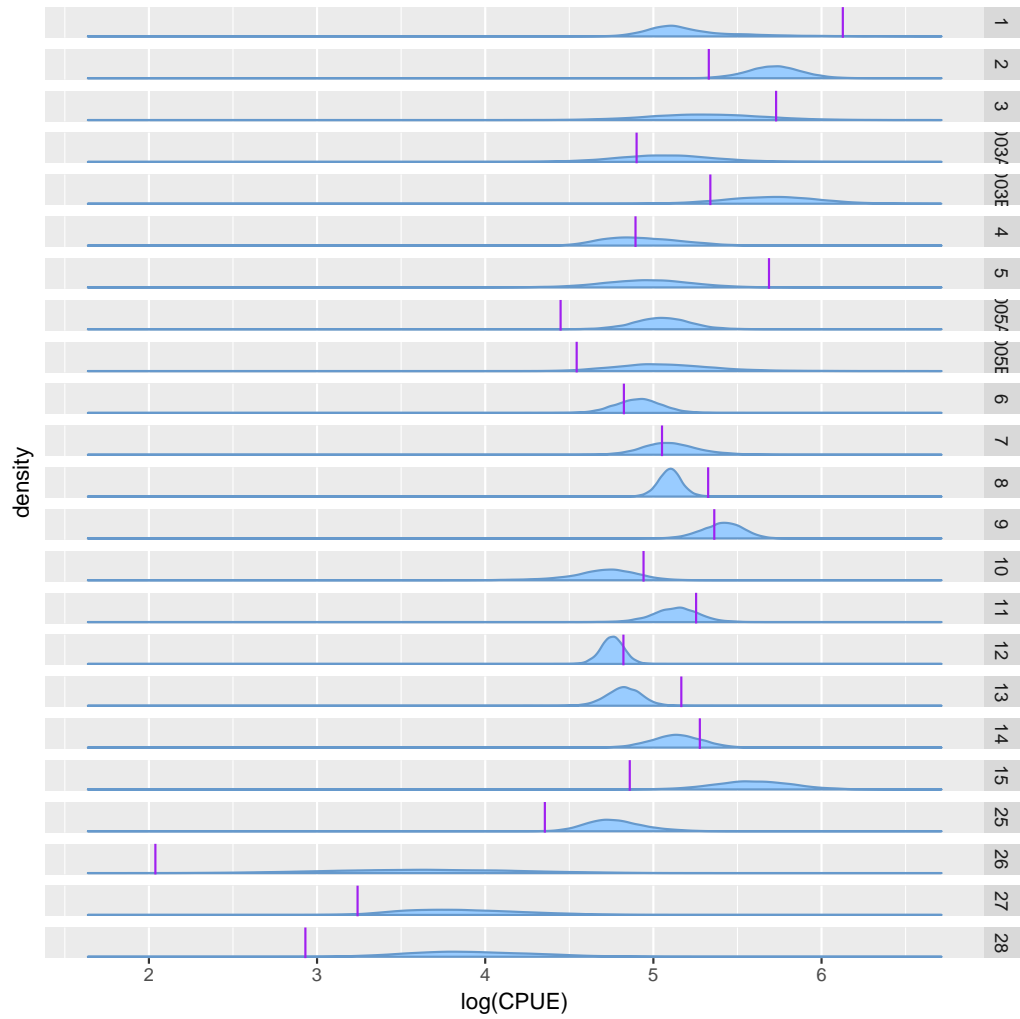
Figure 5.12: The posterior distributions of log CPUE by stratum, as predicted by Model 5. Here, the predictions were aggregated over trawl year. Vertical lines represent the observed log CPUE.

Table 5.13: The median CPUE for each year, as predicted by Model 5 and observed in the dataset

Year	Predicted	95% HPD	Observed
1991	315.85	(261.09, 376.89)	311.38
1992	321.29	(266.38, 383.89)	305.44
1993	388.80	(314.99, 474.19)	276.19
2000	167.67	(135.51, 208.12)	134.09
2001	122.59	(97.96, 152.98)	106.82
2002	105.67	(82.71, 133.32)	96.48
2003	52.87	(40.62, 67.48)	52.92
2004	53.45	(41.46, 67.66)	62.35
2005	65.33	(50.59, 82.62)	63.10
2006	68.11	(51.65, 85.80)	66.57
2007	125.53	(98.26, 157.38)	140.81
2008	153.73	(118.34, 196.74)	152.72

Figure B.15 displays the posterior densities of the rest of the parameters in Model 2. The posterior density for μ represents the model predicted log CPUE for the reference year, 1991, and reference stratum, 1.

The posterior distribution for ρ estimates the correlation in hoki log CPUE between successive trawl years. The 95% highest posterior density interval for the posterior density of ρ is $(0.69, 0.99)$, and the posterior median is 0.86 which, suggests there is strong and positive year-to-year autocorrelation within hoki CPUE and is captured by the model.

The parameter ϕ gives a measure of the total variation of mean CPUE between trawl years. The 95% HPD interval for ϕ is (0.10, 3.56).

The posterior distribution of p has 95% of its highest density between 0.09 and 1. This suggests that there is weak association between the mean CPUEs of each stratum. This suggests that any spatial correlation due to strata has already been captured by κ. The value of η gives a measure of

Table 5.14: The median CPUE for each stratum, as predicted by Model 5 and observed in the dataset

Stratum	Predicted	95% HPD	Observed
1	172.11	(110.14, 337.91)	457.88
2	304.66	(218.67, 407.46)	206.37
3	201.94	(93.37, 362.75)	307.90
003A	159.01	(77.50, 266.32)	134.32
003B	300.98	(156.94, 482.92)	208.25
4	135.92	(90.76, 209.54)	133.46
5	143.12	(74.97, 224.26)	295.18
005A	156.13	(108.75, 211.37)	85.47
005B	156.06	(85.38, 271.93)	94.06
6	136.02	(102.68, 174.71)	124.55
7	162.89	(115.76, 225.99)	156.25
8	163.98	(141.1, 187.42)	205.57
9	225.15	(174.68, 279.00)	213.20
10	112.43	(70.57, 156.71)	139.98
11	170.28	(127.06, 218.50)	191.34
12	116.39	(100.73, 132.25)	124.19
13	124.70	(99.43, 149.44)	175.25
14	170.41	(122.92, 222.98)	195.64
15	268.77	(163.53, 408.40)	128.98
25	116.58	(84.66, 164.73)	77.83
26	37.66	(9.29, 93.96)	7.68
27	46.37	(23.81, 90.62)	25.56
28	49.43	(22.86, 96.81)	18.74

the total variation of mean CPUE between strata. The 95% HPD interval for η is (0.34, 3.07).

The posterior distributions of $\boldsymbol{\kappa}$ displayed in Figure B.13 as a boxplot. The values of κ_j give the spatial correlations between each pair of hoki CPUE observations, within each strata. We see asymmetry, and differences across each κ_j ($j = 1, ..., 23$).

The values of $\boldsymbol{\delta}$ give measures of the precision in hoki CPUE within each strata. The densities for each δ_j are shown the vary across strata.

Residuals were calculated by taking the median of ($\log \text{CPUE} - \log \widehat{\textbf{CPUE}}_{t,w}$), where $\widehat{\textbf{CPUE}}_{t,w}$ is the model predicted CPUE for all iterations in the HMC chain for for year t, and stratum w.

Figure 5.13 displays a plot of the residuals between the model predicted log CPUE and the observed log CPUE. We see a somewhat narrower spread of residuals, with some curvature, and potential outliers in the earlier trawl years. This suggests that some of the variation in CPUE is explained by trawl year and stratum, but there is quite a bit of variation left unexplained. This model appears to explain less variation in hoki CPUE compared to Models 1, 3, and 4, but more than possibly more than Model 2.

5.6   Model Comparison

WAIC and DIC were calculated for Models 1, 2, and, 3. The values are shown below in Table 5.15, along with log posterior density values (lpd) and computed effective numbers of parameters, p. An information criterion could not be applied to Models 4 and 5, because they involve partitioning of the data. In Models 4 and 5, we assumed that hoki CPUE observations were structured, where each pairwise set of observations was assumed to be correlated. DIC and WAIC would break these dependencies.

From Table 5.15, we see that the model with the lowest DIC and WAIC values is Model 3. The model with the highest values is Model 2. We can
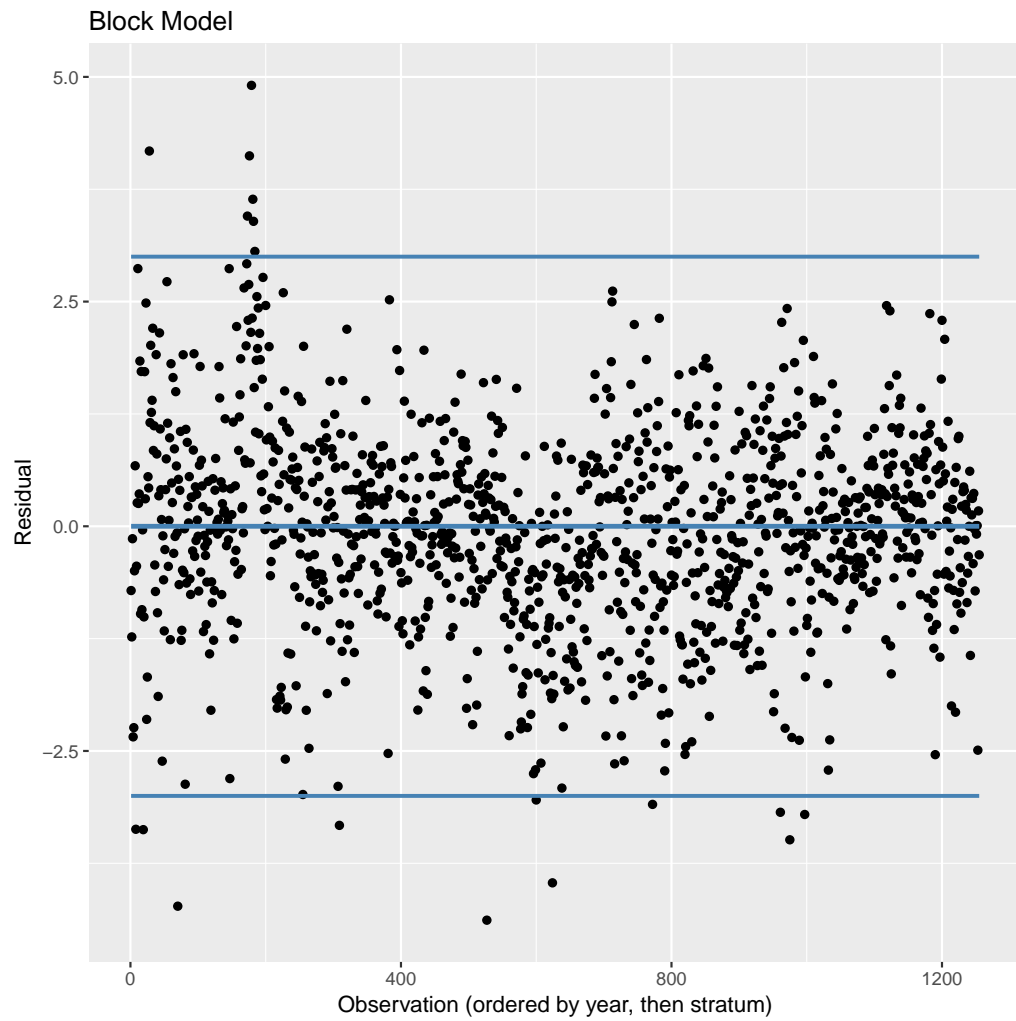
Figure 5.13: Residuals from Model 5, ordered by trawl year.

Table 5.15: Deviance information criteria and Watanabe-Akaike information criteria for Models 1, 2, and 3.

Model	lpd	p_{DIC}	DIC	p_{WAIC}	WAIC
1	-1987	15.3	4037	12.7	4000
2	-2029	31.0	4181	22.4	4103
3	-1841	41.3	3848	34.2	3750

conclude that Model 3 performed the best, providing the best fit to the data compared to Models 1 and 2. Model 2 provided the worst fit. This agrees with our visual assessment of the models, using the residual plots. We now move on to Chapter 6, where we discuss the results presented in this chapter.

Chapter 6

Discussion and Conclusion

The aim of this research was to improve the precision and accuracy of hoki abundance indices by accounting for spatial and temporal autocorrelation within CPUE using Gaussian Markov random fields (GMRFs). In this thesis, we built five models that attempted to explain spatial and temporal variation in hoki CPUE. Trawl year and stratum were modelled using GMRFs with various precision structures. In this Chapter, we discuss the main results outlined in Chapter 5. We start by comparing trawl year and stratum in their abilities to account for temporal and spatial variation respectively. This leads on to discussion of the survey design. We then debate whether depth interval would explain more spatial variation than stratum. Model 3 is then compared to both Models 1 and 2. We then compare and contrast the areal approach with the point referenced approach. Models 4 and 5 are discussed and we debate which dependency structure was more appropriate given the survey design. We conclude the discussion by listing future research considerations after this thesis.

By comparing the plots of residuals for Model 1 (Figure 5.2) to Model 2 (Figure 5.4), we see that trawl year explained much more variation than stratum. This is not surprising when one considers the survey design. The 2-phase random stratification survey design proposed by Francis (1984) attempts to minimise the overall coefficient of variation within hoki catch

weight. This was achieved by allocating more stations to the strata with the most varied catch observations (see Section 2.2). As a result, most of the variation due to strata has been accounted for by sampling design.

A similar result was obtained in Tian et al. (2009). In Tian et al. (2009), survey CPUE of neon flying squid (*Ommastrephes bartramii*) was modelled using both a generalized linear model, and a generalized additive model. In both cases, it was found that the month of trawl explained more variation in CPUE than any other factor (including spatial effects). This result agrees with our conclusion that location factors are less important than time factors when modelling CPUE from a well-designed survey.

The stratum CPUE predictions we made in Chapter 5 were imprecise for Models 2 – 5. This is illustrated in each model's section by the density plots of CPUE by stratum (Figures 5.3, 5.6, 5.9, and 5.12). Small stratum sizes can contribute to the imprecision of CPUE predictions. Table 2.4 in Chapter 2 displays a summary of hoki CPUE by stratum. We can see that the number of trawls with positive catches were low for some strata (for example: 13 trawls in stratum 3; 16 trawls in stratum 5; and only 4 trawls in stratum 26). For comparison, the stratum with the largest number of positive trawls was stratum 12 with 121 trawls. This could have contributed to the imprecise CPUE predictions for strata. The problem of estimating CPUE for strata with few observations was also encountered in Shelton et al. (2014). We propose that the problem could be resolved by amalgamating some strata.

The issue of small numbers of observations in strata could be addressed to improve CPUE prediction by using a different measure of spatial distribution, such as depth intervals. As discussed in Section 2.1, stratification in the summer research trawl surveys was based on depth intervals of 300 – 600 m, 600 – 800 m, and 800 – 1000 m. By replacing the stratum factor with depth interval in our models, we would decrease the number of estimated parameters and increase the number of observations within each group. This may improve the precision of CPUE prediction, and possibly explain

more spatial variation in hoki CPUE, than stratum. Francis et al. (2002) and Fujita (2016) found that depth was a statistically significant predictor of CPUE for species on the Chatam Rise. Depth interval could be explored with further research.

Model 3 was a combination of both Models 1 and 2, in that it included both trawl year and stratum as factors to account for temporal and spatial variation in hoki CPUE. Table 5.15 shows that Model 3 had the lowest DIC and WAIC value of the first three models. We therefore conclude that a model with both factors included explains more variation than a model with only one of these factors. Even though stratum explains much less variation in hoki CPUE, it still contributes to Model 3. This suggests that trawl year alone does not account for all the variation in hoki CPUE collected from scientific surveys; a model including stratum as an additional factor is more suitable.

Results from Models 3 and 4 can be compared to investigate the differences between the areal approach and point referenced approach to account for spatial and temporal variation. We hypothesised that modelling point referenced CPUE would lead to more precise predictions than with an areal approach. However, our results indicate that this was not the case. The first noticeable difference can be seen in the plots of predicted CPUE by year and by strata (Figures 5.5 and 5.6 for Model 3, and Figures 5.8 and 5.9 for Model 4). The precision and accuracy of Model 4 predictions appear to be lower compared to that of Model 3. The density plots of predicted CPUE for Model 4 are more spread than those of Model 3. In addition, the Model 4 predicted CPUE densities have less overlap with observed CPUE values than can be seen for Model 3. The dependence structure we assumed for CPUE in Model 4 may not have been appropriate.

Model 4 assumed that each pair of hoki CPUE observations were spatially correlated depending on their distance. The spatial correlation also depended on two parameters: a correlation parameter κ; and a precision parameter δ. The parameters κ and δ were kept constant across each pair of

hoki CPUE observations. This meant that a pair of observations within one stratum had the same covariance as a pair of observations in another stratum, provided the each pair of observations were the same distance apart. Model 5 assumed that each pair of hoki CPUE observations were associated within a stratum, but independent between them. In other words, hoki CPUE had a different dependence structure within each stratum. We hypothesized that by modelling spatial autocorrelation of hoki CPUE separately for each stratum, we would arrive at more precise CPUE predictions compared to Model 4. By comparing the residual plots of Model 4 (Figure 5.10) and Model 5 (Figure 5.13), we can conclude that Model 5 provided a worse fit to the data, thus disproving our hypothesis. However, this conclusion may be invalid, since we have serious reservations about parameter convergence in Model 5 (see Section 4.2.1, and the diagnostic plots in the appendices). By running Model 5 for more iterations, we may see it providing a better fit. This should be considered for future research.

We compared our results to the analysis on commercial scallop (*Pecten maximus*) CPUE performed by Murray et al. (2013). Murray et al. (2013) modelled scallop CPUE using commercial fisheries data from the Isle of Man. They constructed a series of generalized linear models that included factors to account for differences in vessels, location, and time. Models that contained a location factor explained more variation in scallop CPUE than models with a time factor. From our models, we found that trawl year explained more variation in hoki CPUE than stratum. We deduce that the difference between our results and the results of Murray et al. (2013) is due to survey design. As discussed previously, our survey was designed to account for most of the spatial variation in hoki CPUE. As such, any stratum effect we include would not add much explanatory power to model of survey CPUE. However, commercial data is not sampled in the same way as survey data. Commercial data is "presence-only" data, meaning they are collected in areas and at times when fish are available. This does not account for spatial variation in the same our survey does. As such, loca-

tion factors can explain much more spatial variation in commercial CPUE compared to survey CPUE.

In this thesis, we found that both trawl year and stratum (to a lesser extent) are important factors to be considered when modelling hoki CPUE. By modelling both factors using GMRFs, we were able to explain a significant amount of spatial and temporal variation in hoki CPUE. Furthermore, we found that using the areal approach to model hoki CPUE was better in that it accounted for more spatial and temporal variation than the point referenced approach. We illustrated that using an appropriate and well designed survey can significantly account for spatial variation without the need for including a spatial effect.

In addition to the results achieved in this thesis, there were a multitude of areas where improvement could be made. Future research considerations include:

- Running the HMC chains of Models 4 and 5 for more iterations. Conclusions drawn from the current Model 5 results 5 are likely to be incorrect due to the lack of convergence as seen in the diagnostic plots.

- Replacing stratum with depth interval as a different form of spatial effect in the hoki CPUE models. Our hypothesis is that a model with depth interval will explain more spatial variation in hoki CPUE than a model with stratum.

- Repeating the analysis for a fish species that the survey was not designed for. We hypothesise that stratum would explain more spatial variation in the CPUE of another species. This is because the survey will not account for as much spatial variation as it does for hoki.

- Repeating the analysis using catch and effort data from commercial fisheries, where other factors (for example, vessel and gear effects) affect CPUE. We would be able to investigate how much variation time and space explain relative to other factors.

- Investigating the use of stochastic partial differential equations (SPDE) for the approximation of Gaussian random fields, as opposed to the conditionally autoregressive GMRFs we used in this thesis. Gaussian random fields allow for modelling in continuous space, which may be more appropriate for point reference data (Lindgren et al., 2011)

- Investigating how to account for the interaction of spatial and temporal autocorrelation. In Chapter 2, we found evidence for the existence of spatio-temporal autocorrelation within hoki CPUE. With more data (for example, commercial data), we could implement a GMRF to model the spatio-temporal autocorrelation in hoki CPUE. This has been researched by Thorson et al. (2015).

- Investigating pivotal discrepancy measures (Yuan & Johnson, 2012) and other model comparison techniques to compare structured Bayesian hierarchical models. We could not compute an information criterion for Models 4 and 5. This was due to the dependence structure on hoki CPUE that both models assumed. Pivotal discrepancy measures can be used to compare hierarchical models, however, it is still unknown whether they can be extended to models with structured dependencies.

Appendix A

Trace Plots

A.1   Model 1: Temporal Effects
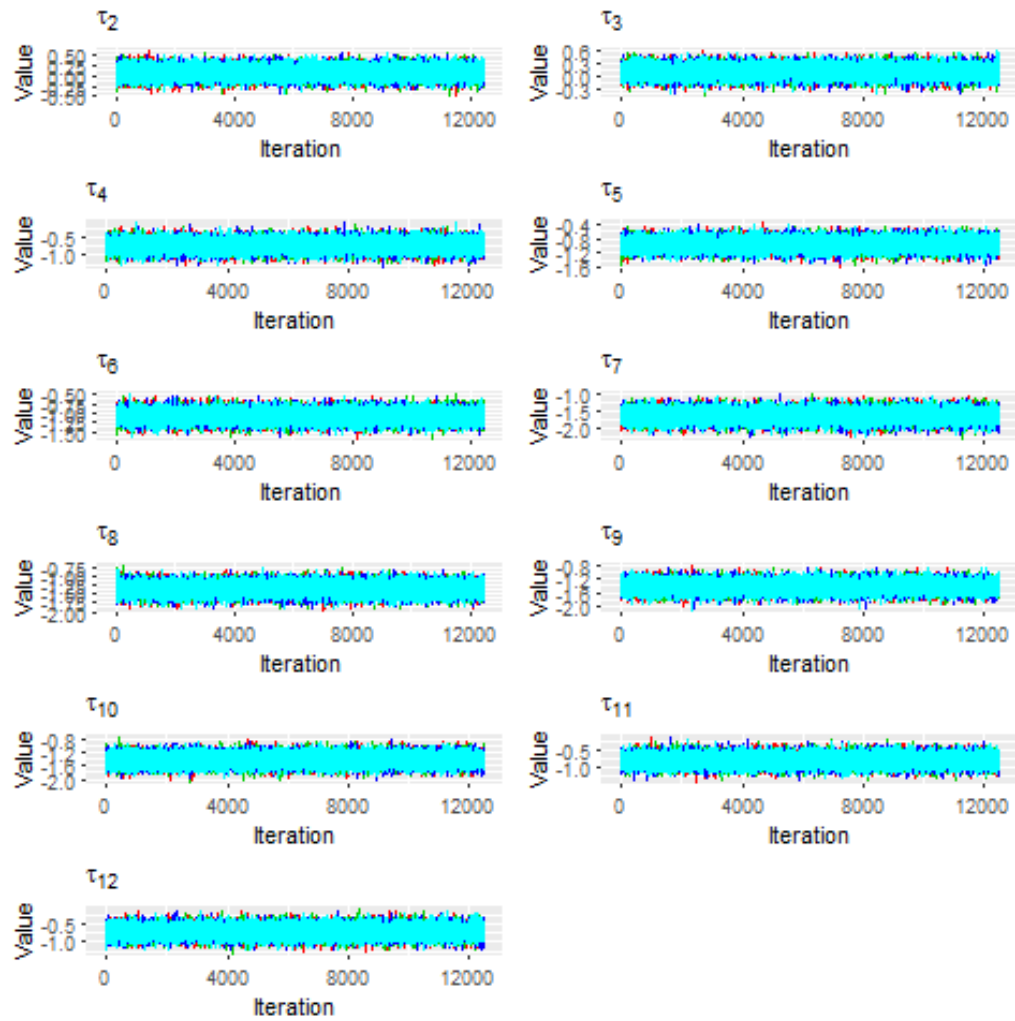
Figure A.1: Trace plots for the posterior distributions of $\tau_2, ..., \tau_{12}$ in Model 1.
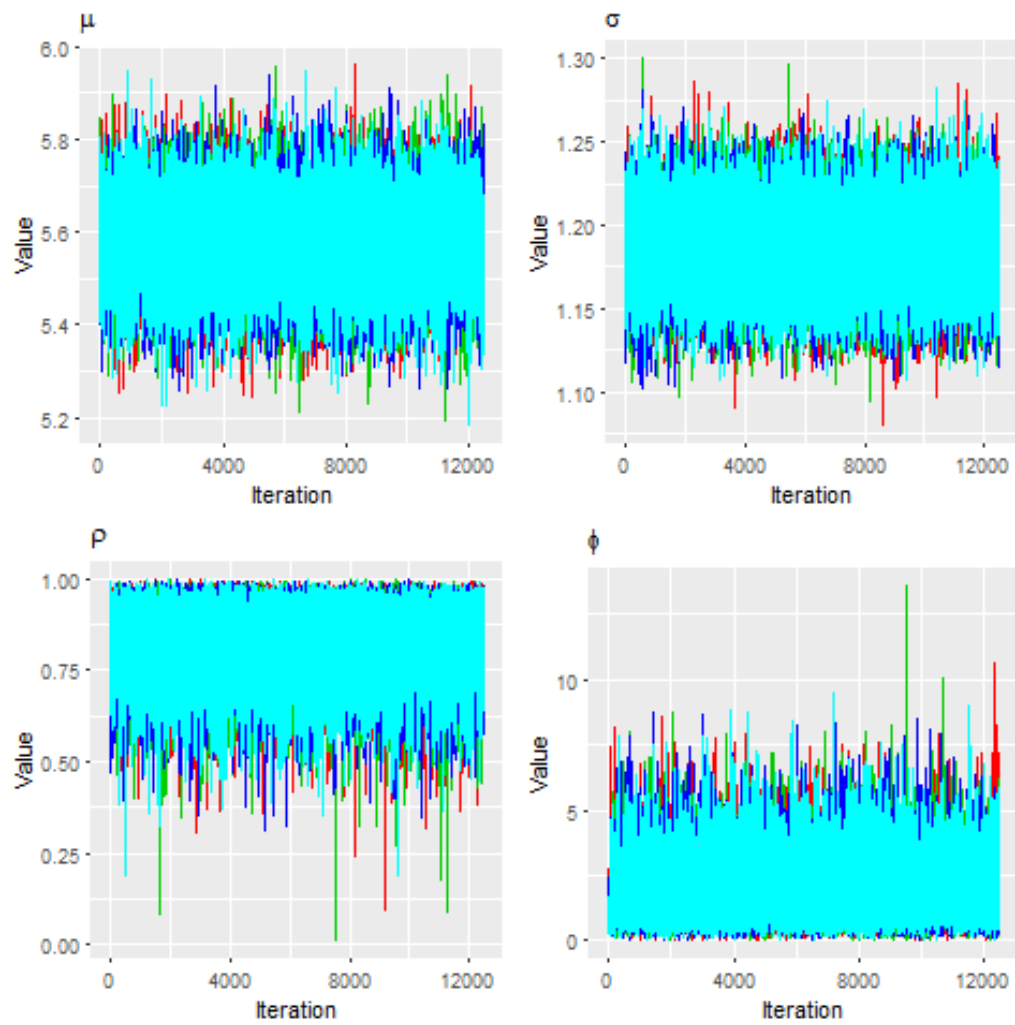
Figure A.2: Trace plots for the posterior distribution of μ, σ, ϕ, and ρ in Model 1

A.2   Model 2: Areal Spatial Effects

Figure A.3: Trace plots for the posterior distributions of $\omega_2, ..., \omega_{23}$ in Model 2.
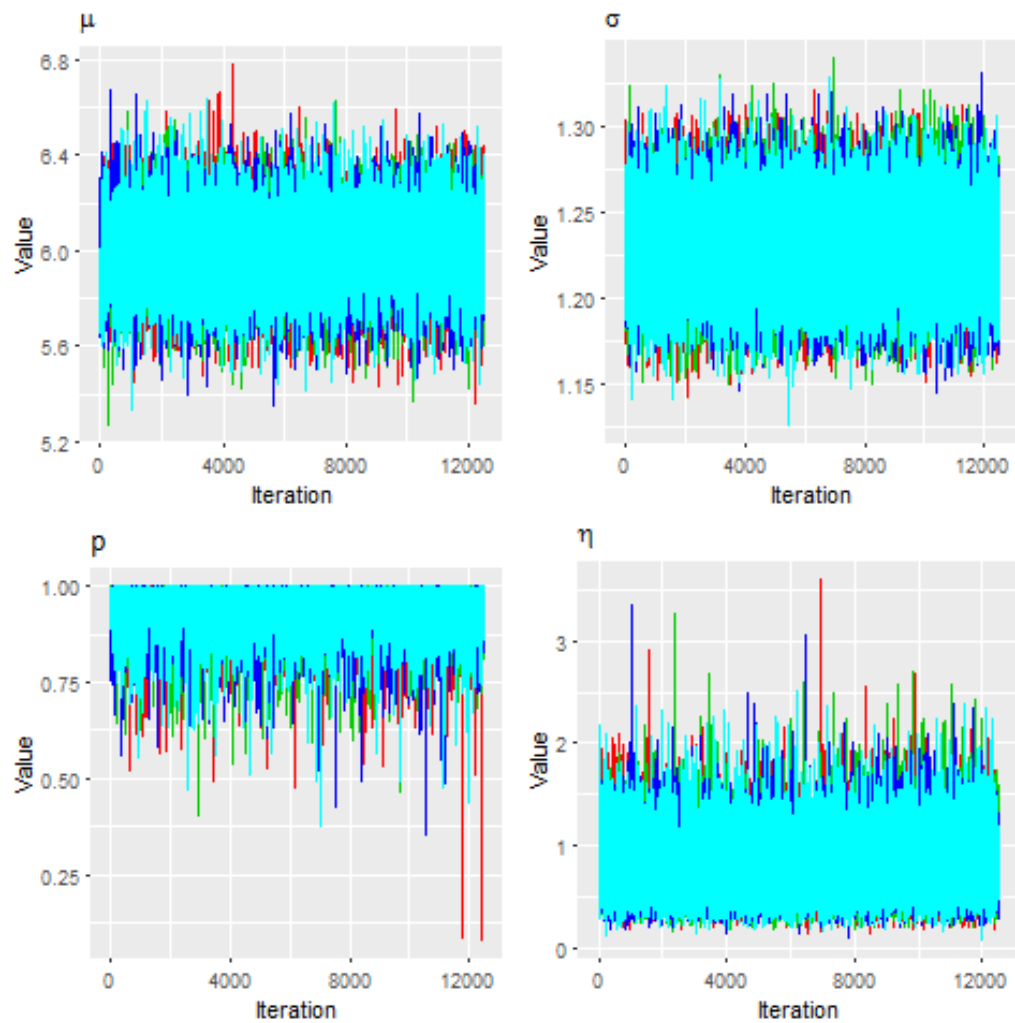
Figure A.4: Trace plots for the posterior distributions of μ, σ, p, and η in Model 2.

A.3   Model 3: Temporal and Areal Spatial Effects
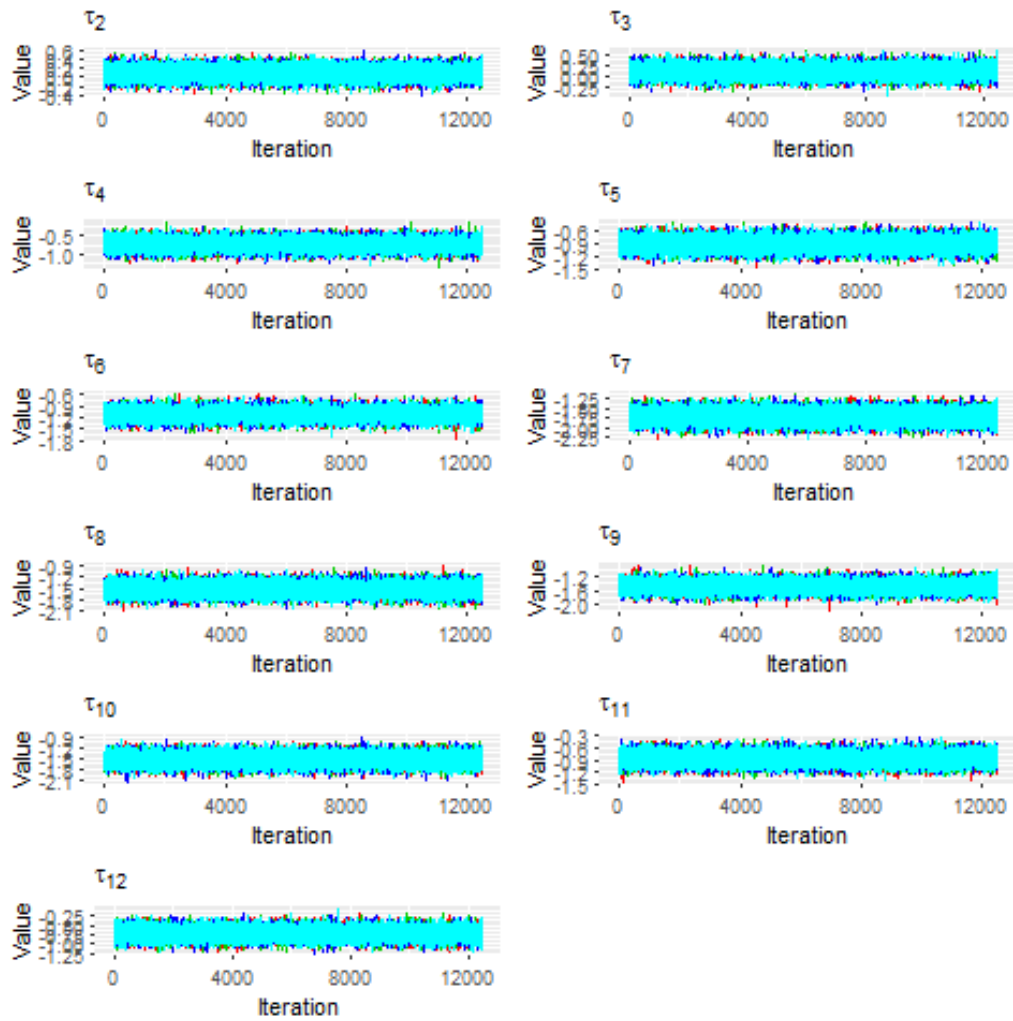
Figure A.5: Trace plots for the posterior distributions of $\tau_2, ..., \tau_{12}$ in Model 3.

Figure A.6: Trace plots for the posterior distributions of $\omega_2, ..., \omega_{23}$ in Model 3.
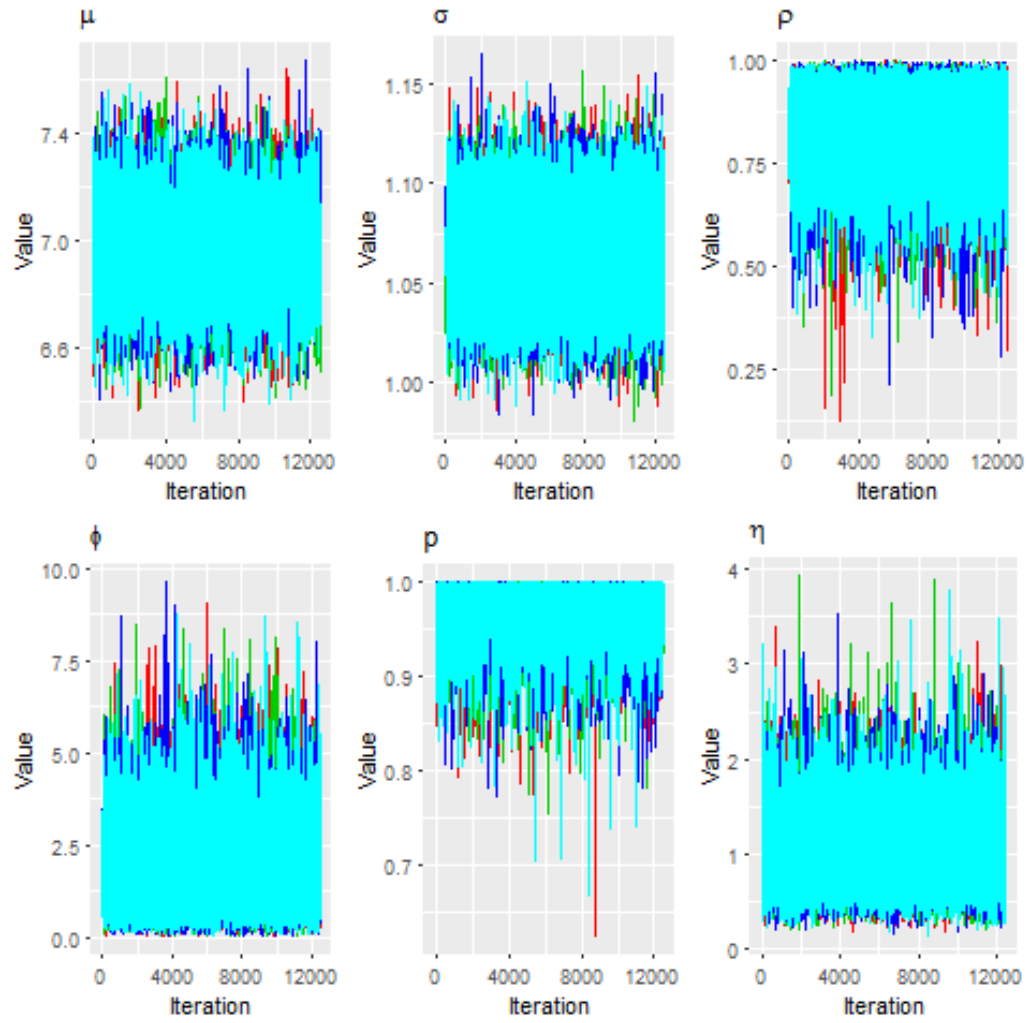
Figure A.7: Trace plots for the posterior distributions of μ, σ, ρ, ϕ, p, and η in Model 3.

A.4 Model 4: Point Referenced CPUE Model
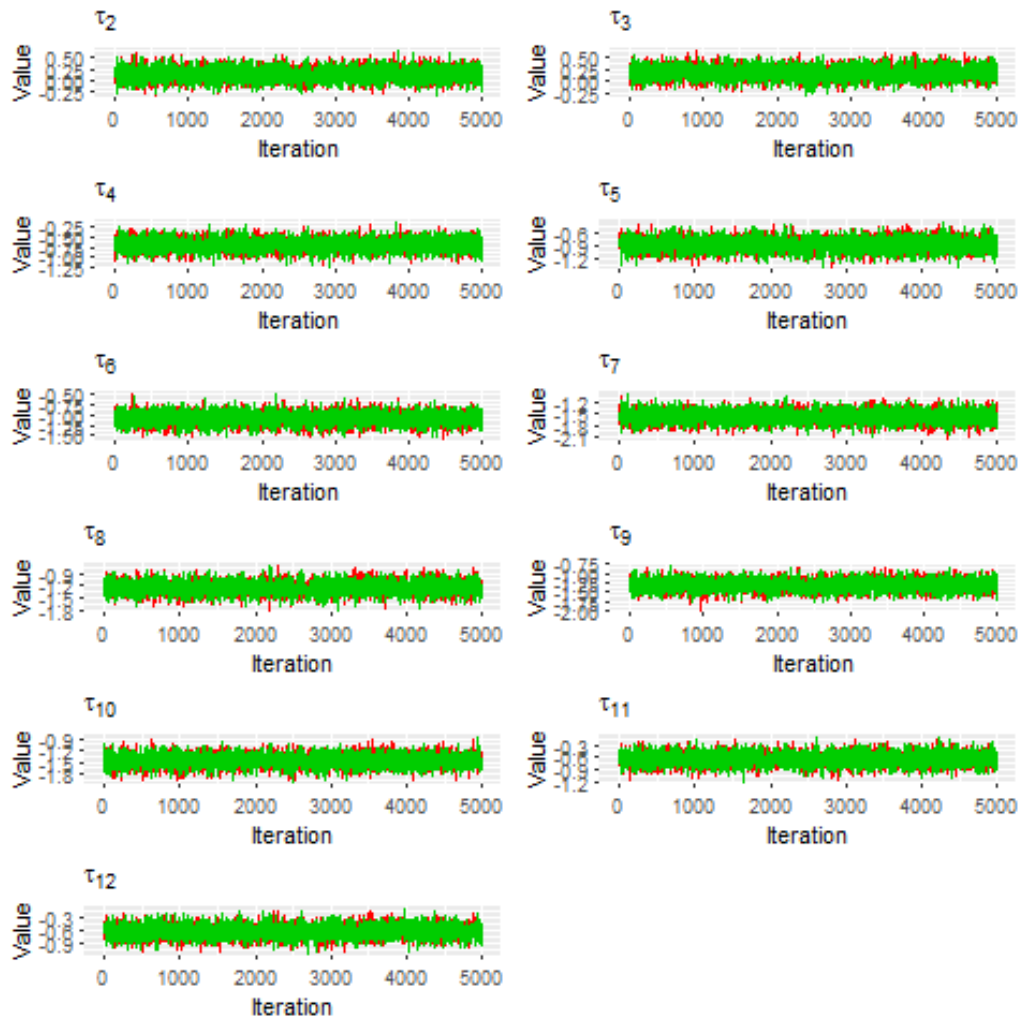
Figure A.8: Trace plots for the posterior distributions of $\tau_2, ..., \tau_{12}$ in Model 4.
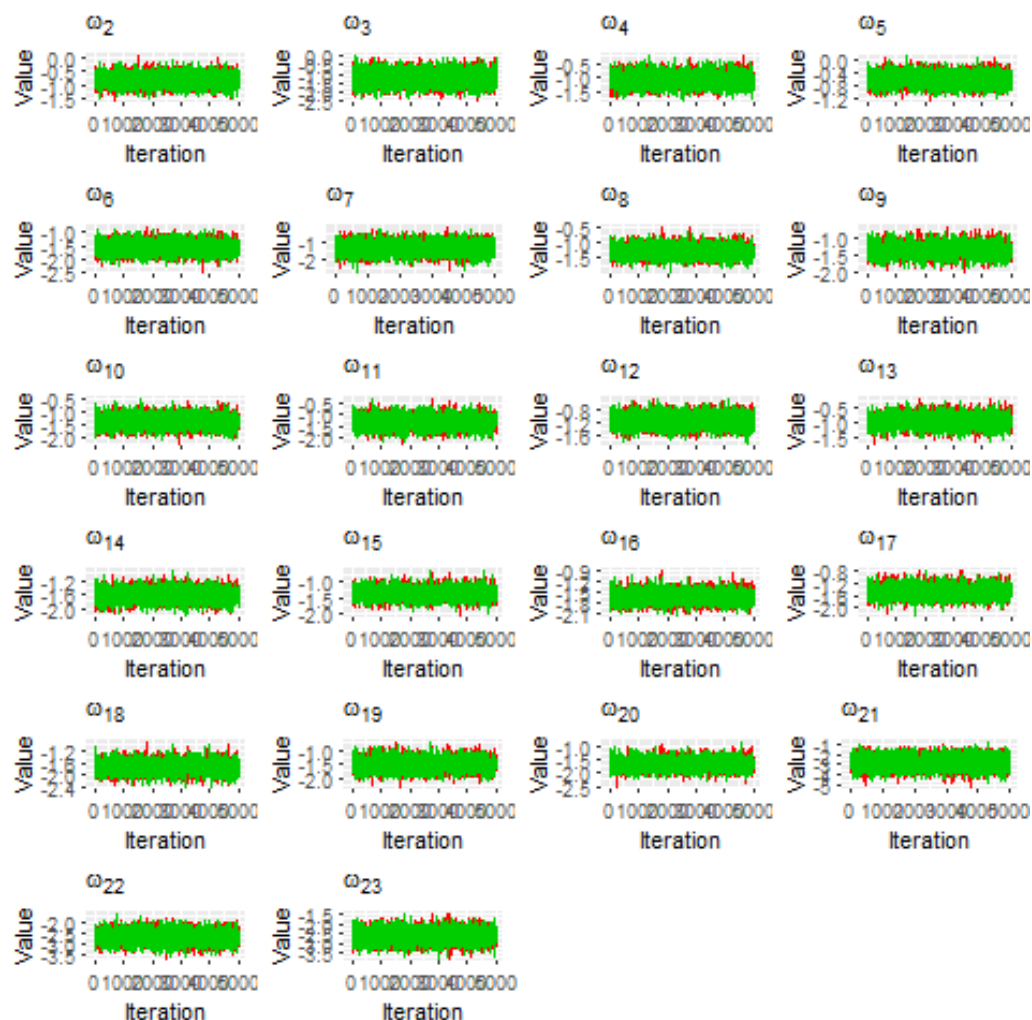
Figure A.9: Trace plots for the posterior distributions of $\omega_2, ..., \omega_{23}$ in Model 4.
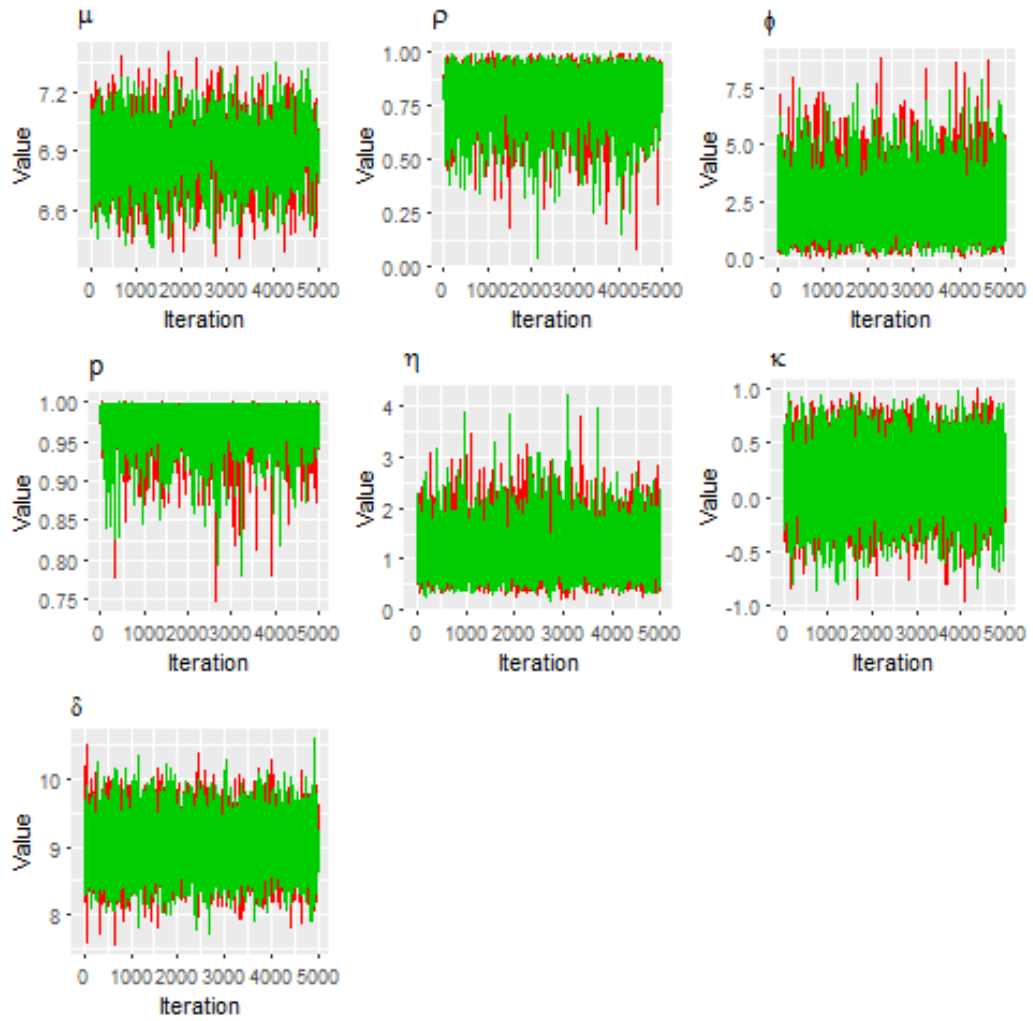
Figure A.10: Trace plots for the posterior distributions of μ, ρ, ϕ, p, η, κ, and δ in Model 4.

A.5   Model 5: Blocked Point Referenced CPUE Model
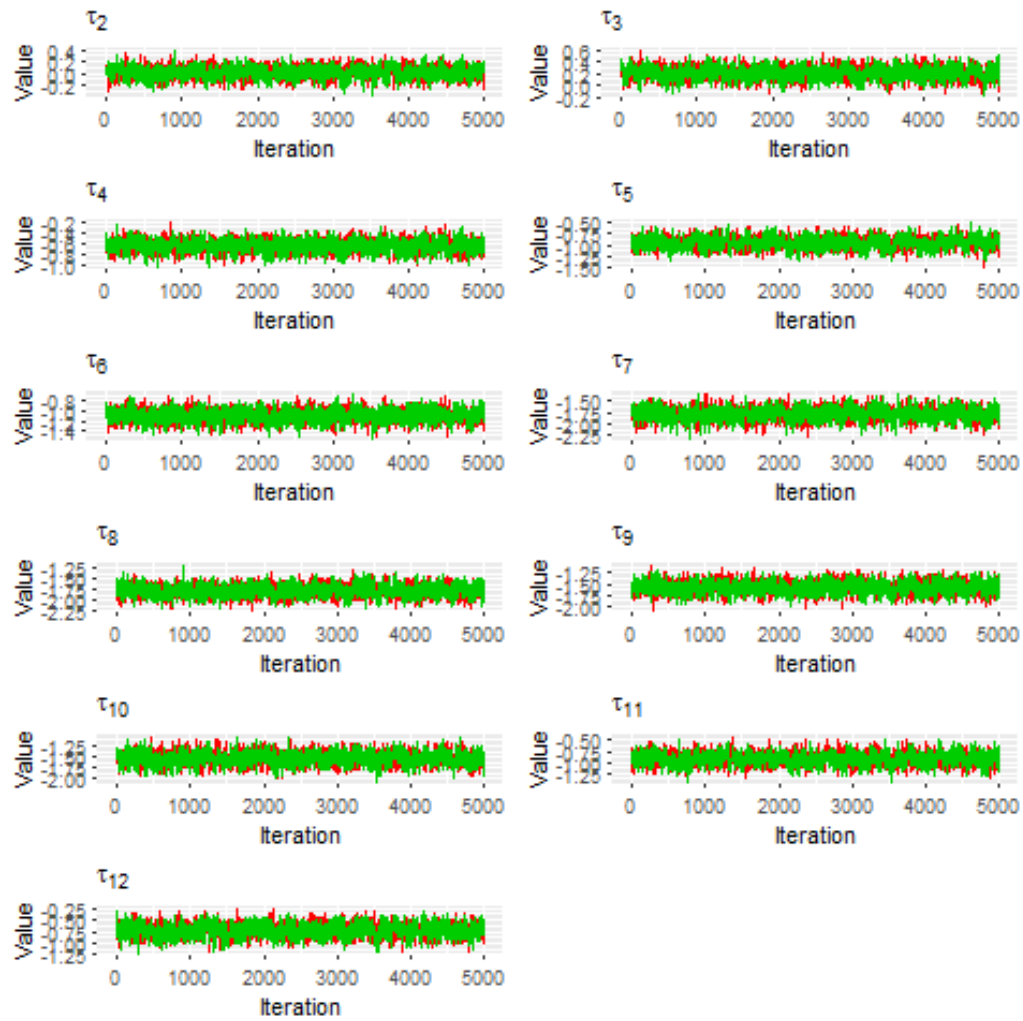
Figure A.11: Trace plots for the posterior distributions of $\tau_2, ..., \tau_{12}$ in Model 5.
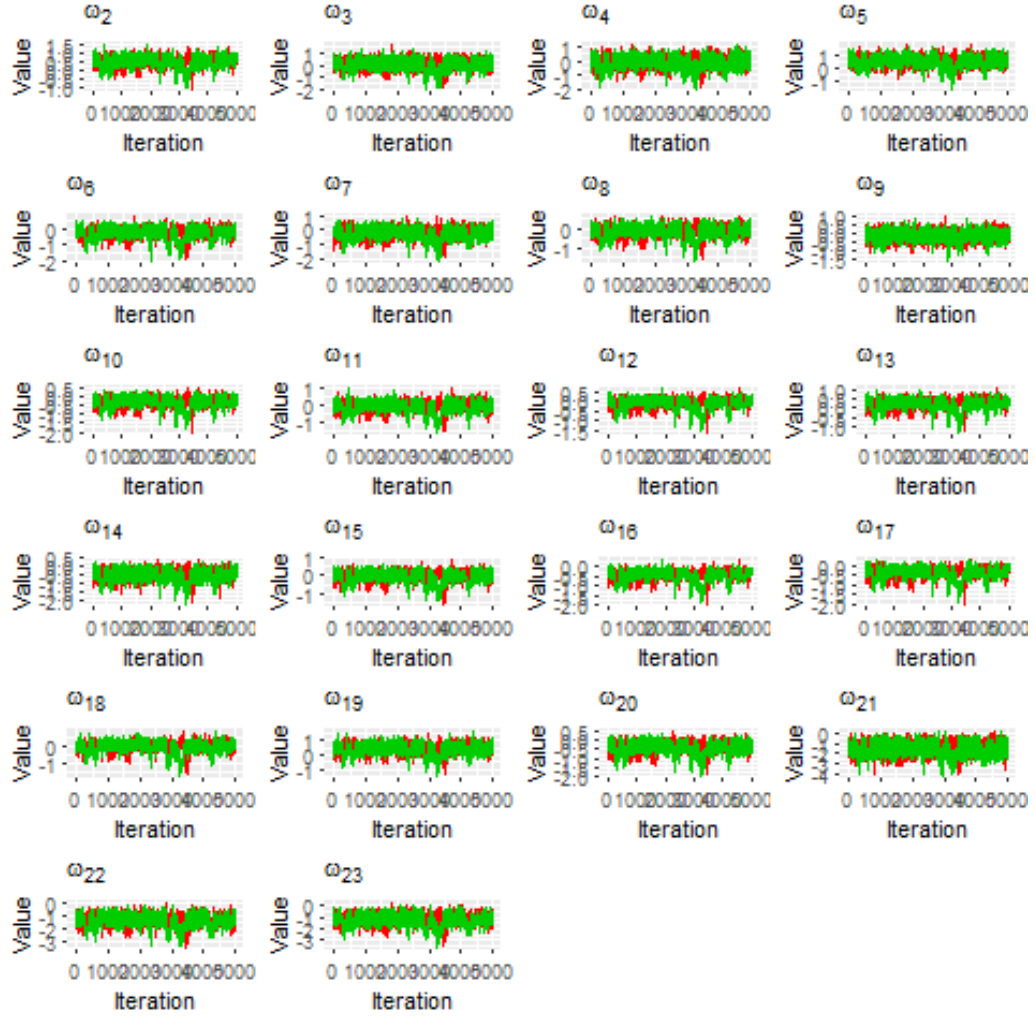
Figure A.12: Trace plots for the posterior distributions of $\omega_2, ..., \omega_{23}$ in Model 5.
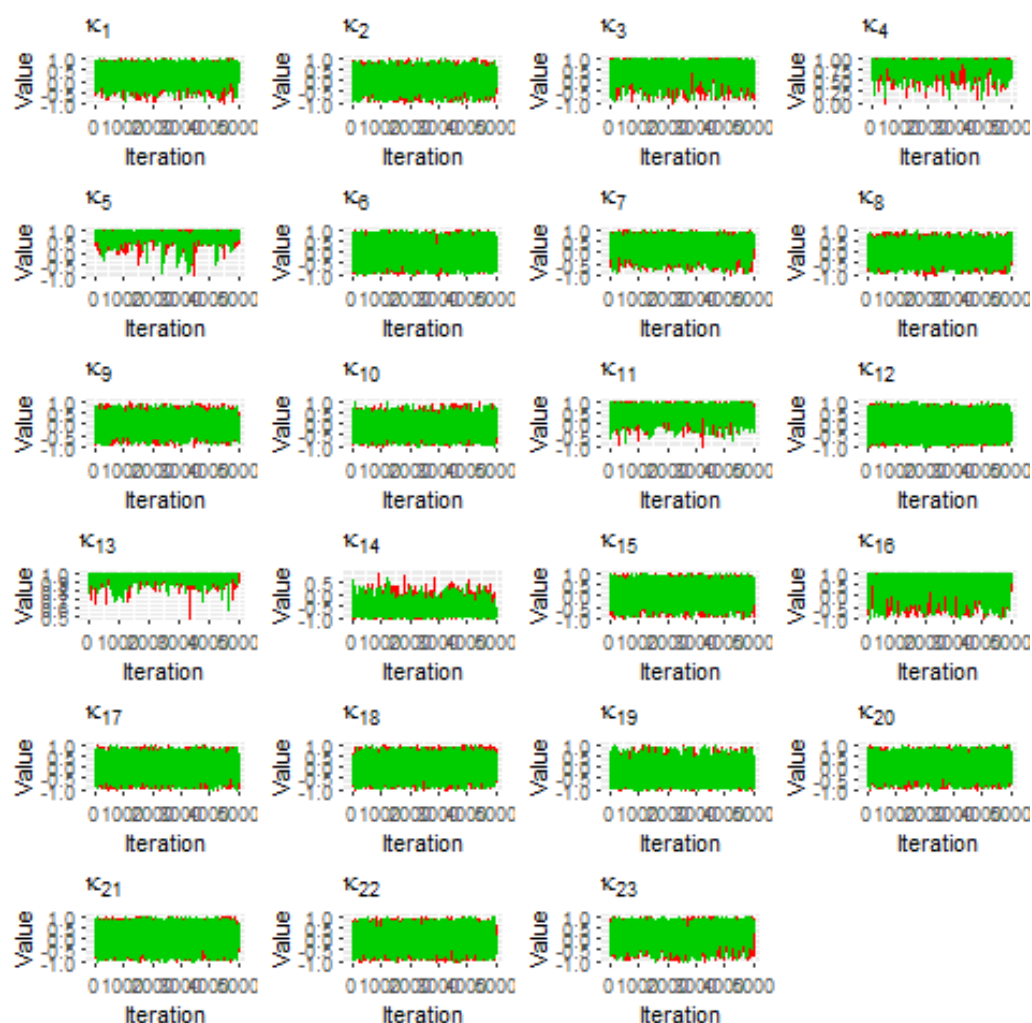
Figure A.13: Trace plots for the posterior distributions of $\kappa_1, ..., \kappa_{23}$ in Model 5.

Figure A.14: Trace plots for the posterior distributions of $\delta_1, ..., \delta_{23}$ in Model 5.
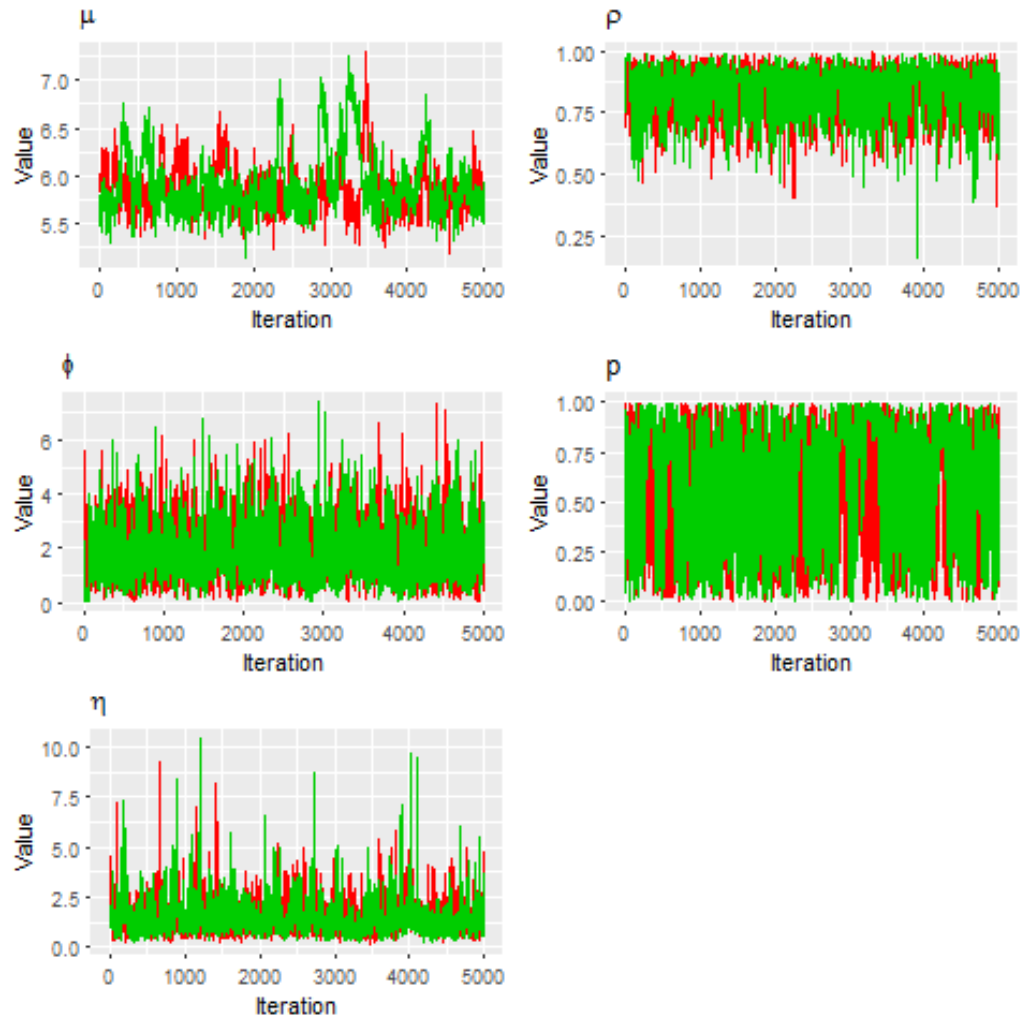
Figure A.15: Trace plots for the posterior distributions of μ, ρ, ϕ, p, and η in Model 5.

Appendix B

Posterior Density Plots

B.1 Model 1: Temporal Effects
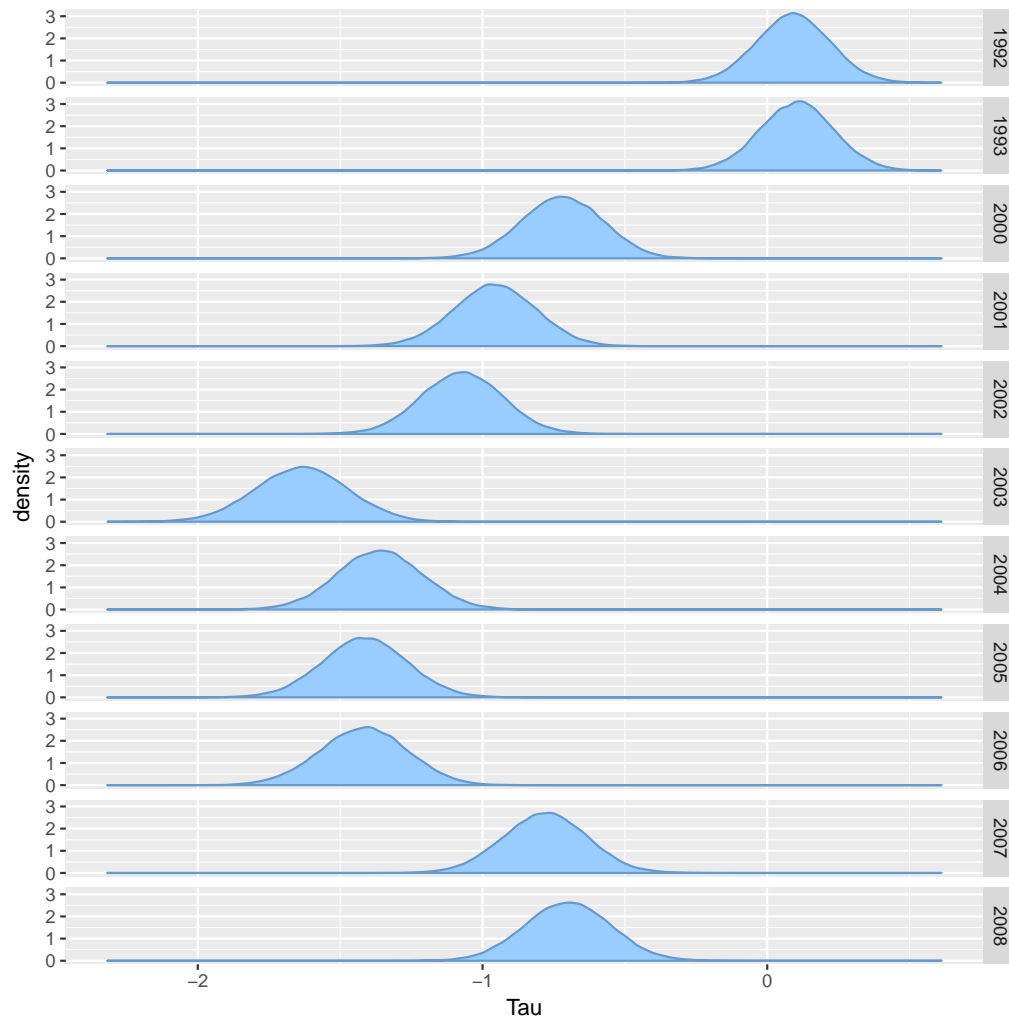
Figure B.1: The posterior distributions for the 11 parameters, $\tau_2, ..., \tau_{12}$, representing the effects on hoki CPUE due to year. The reference level was set to the year 1991.
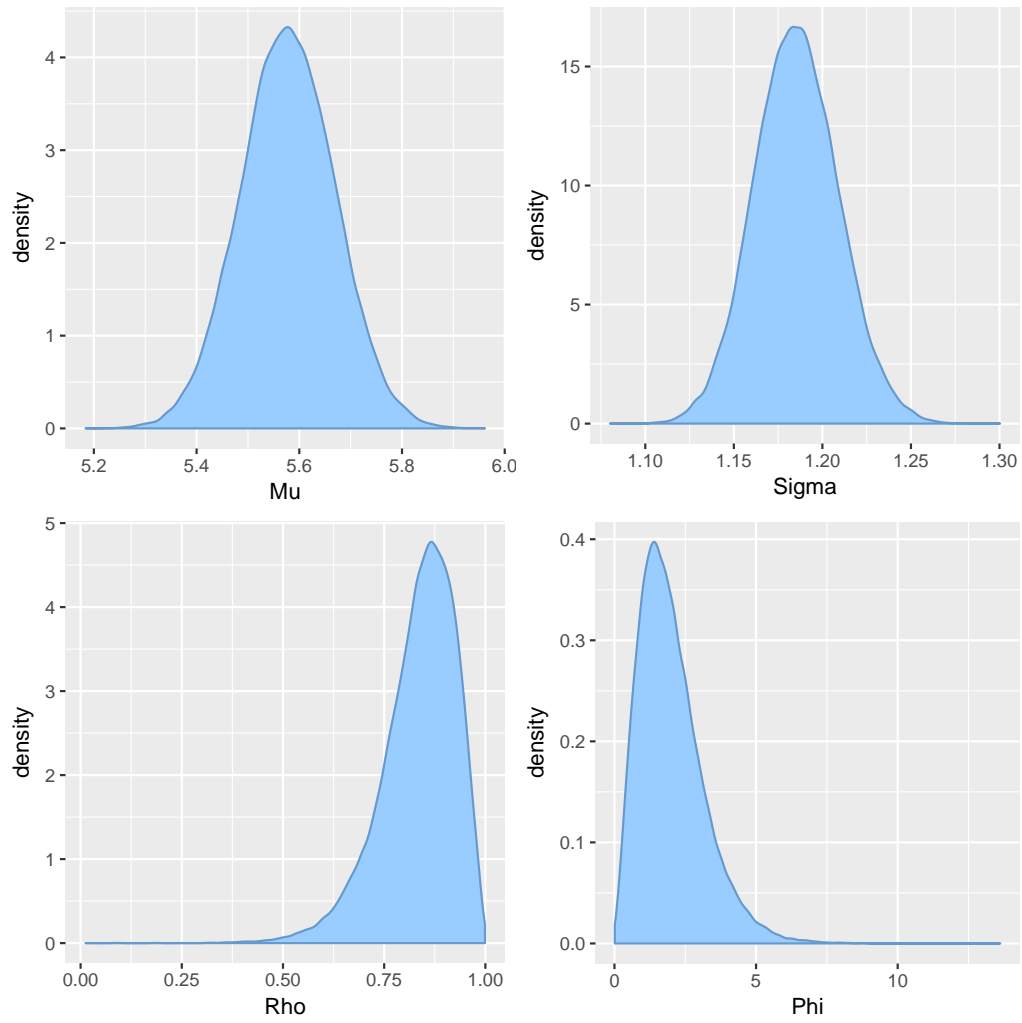
Figure B.2: The posterior distributions for the parameters μ, σ, ρ, and ϕ.

B.2   Model 2: Areal Spatial Effects
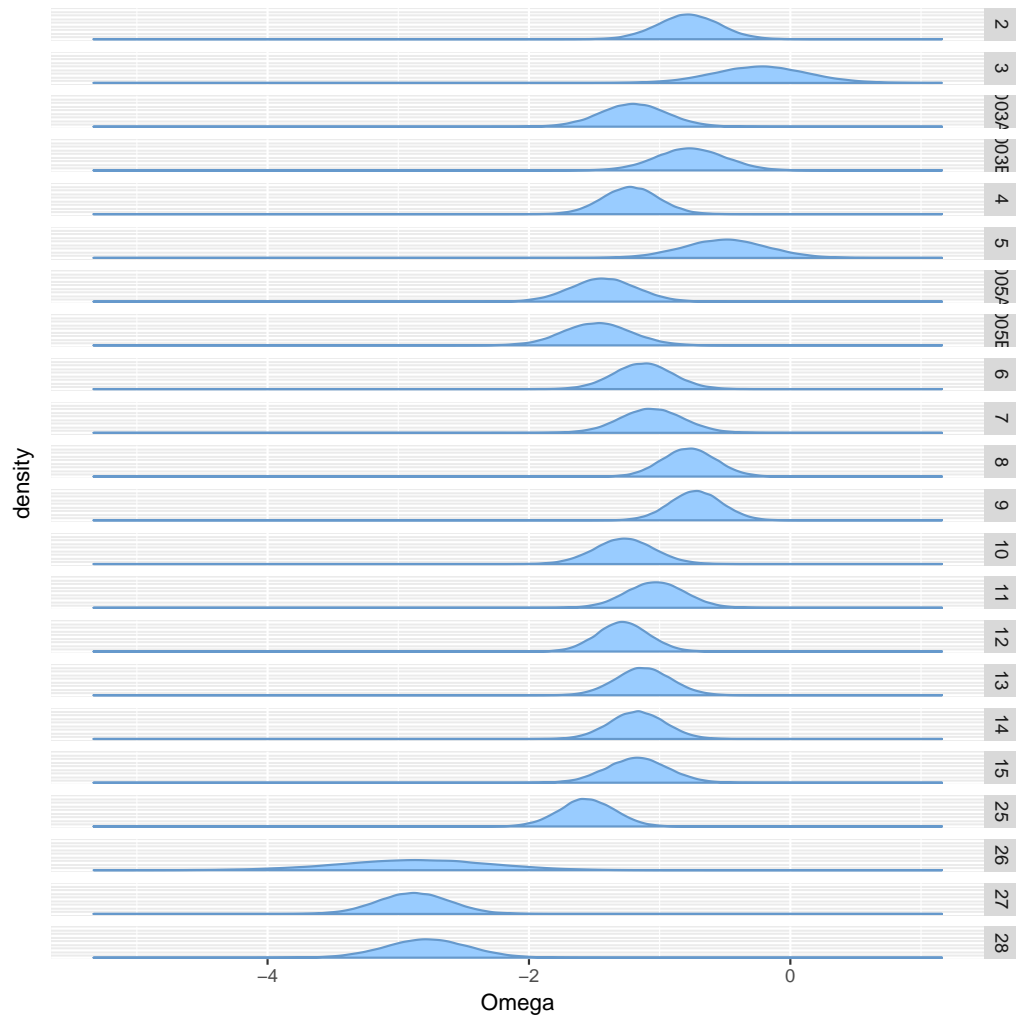
Figure B.3: The posterior distributions for the 22 parameters, $\omega_2, ..., \omega_{23}$, representing the effects on hoki CPUE due to stratum with reference to stratum 1.
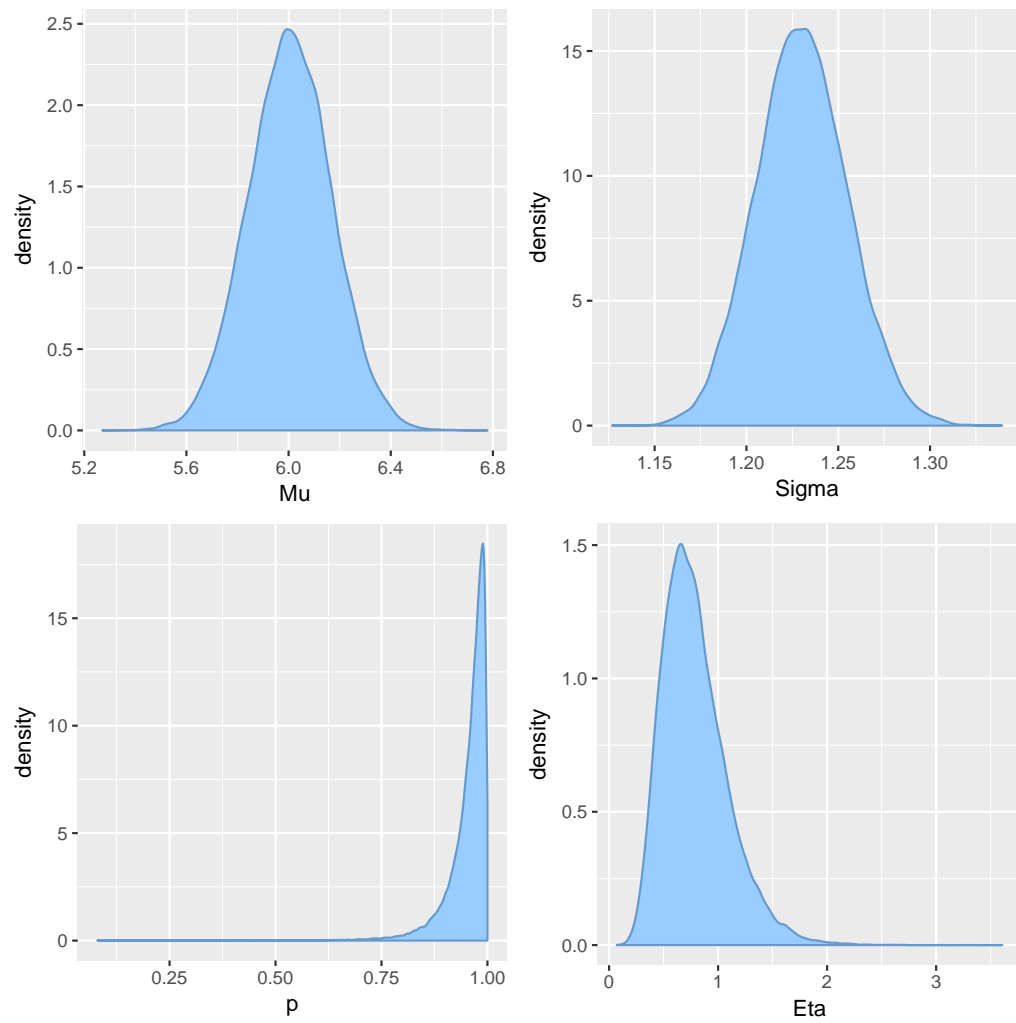
Figure B.4: The posterior distributions for the parameters μ, σ^2, p, and η.

B.3 Model 3: Temporal and Areal Spatial Effects
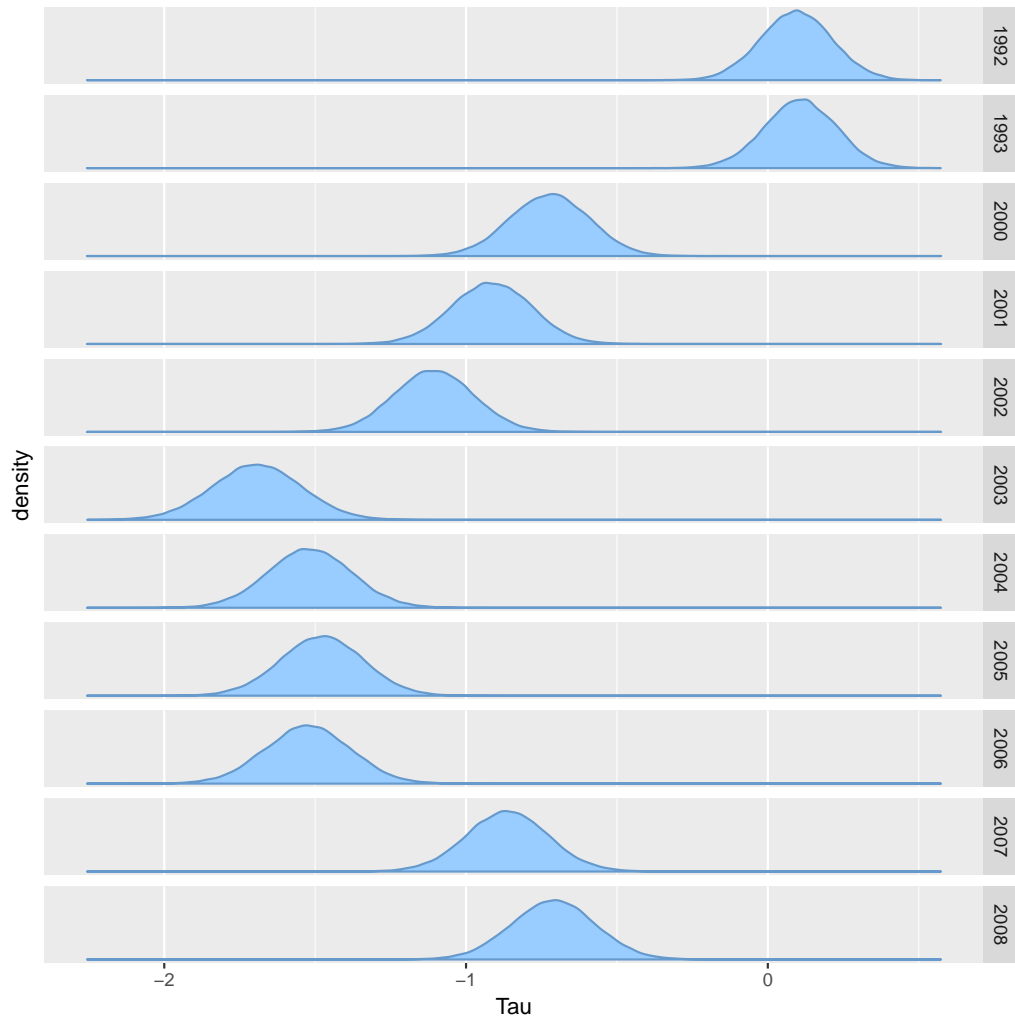
Figure B.5: The posterior distributions for the 11 parameters, $\tau_2, ..., \tau_{12}$, representing the effects on hoki CPUE due to year.
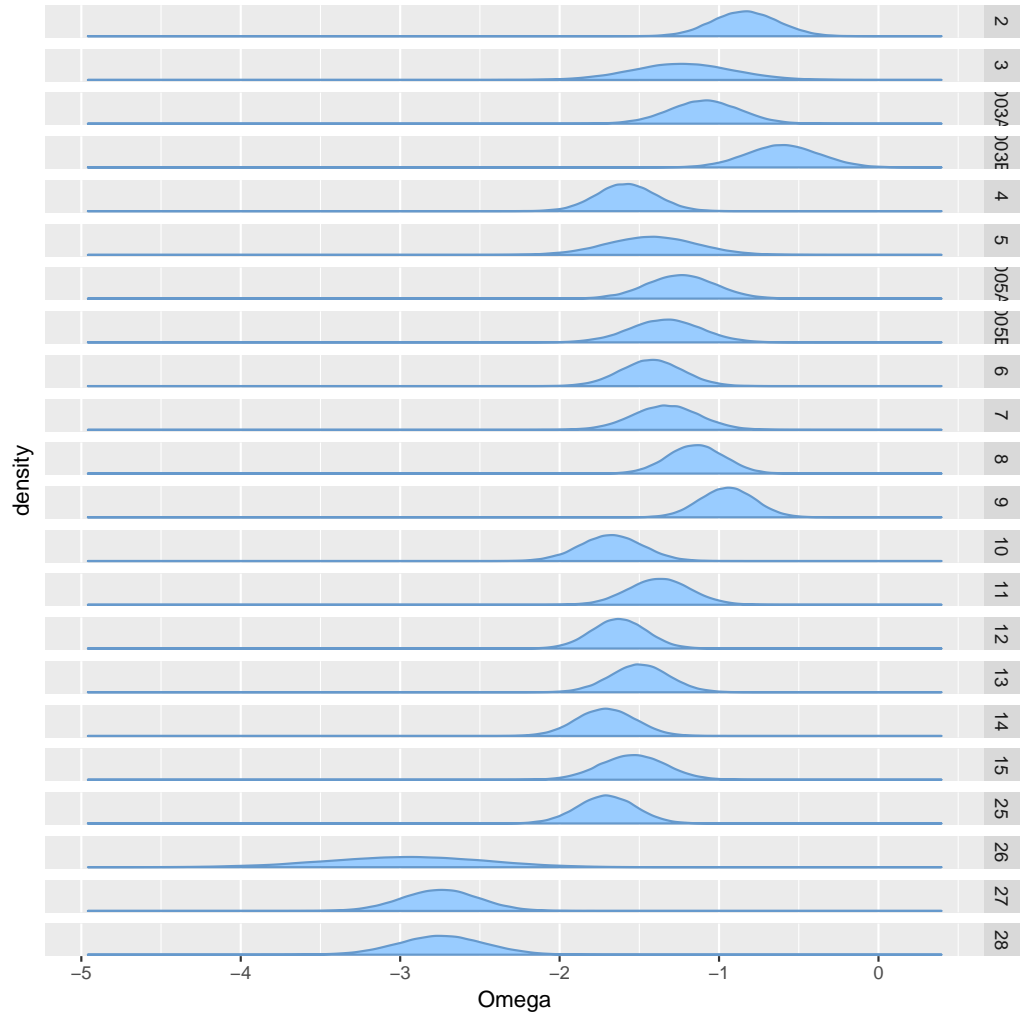
Figure B.6: The posterior distributions for the 22 parameters, $\omega_2, ..., \omega_{23}$, representing the effects on hoki CPUE due to stratum.
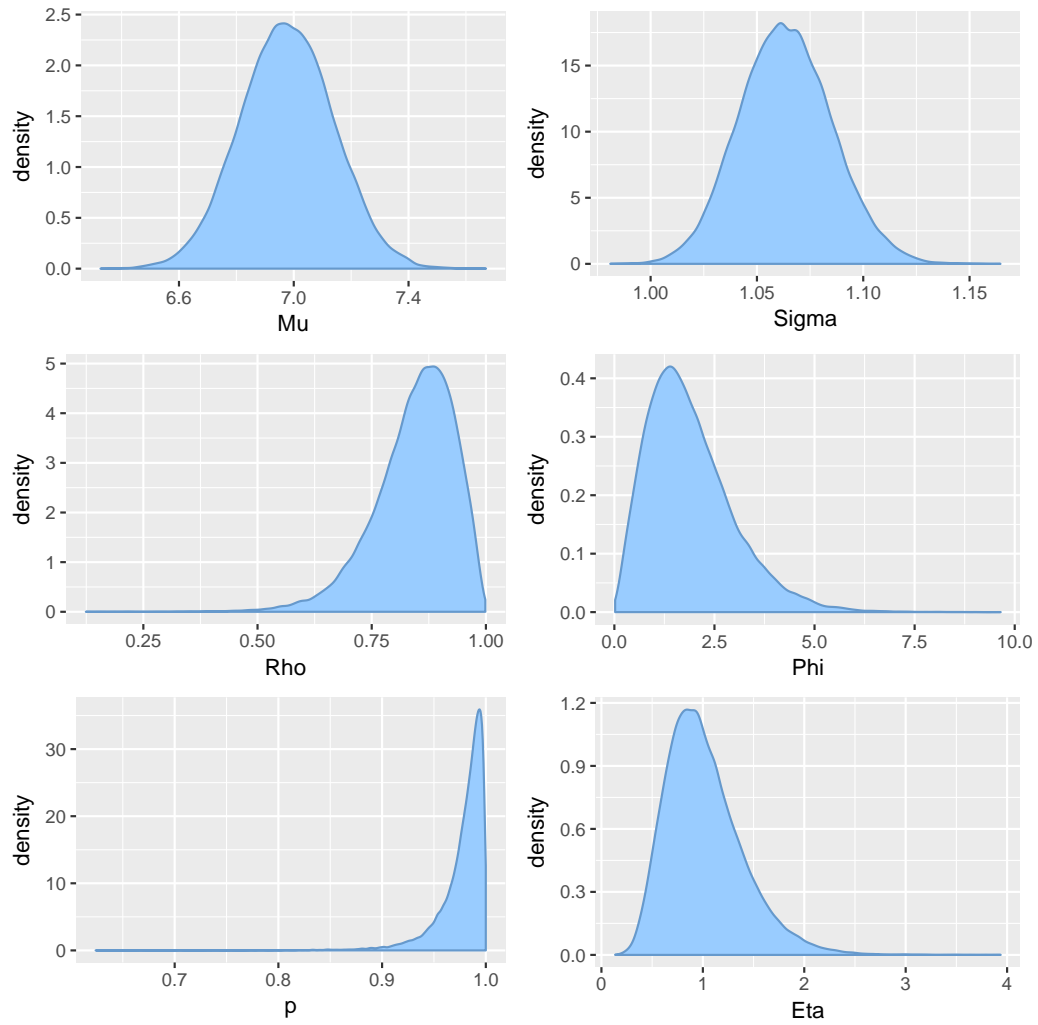
Figure B.7: The posterior distributions for the parameters μ, σ^2, ρ, ϕ, p, and η.

B.4 Model 4: Point Referenced CPUE Model
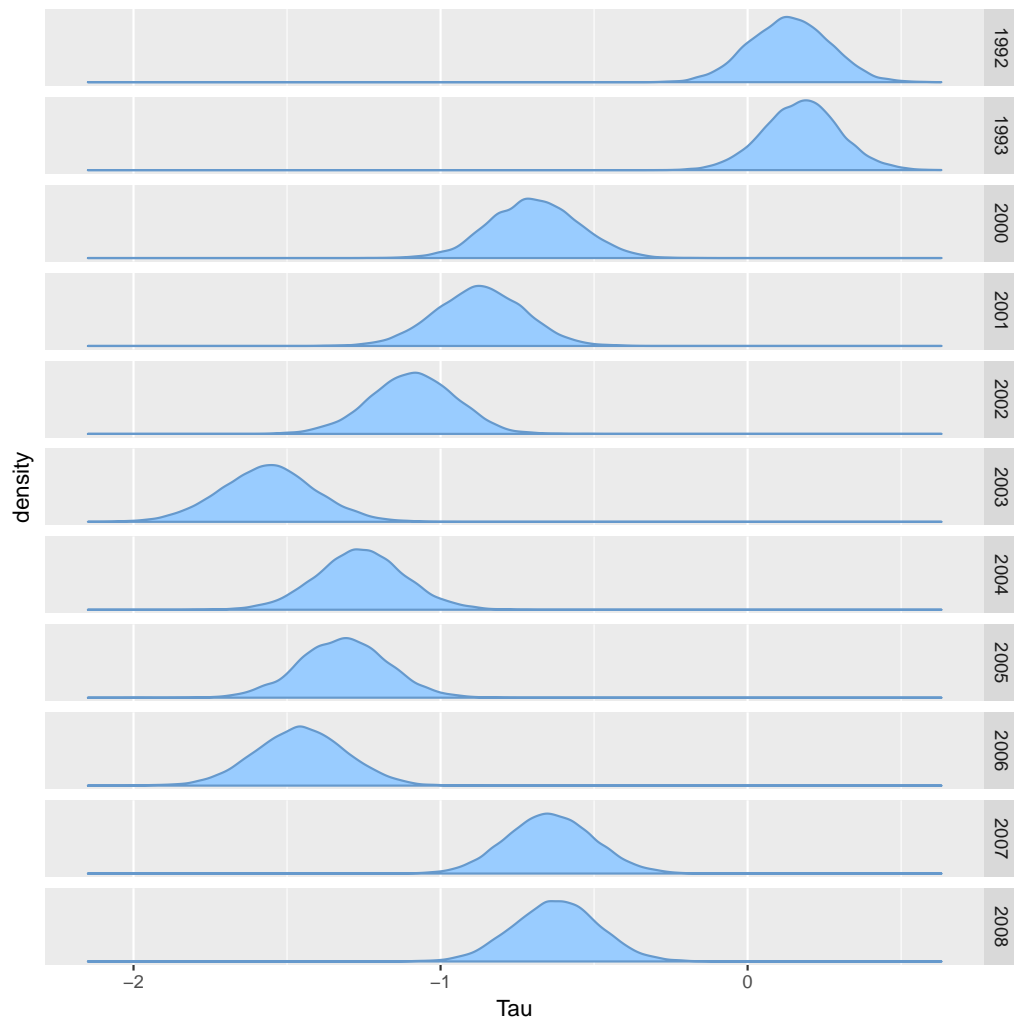
Figure B.8: The posterior distributions for the 11 parameters, $\tau_2, ..., \tau_{12}$, representing the effects on hoki CPUE due to year in Model 4.
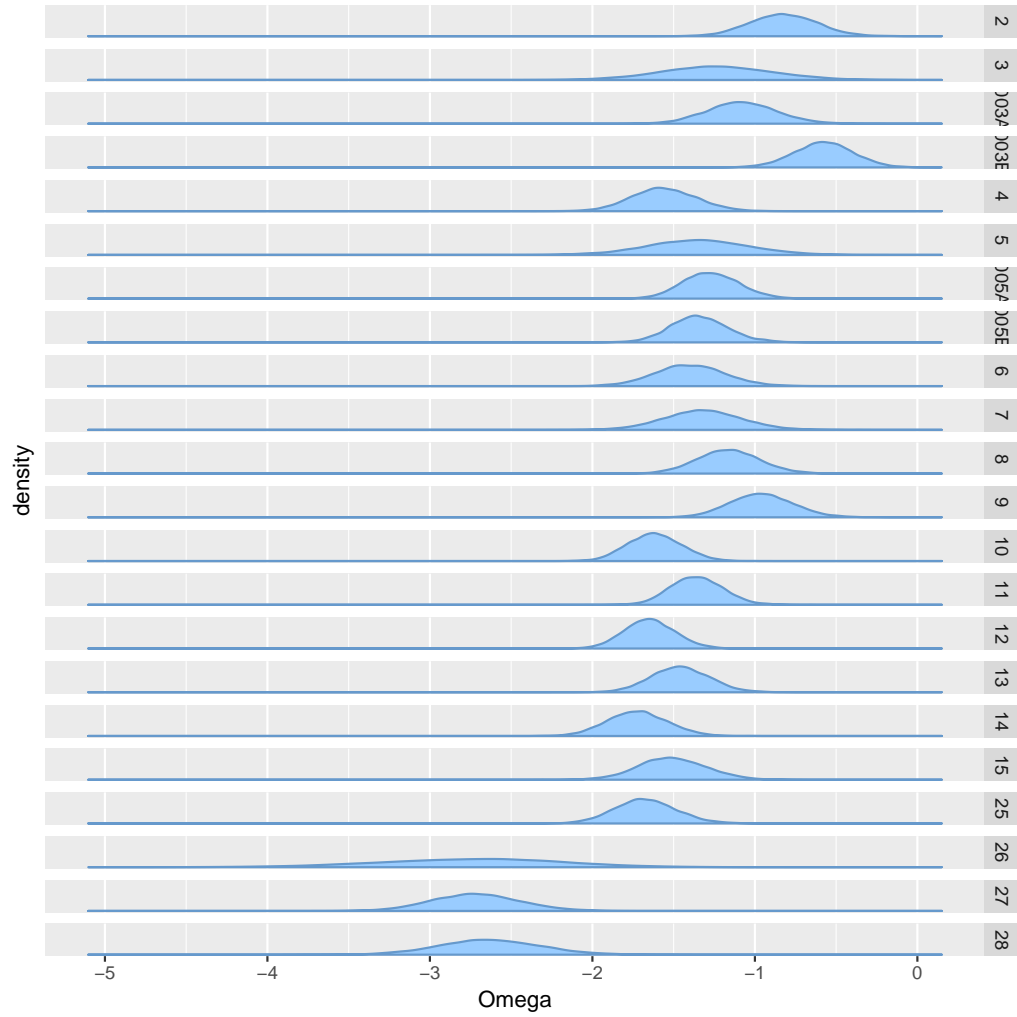
Figure B.9: The posterior distributions for the 22 parameters, $\omega_2, ..., \omega_{23}$, representing the effects on hoki CPUE due to stratum in Model 4.
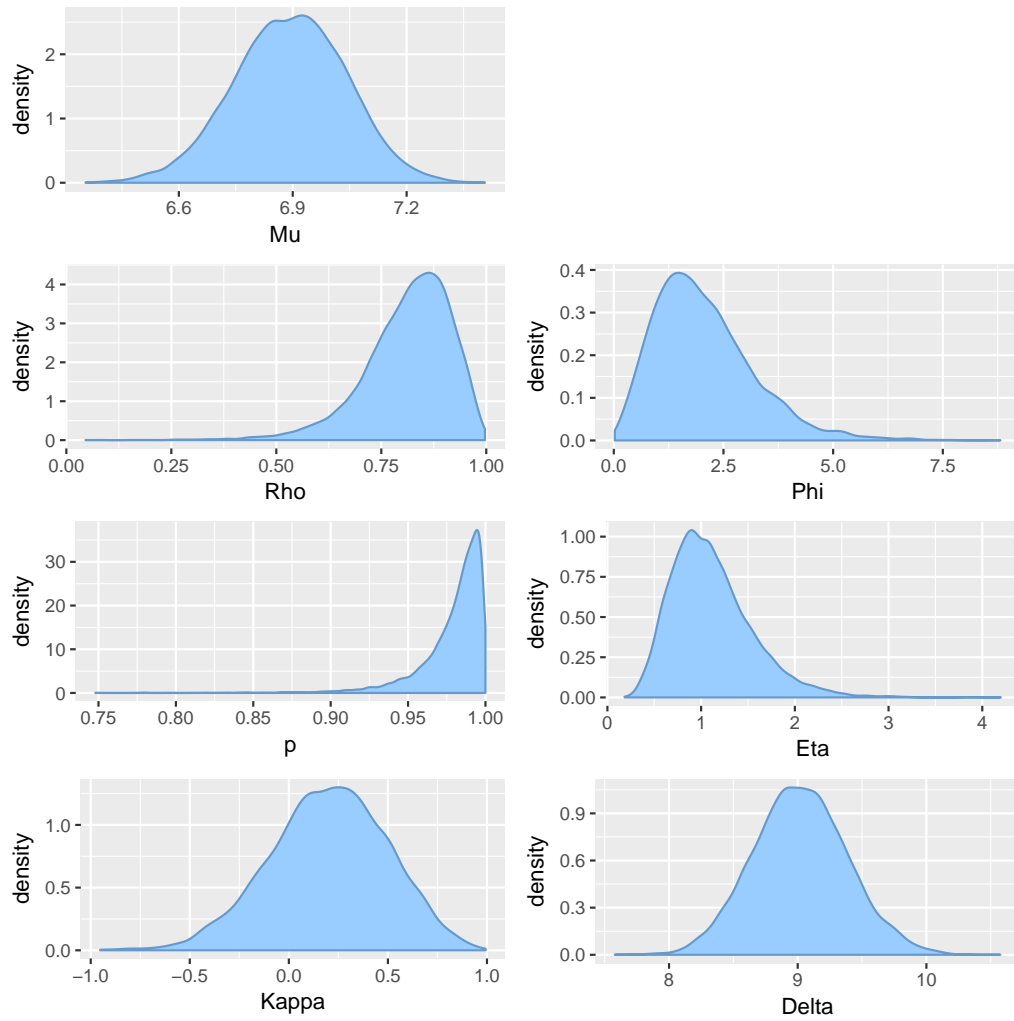
Figure B.10: The posterior distributions for the parameters μ, ρ, ϕ, p, η, κ, and δ in Model 4.

B.5　Model 5: Blocked Point Referenced CPUE Model
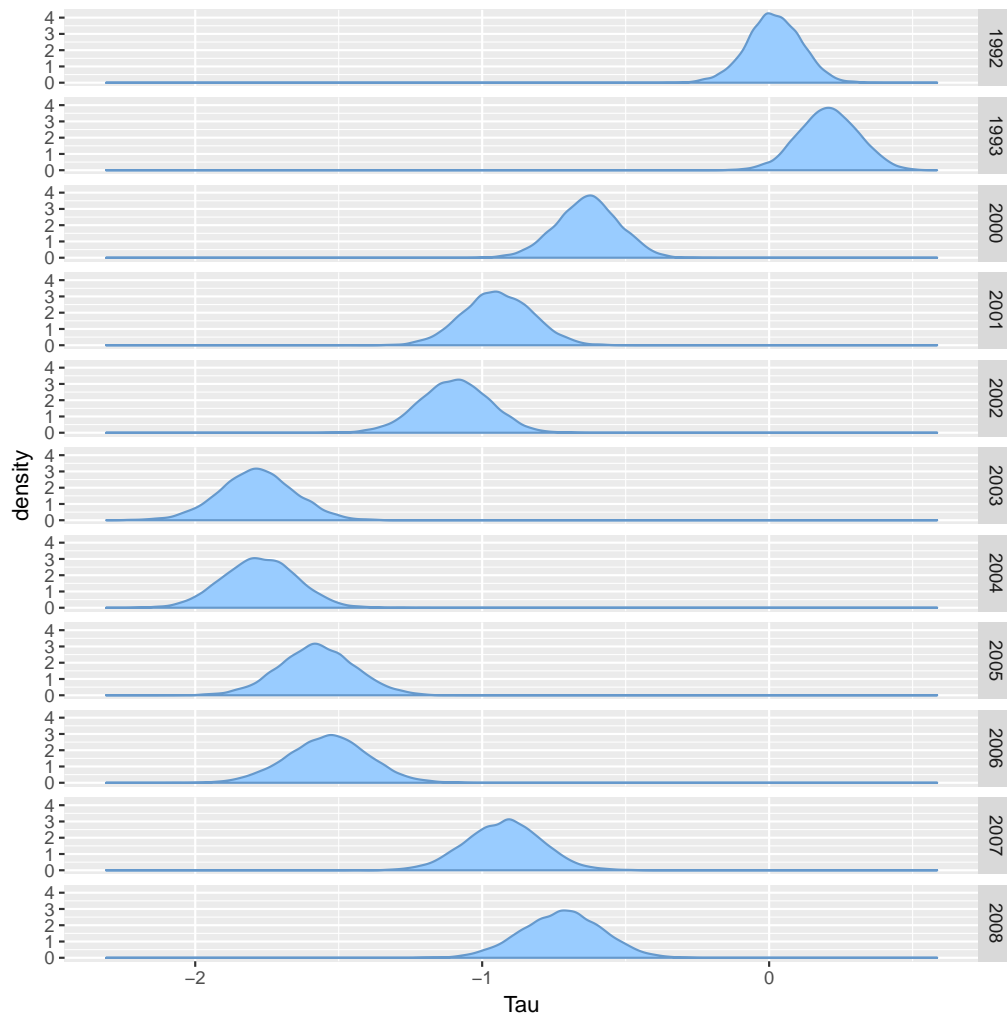
Figure B.11: The posterior distributions for the 11 parameters, $\tau_2, ..., \tau_{12}$, representing the effects on hoki CPUE due to year.
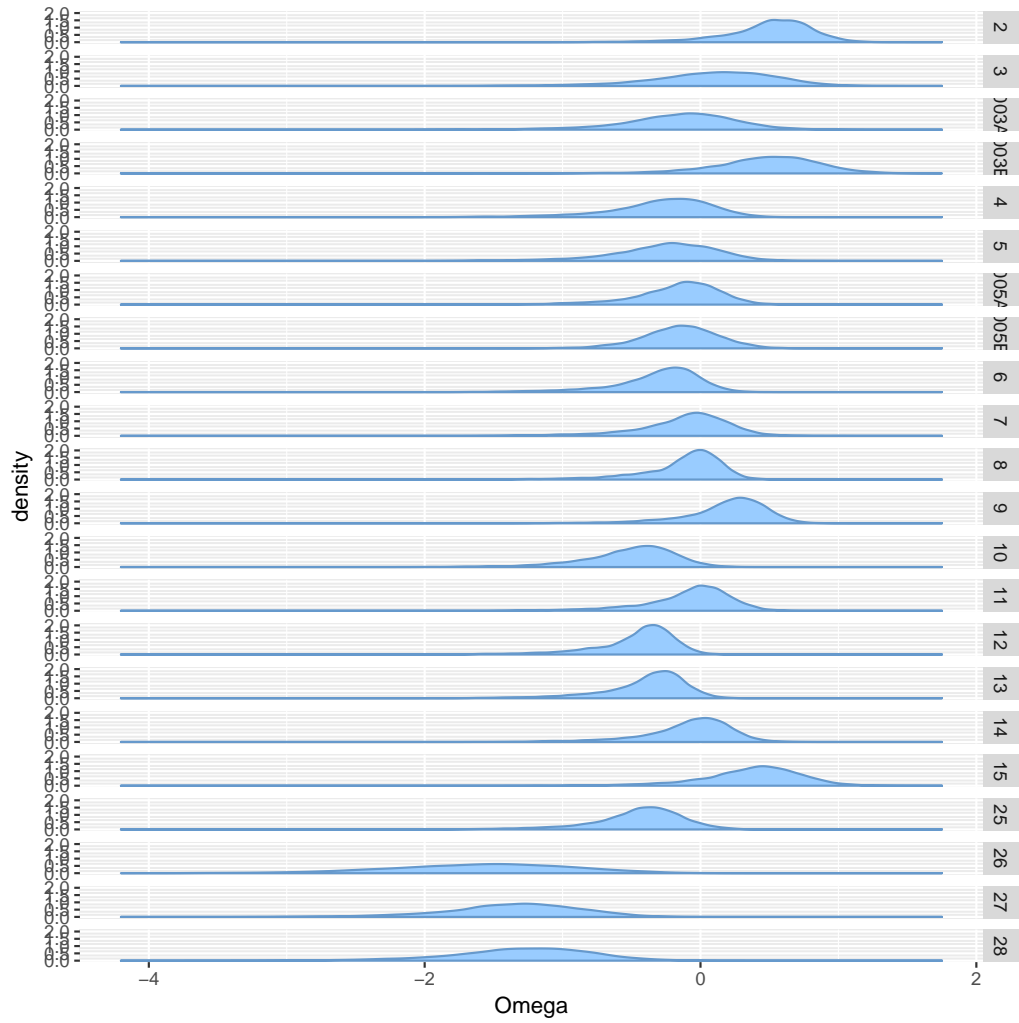
Figure B.12: The posterior distributions for the 22 parameters, $\omega_2, ..., \omega_{23}$, representing the effects on hoki CPUE due to stratum.
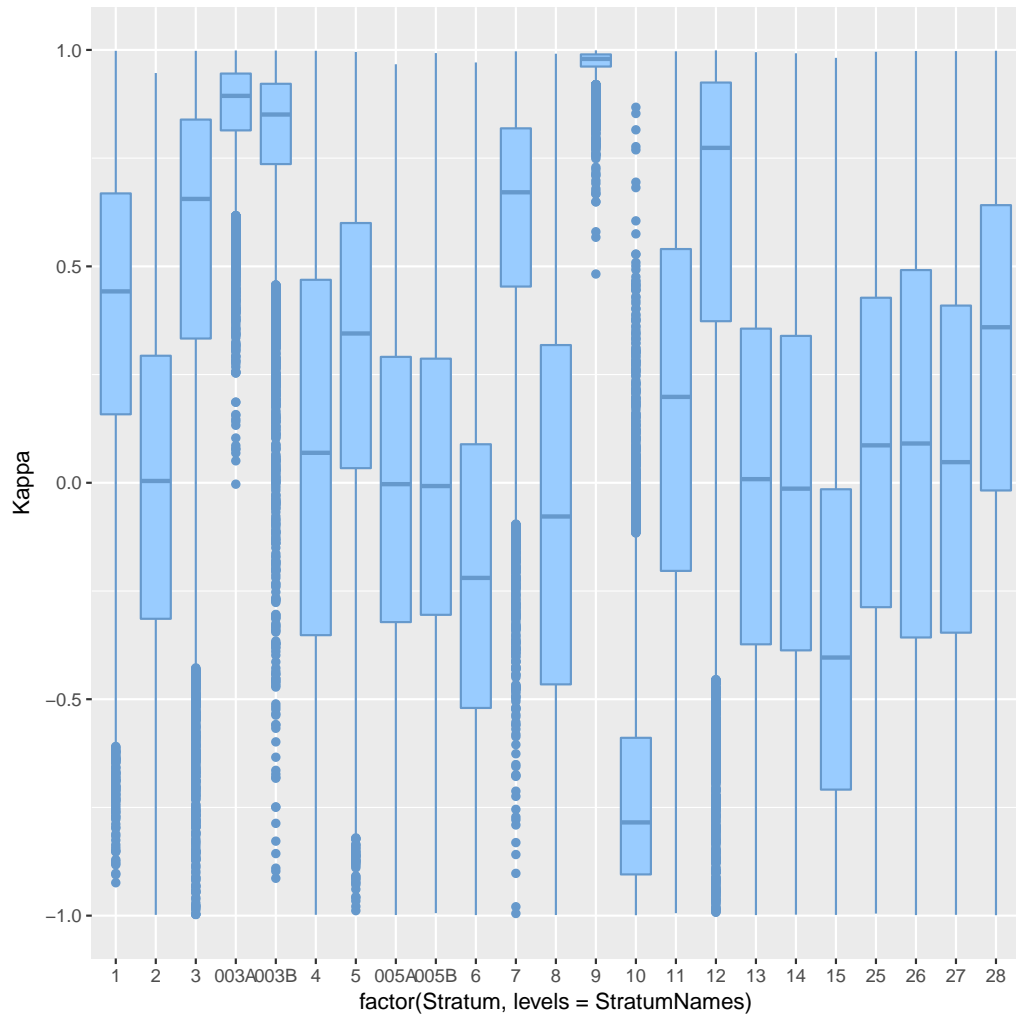
Figure B.13: The posterior distributions for the 23 parameters, $\kappa_1, ..., \kappa_{23}$, representing the correlations between strata.
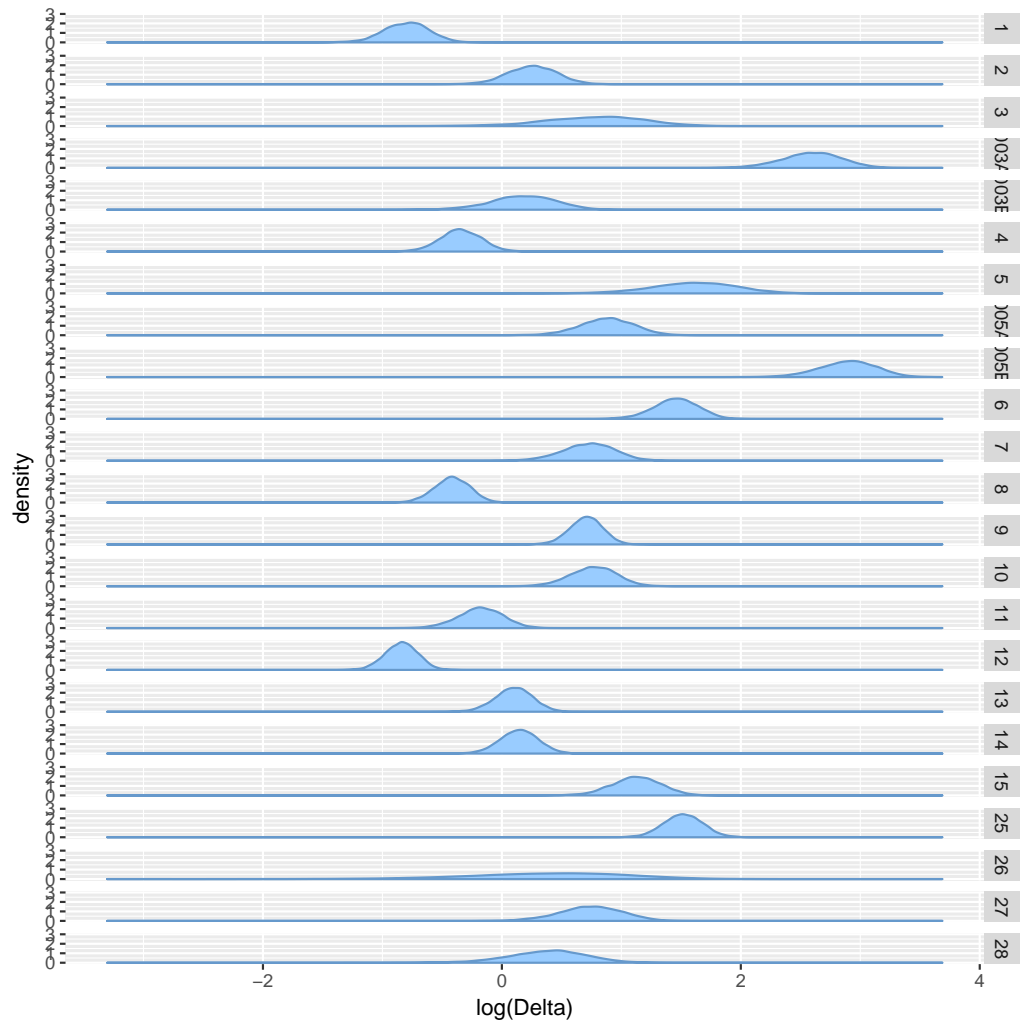
Figure B.14: The posterior distributions for the 23 parameters, $\delta_1, ..., \delta_{23}$, representing the precisions.
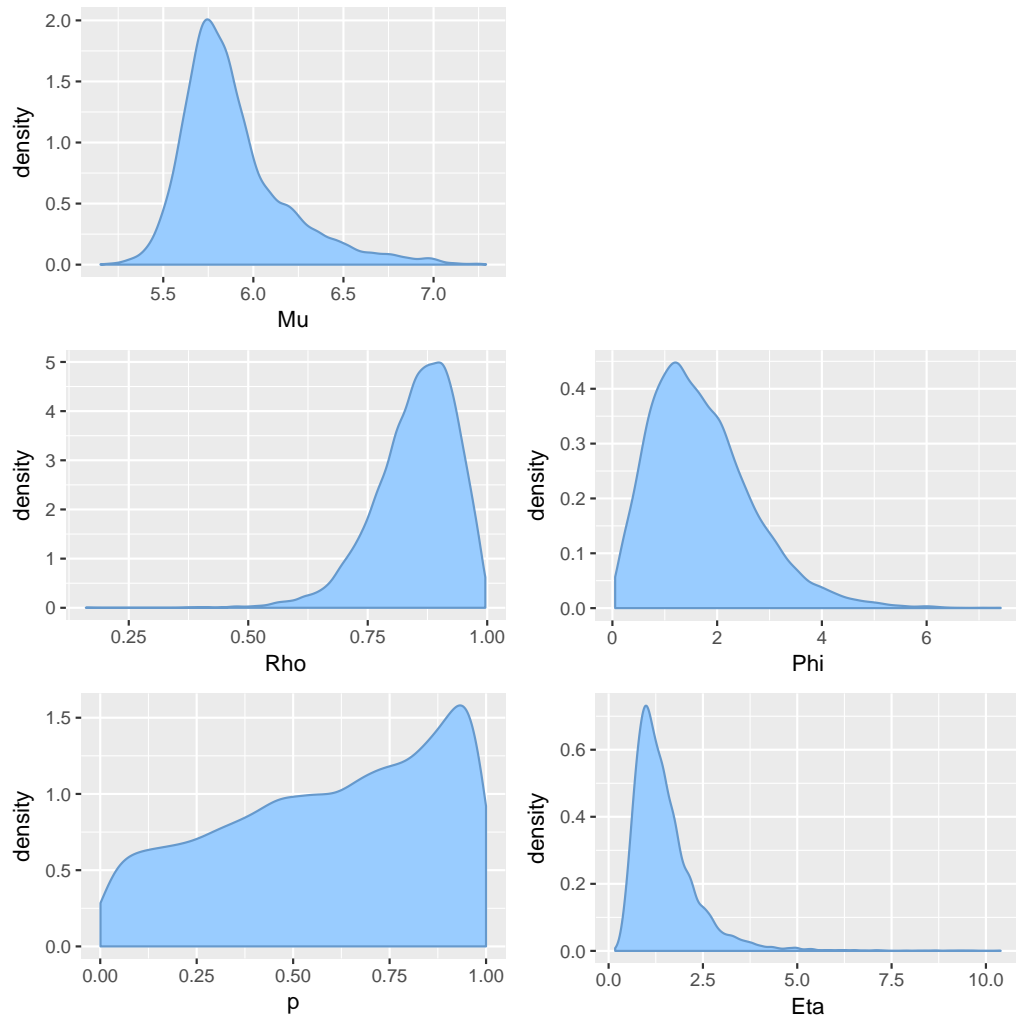
Figure B.15: The posterior distributions for the parameters μ, ρ, ϕ, p, η.

Appendix C

Autocorrelation Plots

C.1   Model 1: Temporal Effects
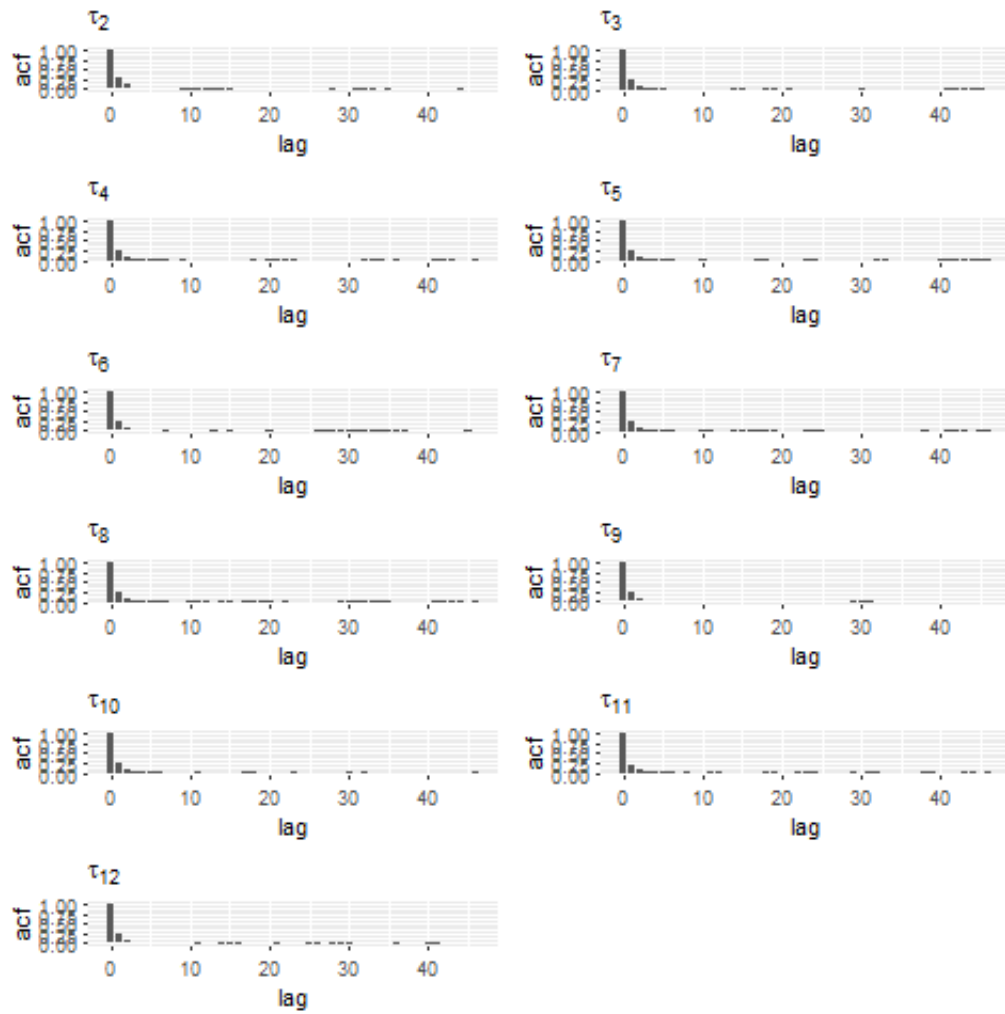
Figure C.1: Autocorrelation plots for $\tau_2, ..., \tau_{12}$ in Model 1.
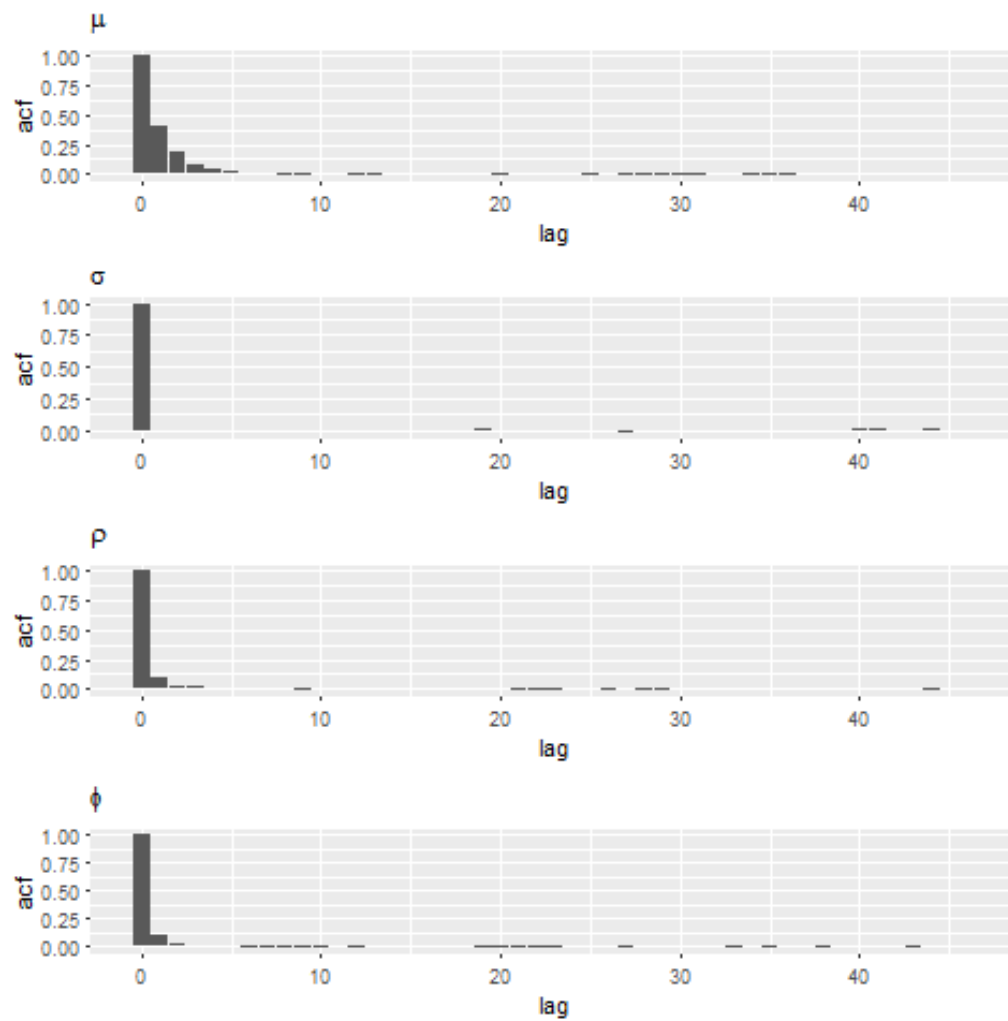
Figure C.2: Autocorrelation plots for μ, σ, ϕ, and ρ in Model 1.

C.2 Model 2: Areal Spatial Effects
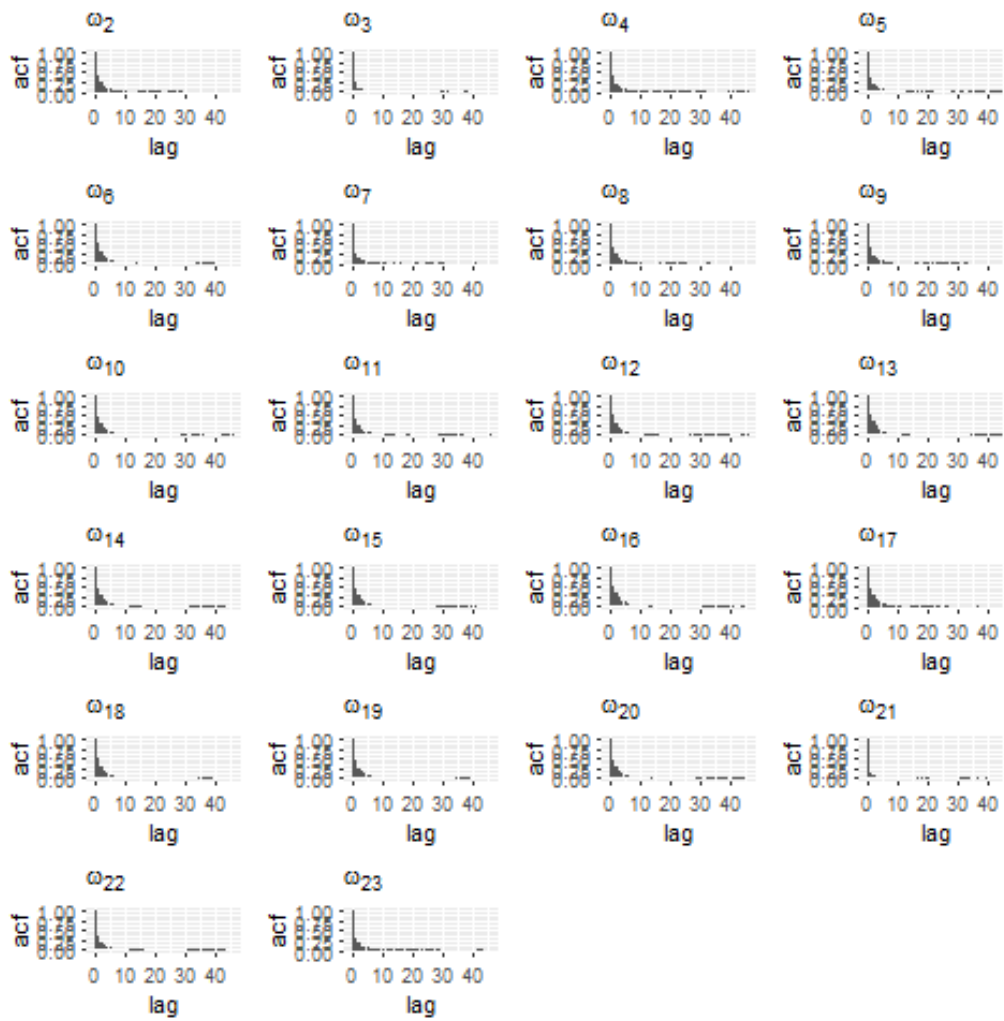
Figure C.3: Autocorrelation plots for $\omega_2, ..., \omega_{23}$ in Model 2.
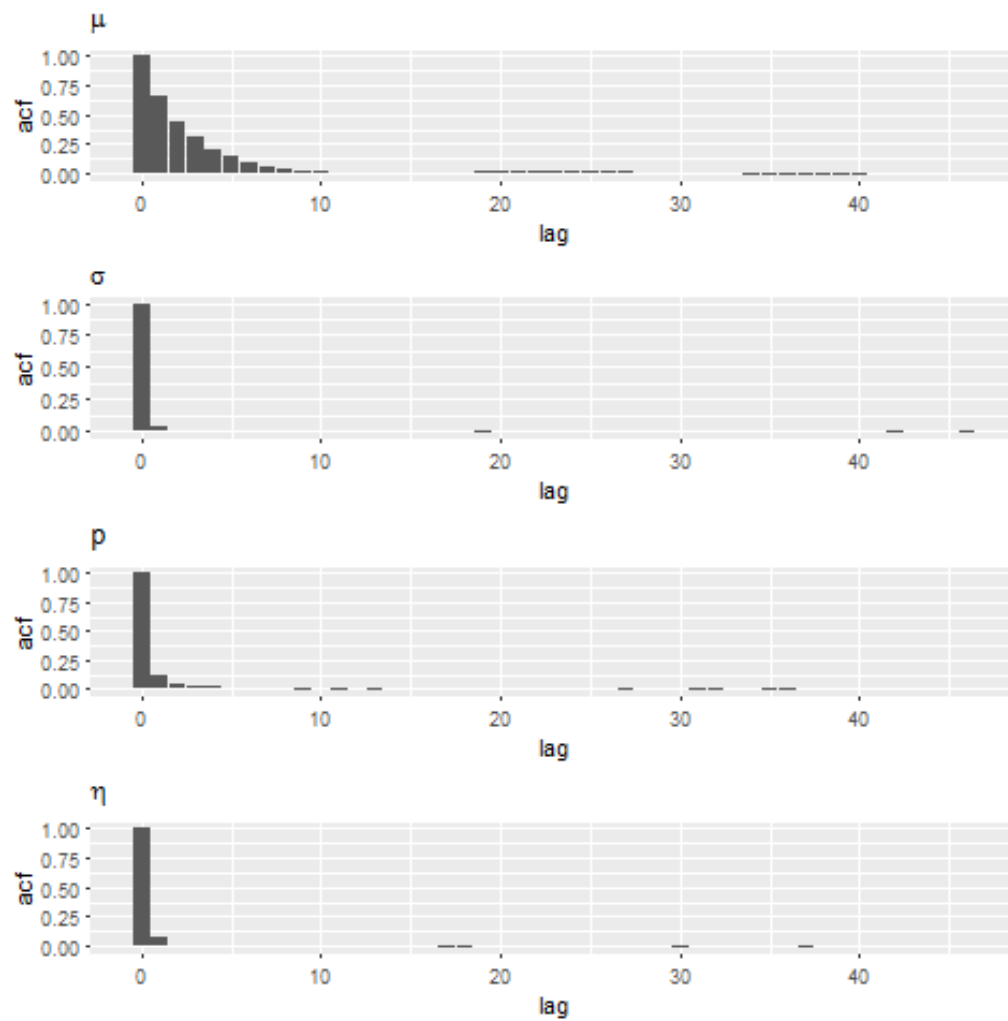
Figure C.4: Autocorrelation plots for μ, σ, p, and η in Model 2.

C.3   Model 3: Temporal and Areal Spatial Effects
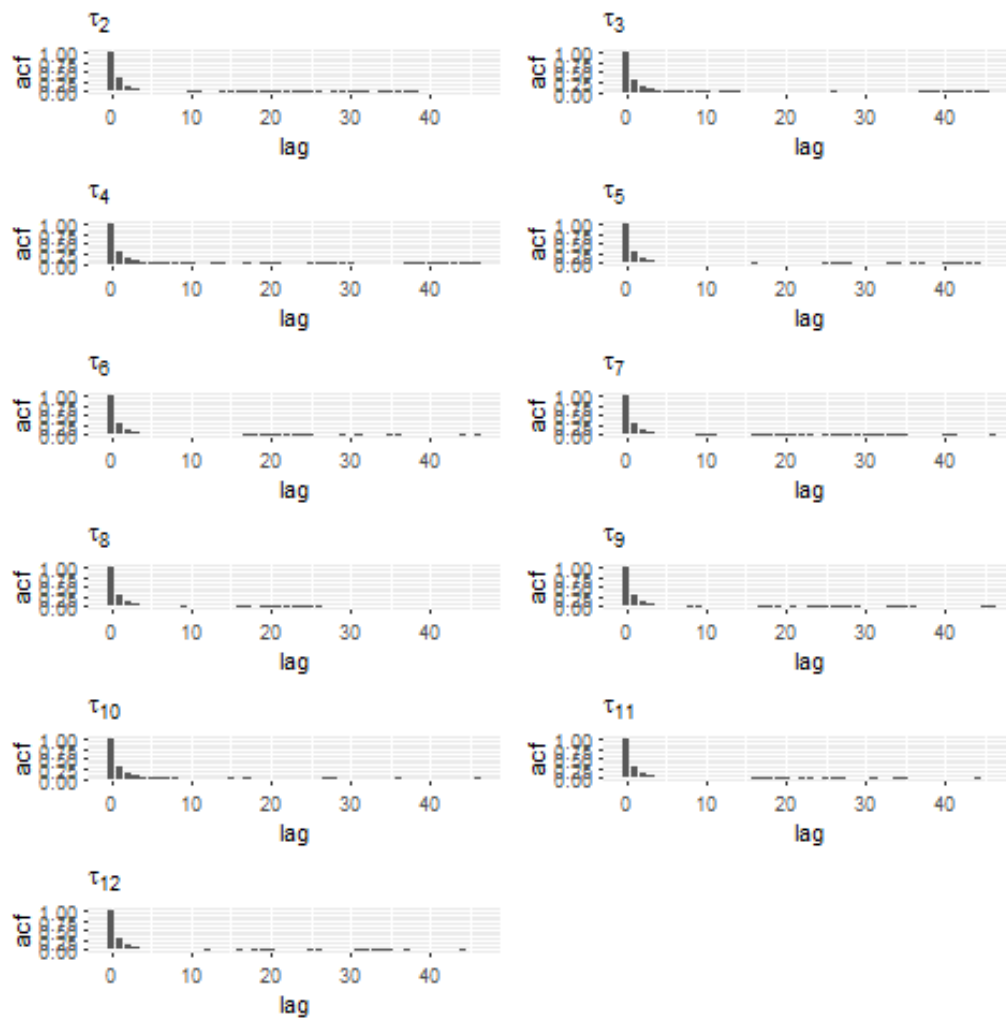
Figure C.5: Autocorrelation plots for $\tau_2, ..., \tau_{12}$ in Model 3.

Figure C.6: Autocorrelation plots for $\omega_2, ..., \omega_{23}$ in Model 3.
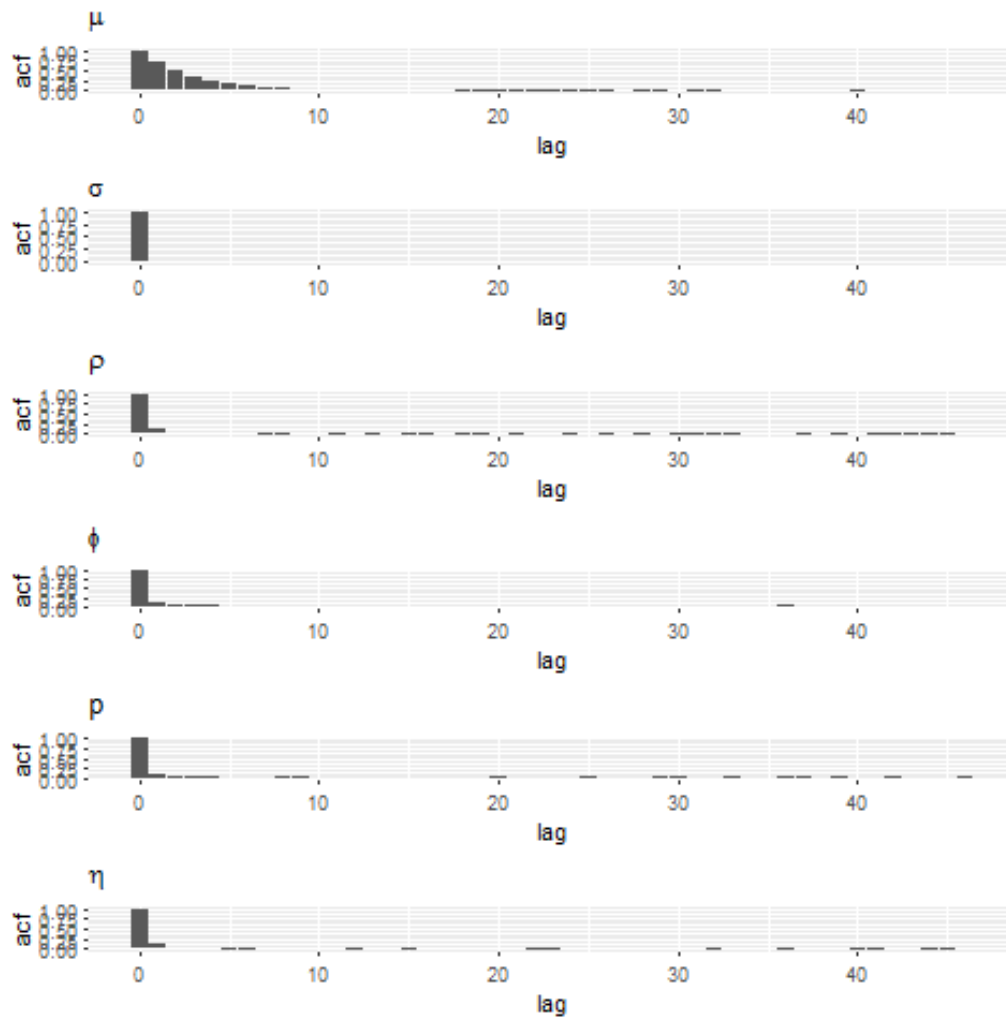
Figure C.7: Autocorrelation plots for μ, σ, ρ, ϕ, p, and η in Model 3.

C.4 Model 4: Point Referenced CPUE Model
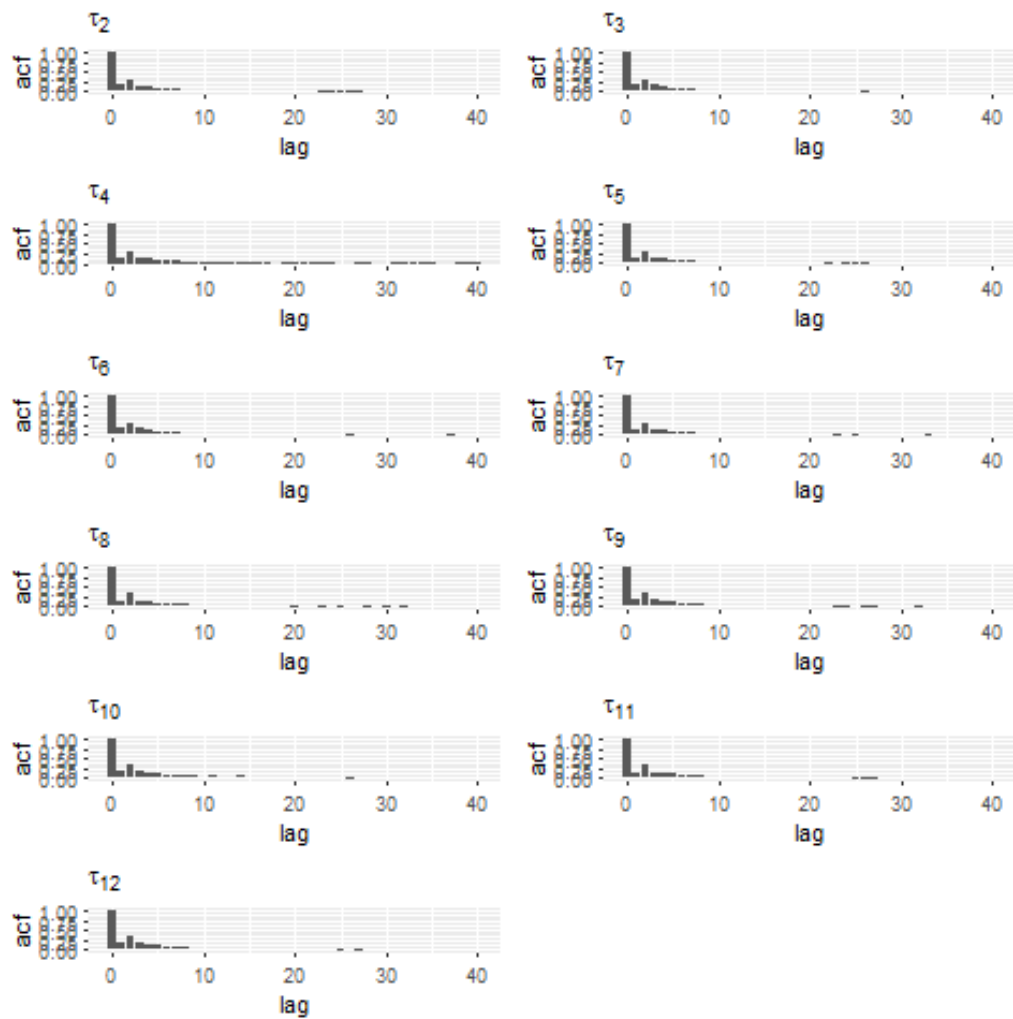
Figure C.8: Autocorrelation plots for $\tau_2, ..., \tau_{12}$ in Model 4.
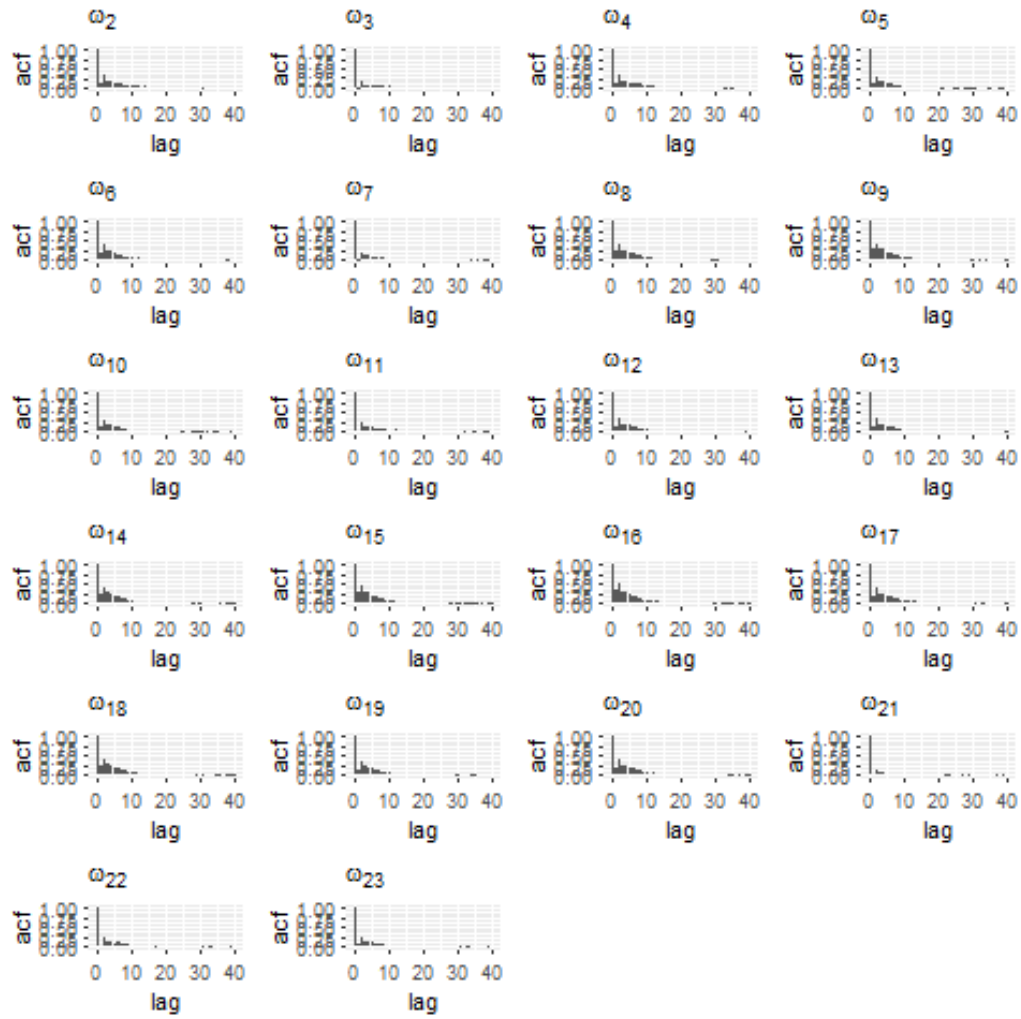
Figure C.9: Autocorrelation plots for $\omega_2, ..., \omega_{23}$ in Model 4.
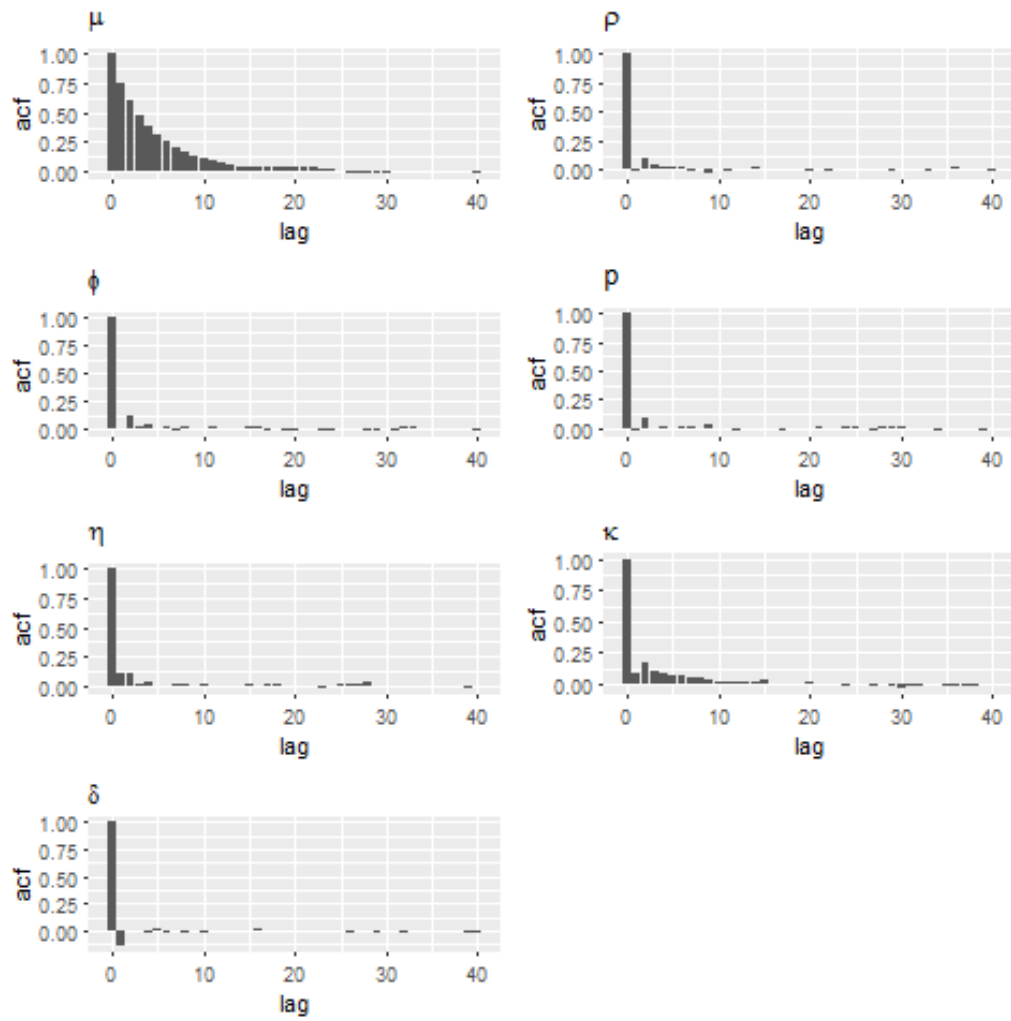
Figure C.10: Autocorrelation plots for μ, ρ, ϕ, p, η, κ, and δ in Model 4.

C.5 Model 5: Blocked Point Referenced CPUE Model
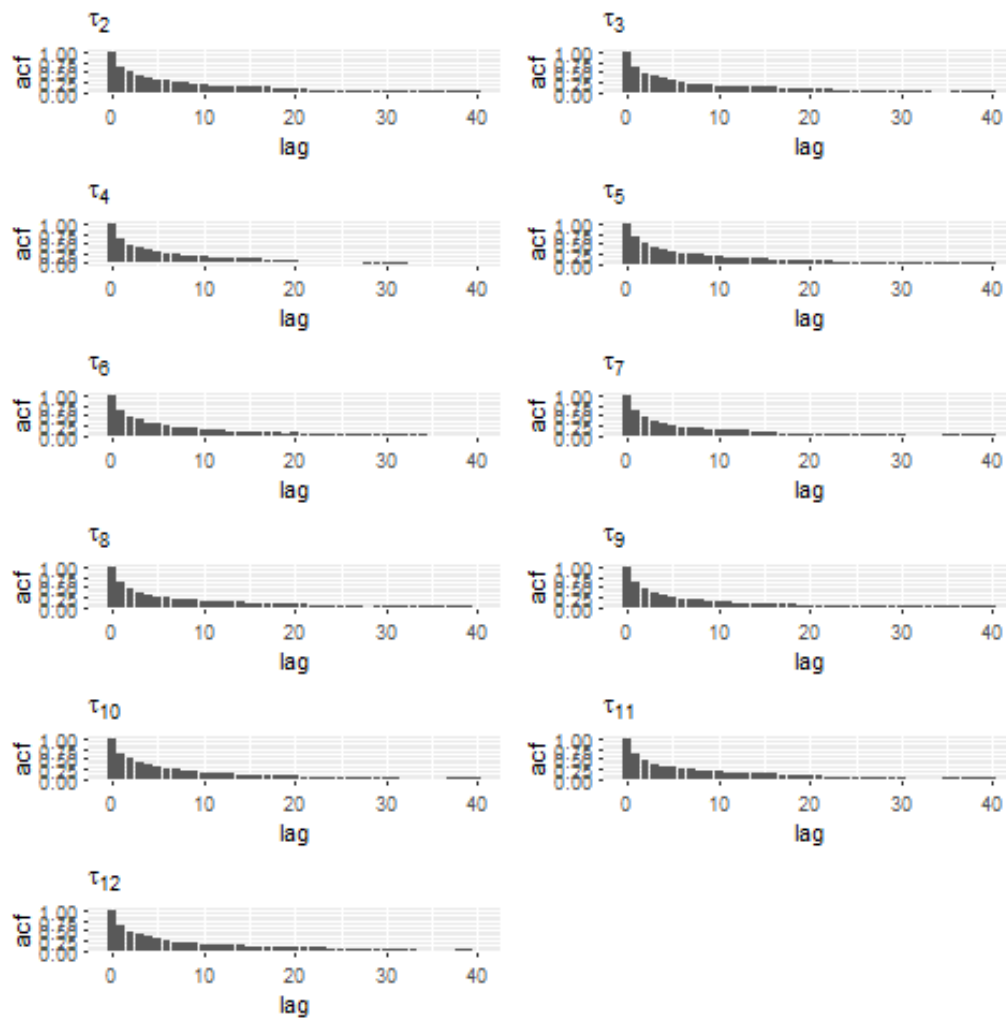
Figure C.11: Autocorrelation plots for $\tau_2, ..., \tau_{12}$ in Model 5.
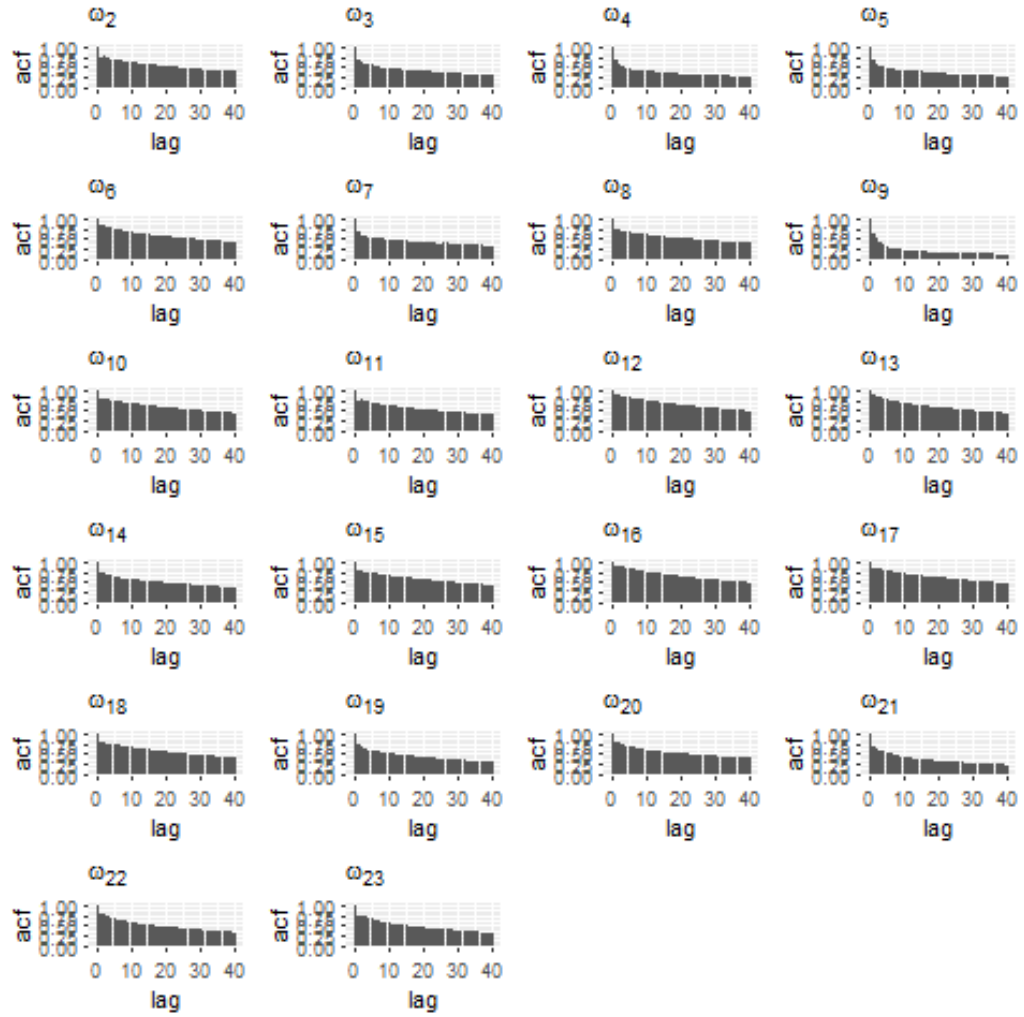
Figure C.12: Autocorrelation plots for $\omega_2, ..., \omega_{23}$ in Model 5.
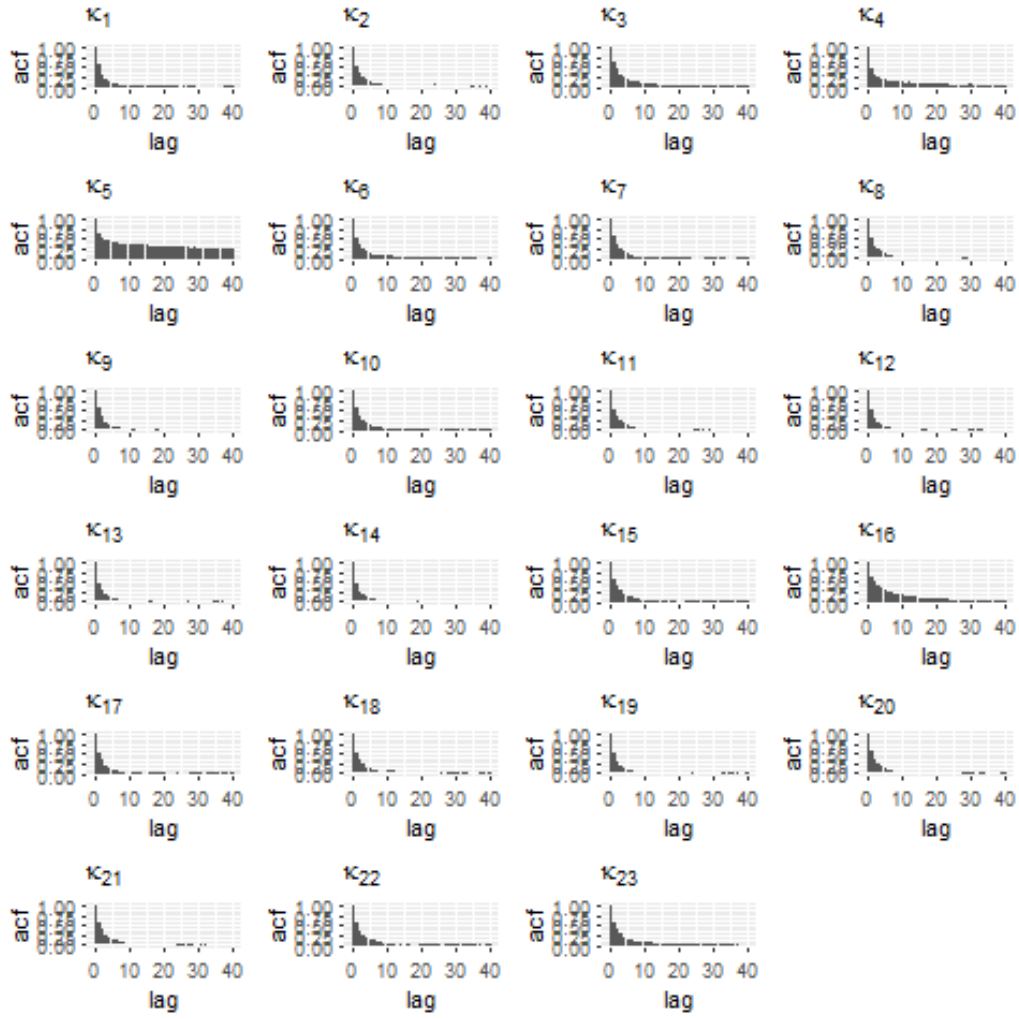
Figure C.13: Autocorrelation plots for $\kappa_1, ..., \kappa_{23}$ in Model 5.

Figure C.14: Autocorrelation plots for $\delta_1, ..., \delta_{23}$ in Model 5.
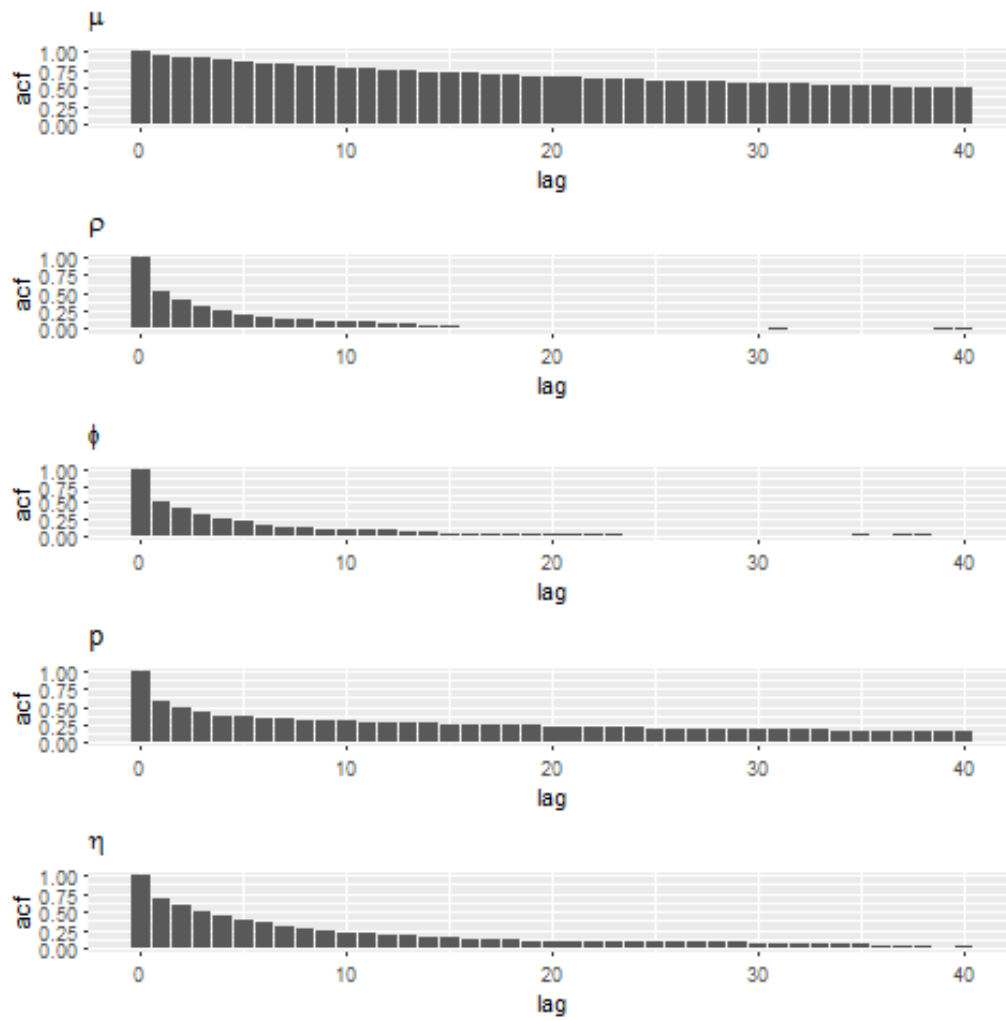
Figure C.15: Autocorrelation plots for μ, ρ, ϕ, p, and η in Model 5.

Appendix D

Gelman – Rubin Plots

D.1 Model 1: Temporal Effects
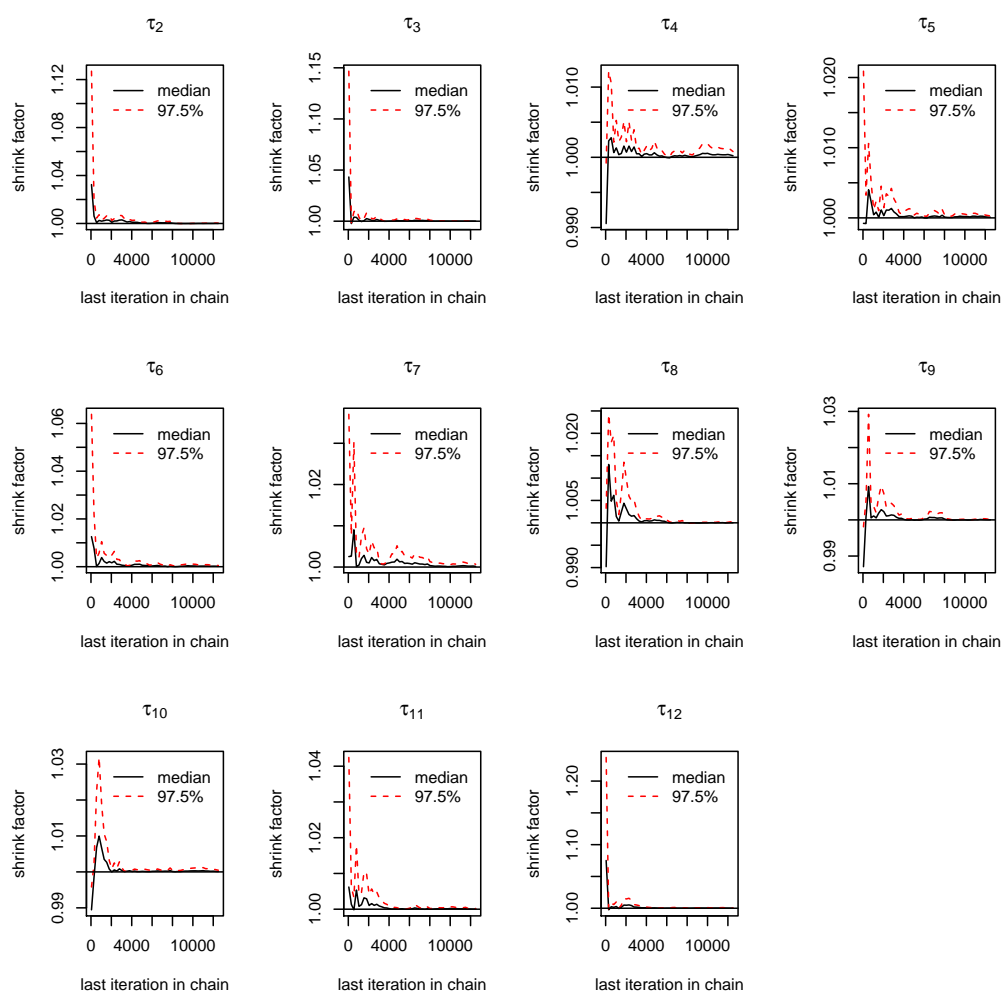
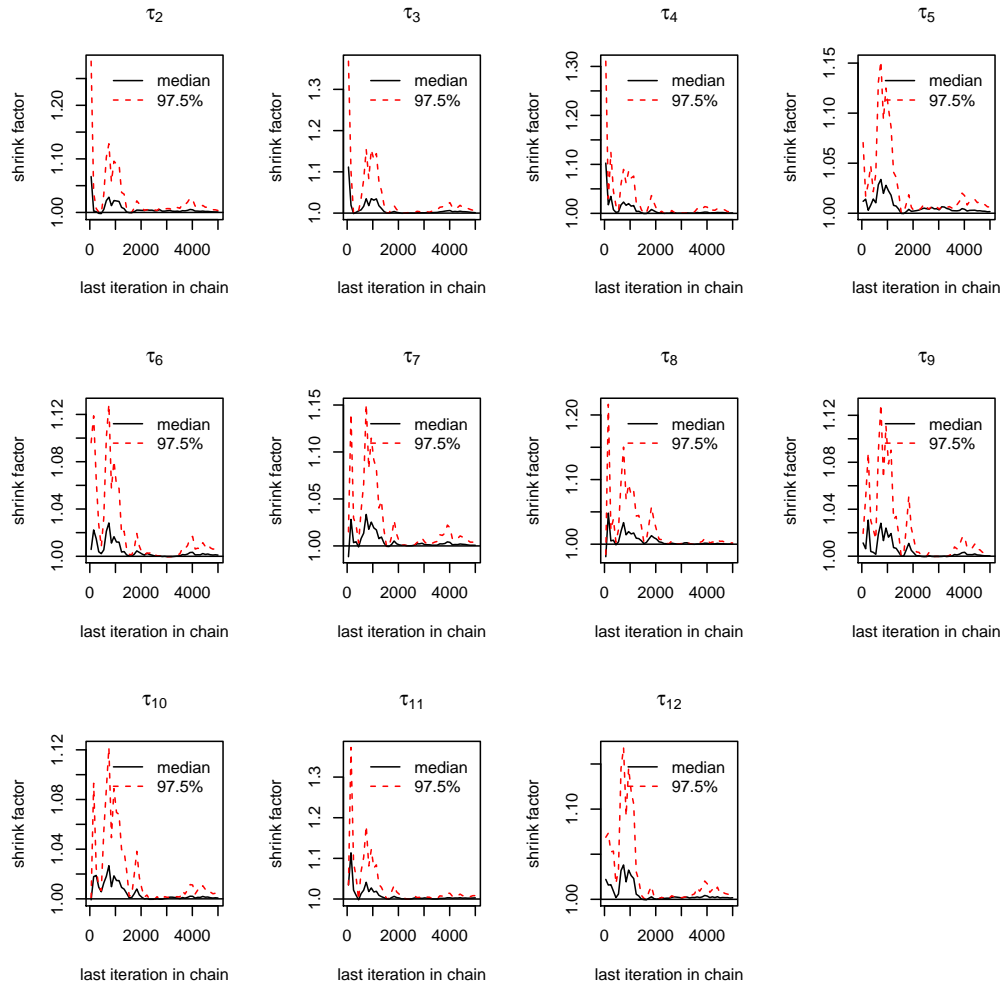Figure D.1: Gelman-Rubin plots for the parameters in Model 1.
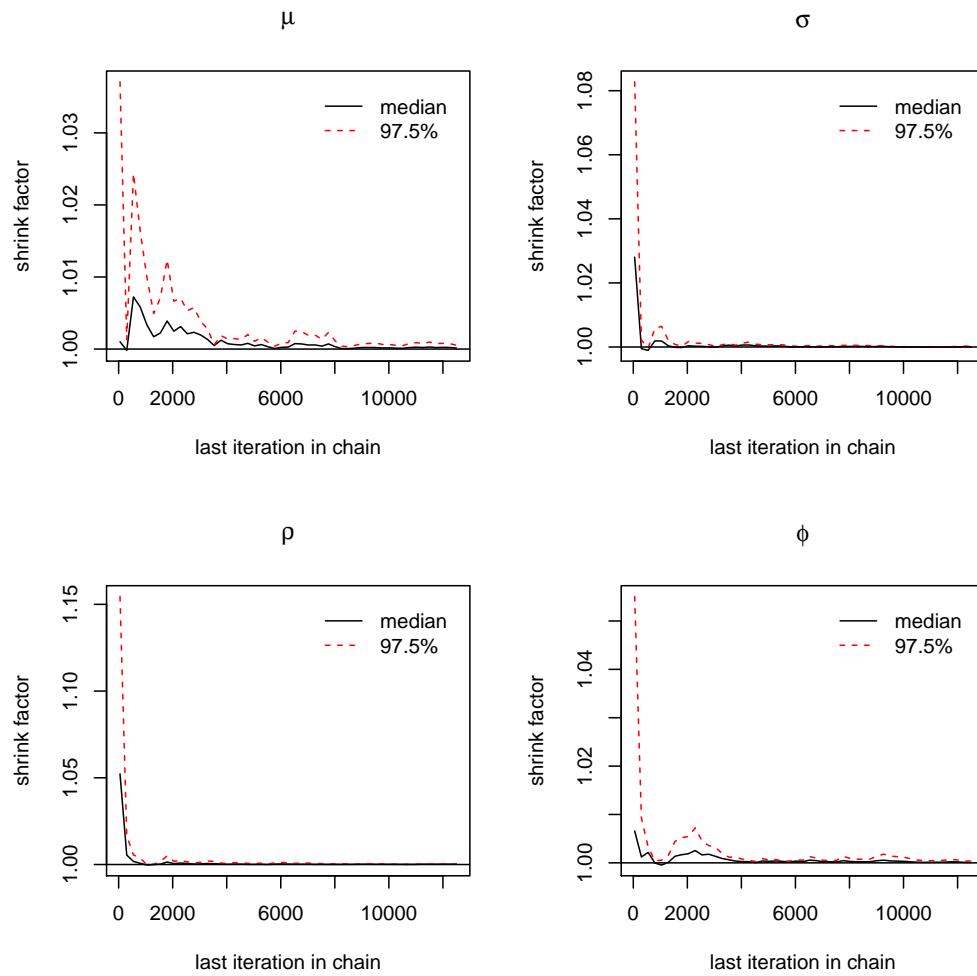
Figure D.2: Gelman-Rubin plots for the parameters in Model 1.

D.2   Model 2: Areal Spatial Effects
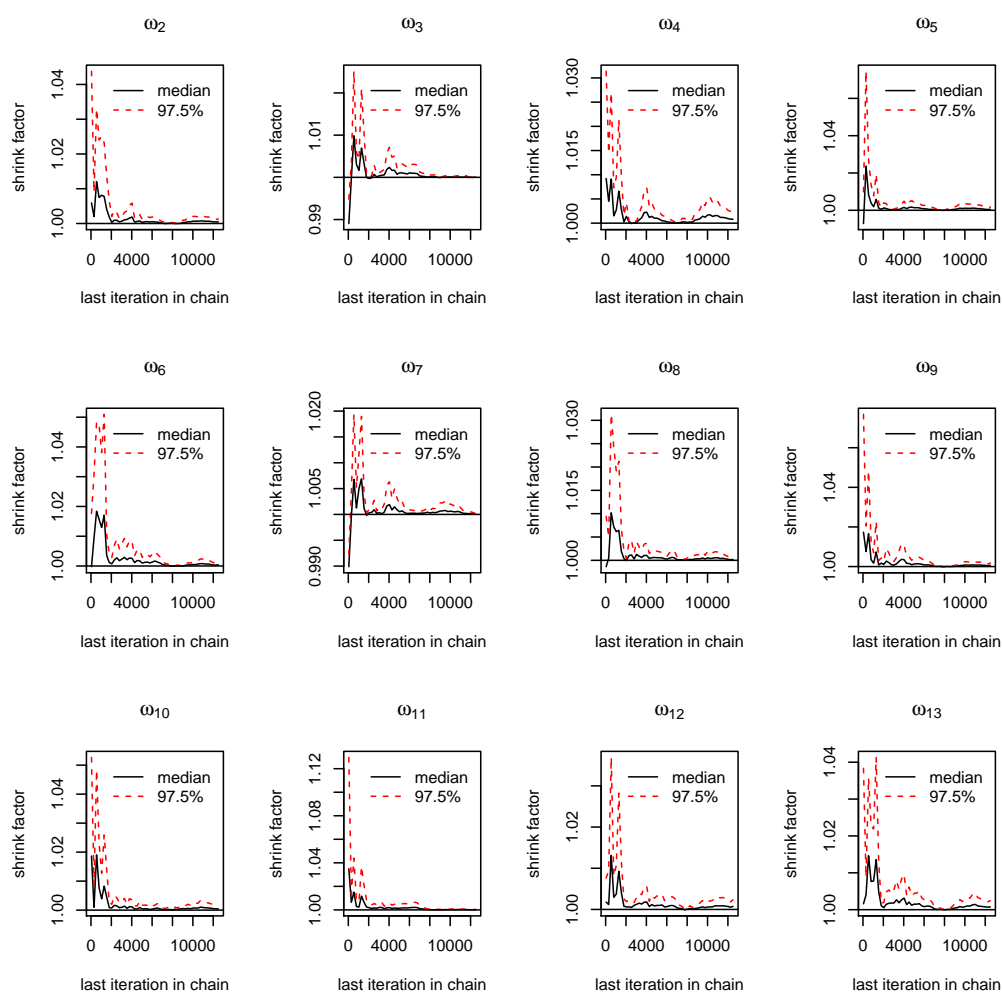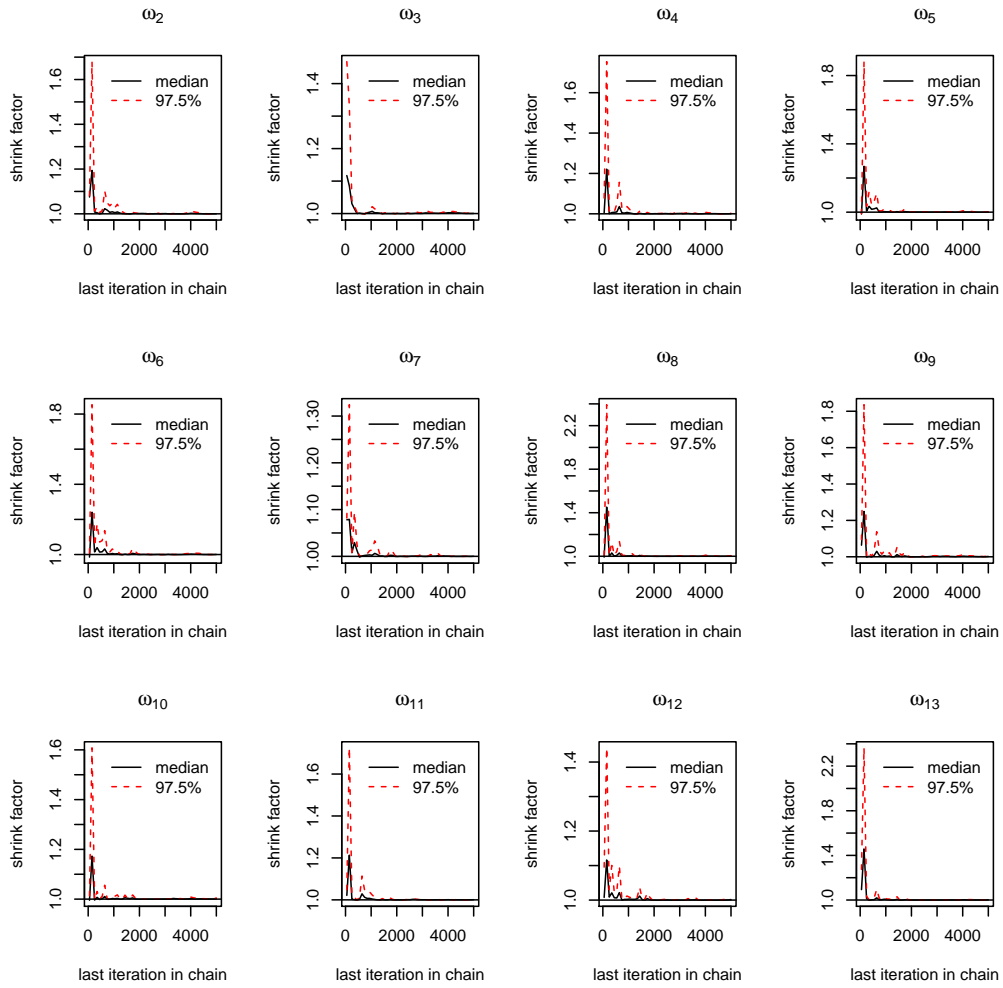

Figure D.3: Gelman-Rubin plots for the parameters in Model 2.
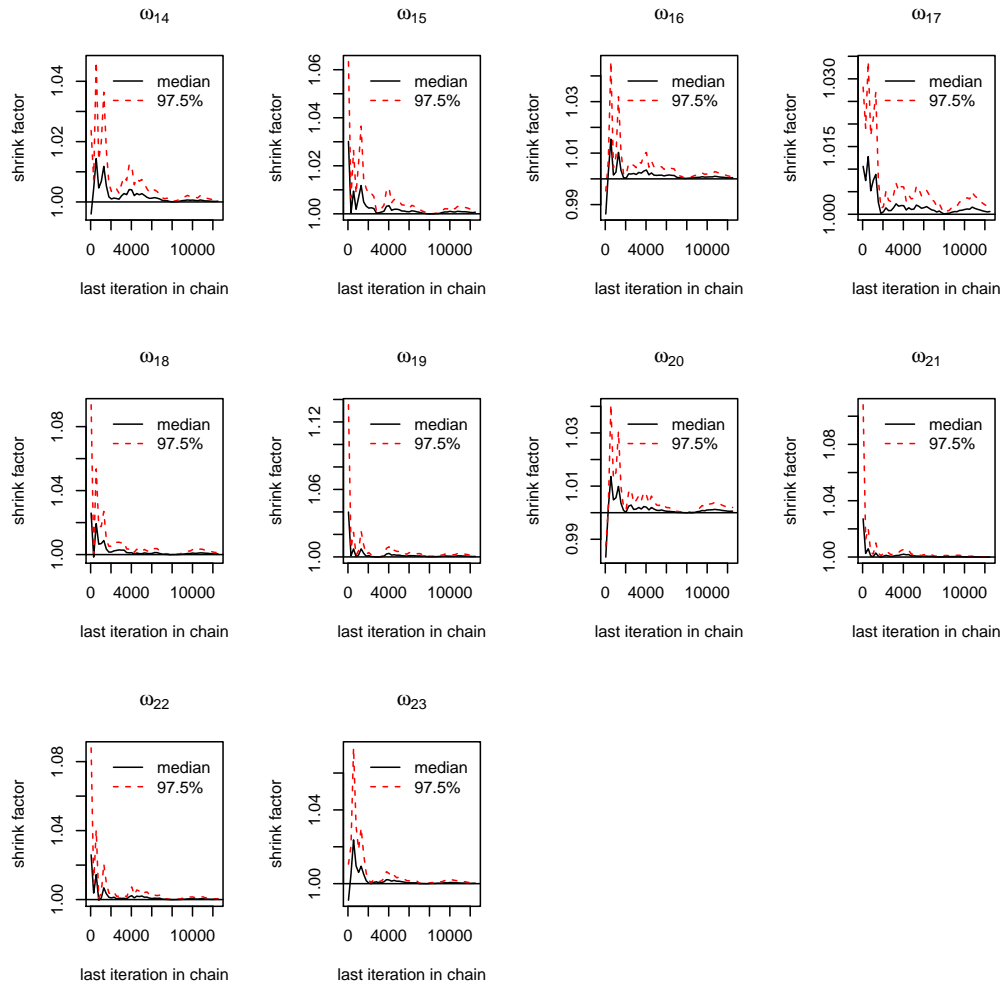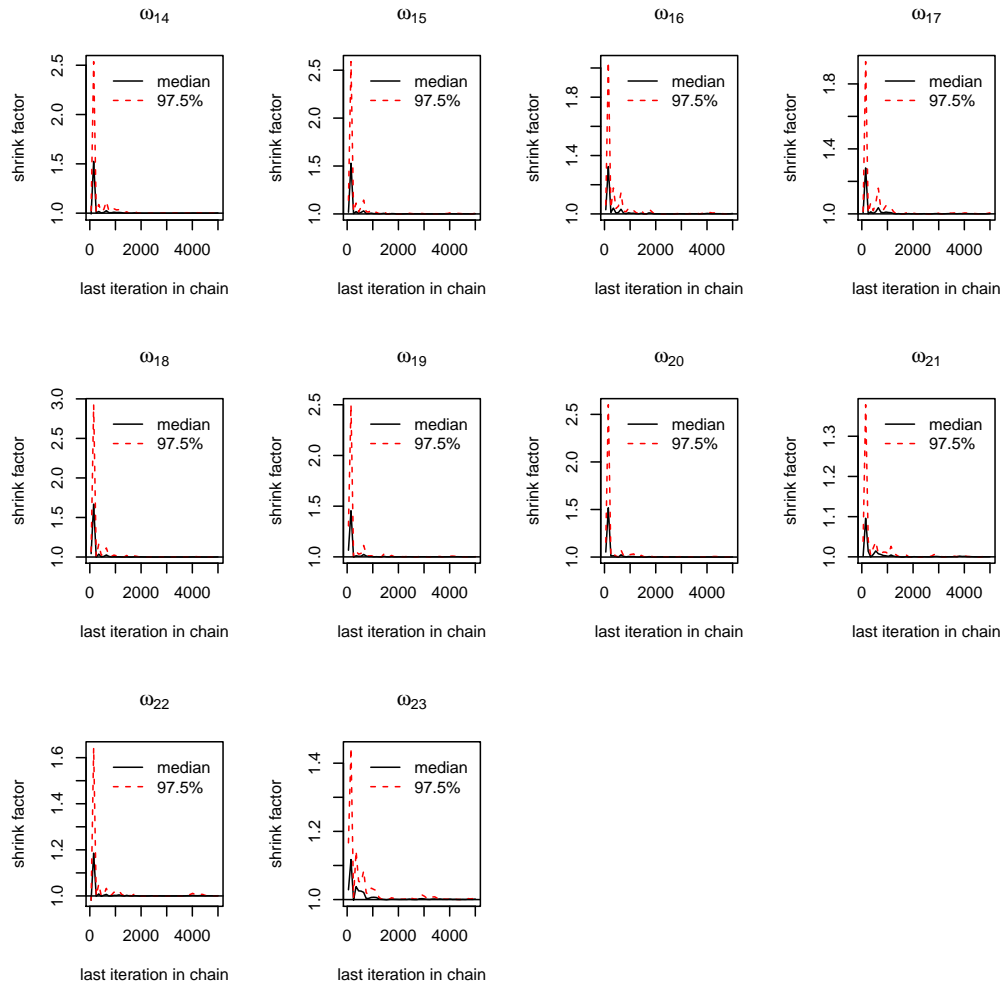
Figure D.4: Gelman-Rubin plots for the parameters in Model 2.
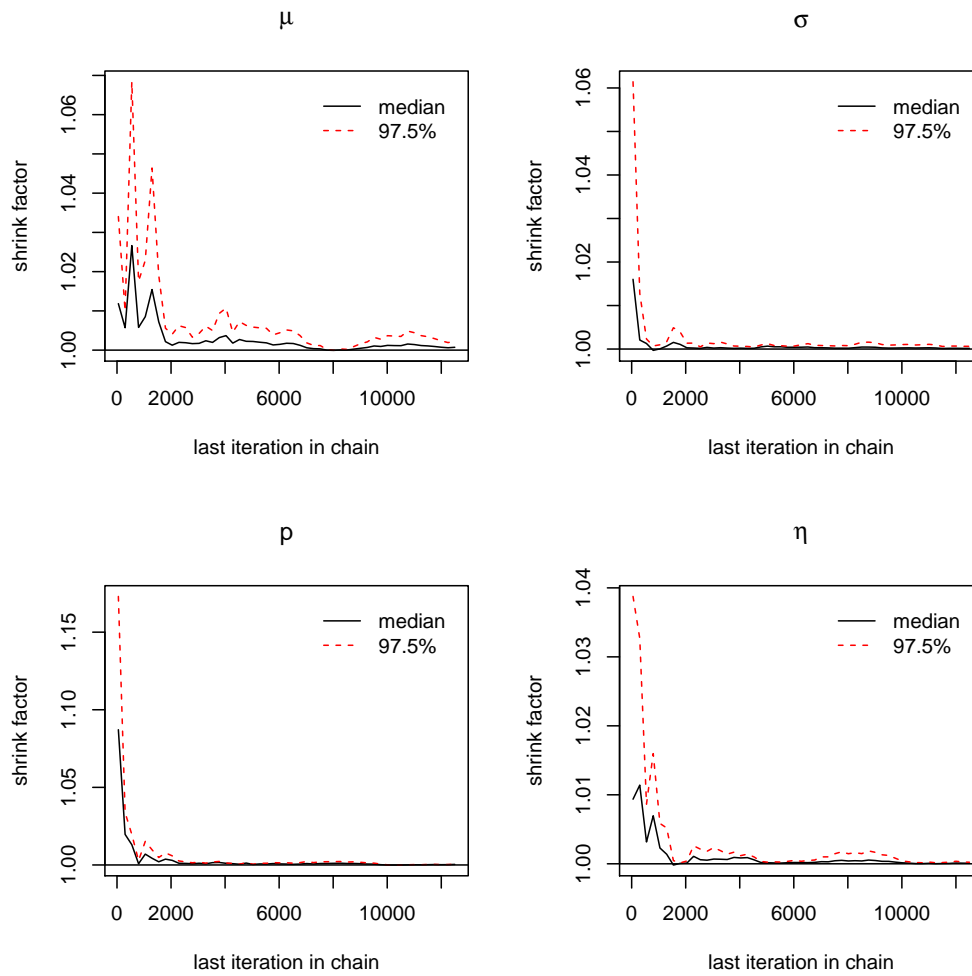
Figure D.5: Gelman-Rubin plots for the parameters in Model 2.

D.3 Model 3: Temporal and Areal Spatial Effects
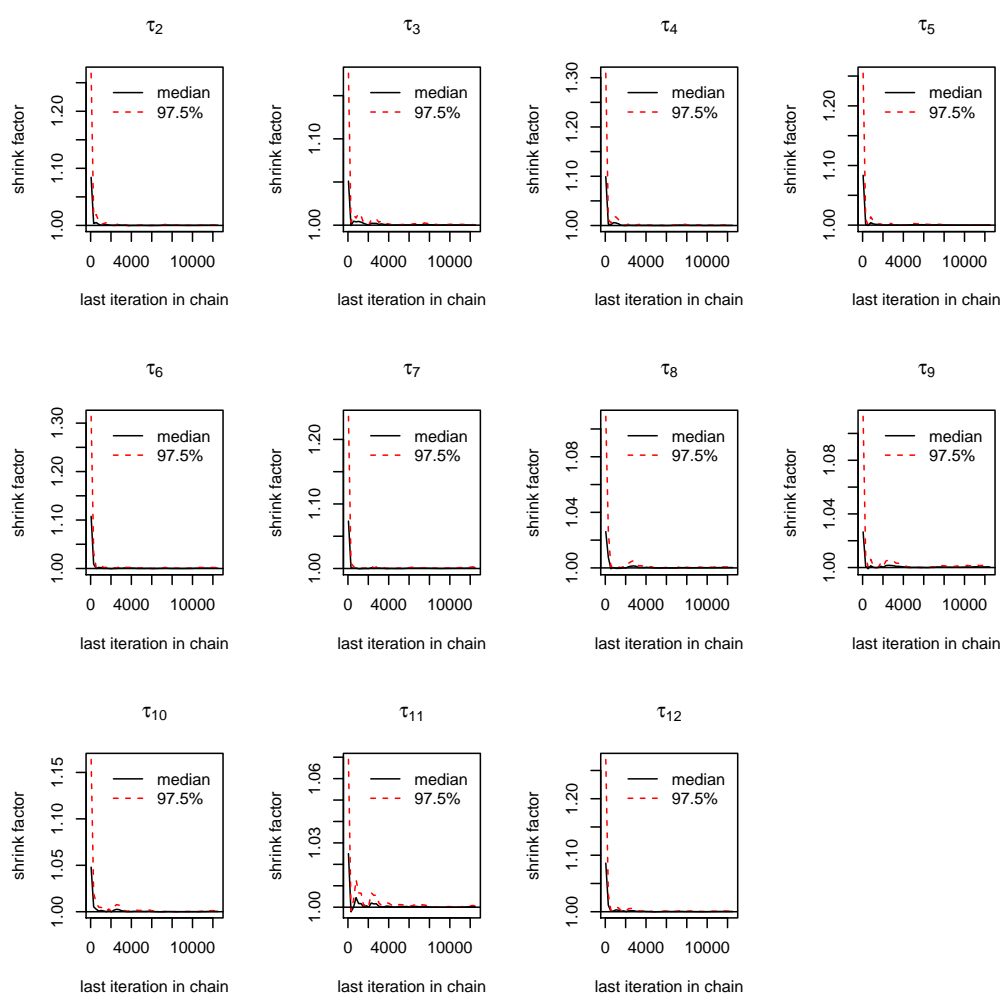
Figure D.6: Gelman-Rubin plots for the parameters in Model 3.
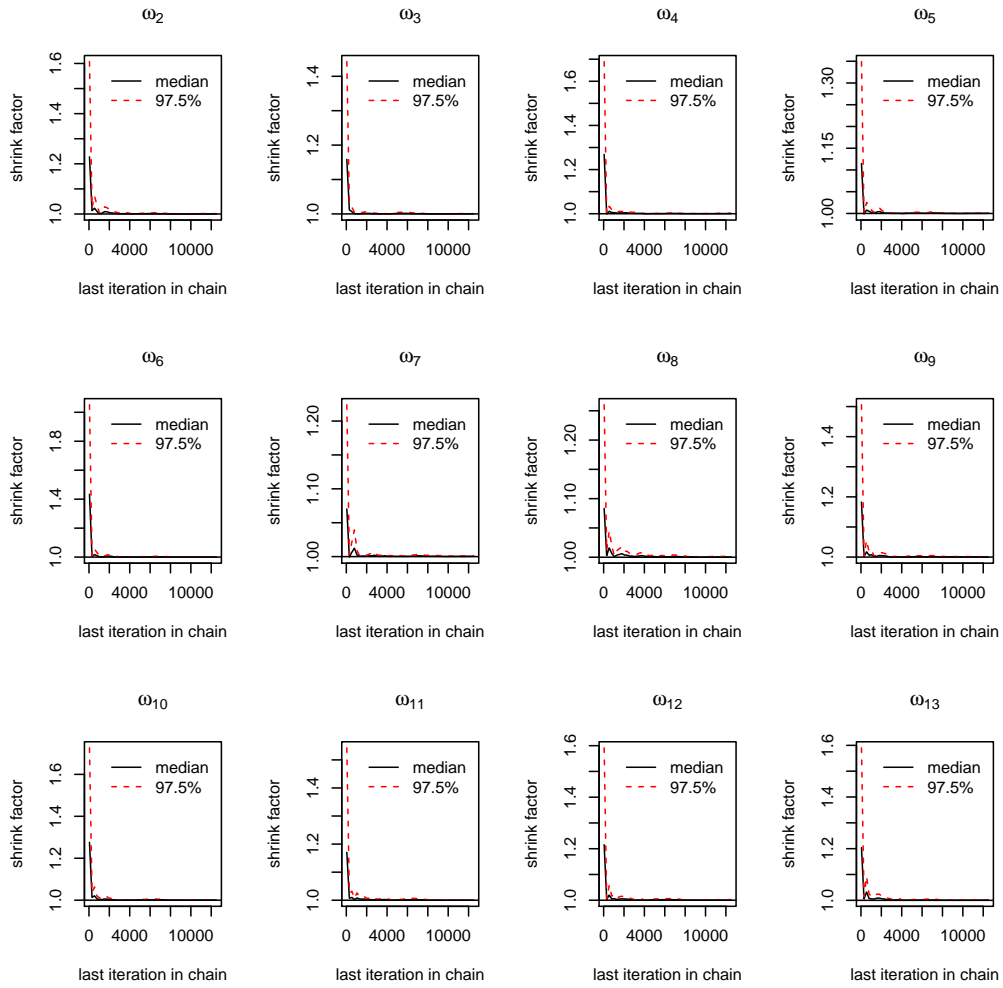
Figure D.7: Gelman-Rubin plots for the parameters in Model 3.
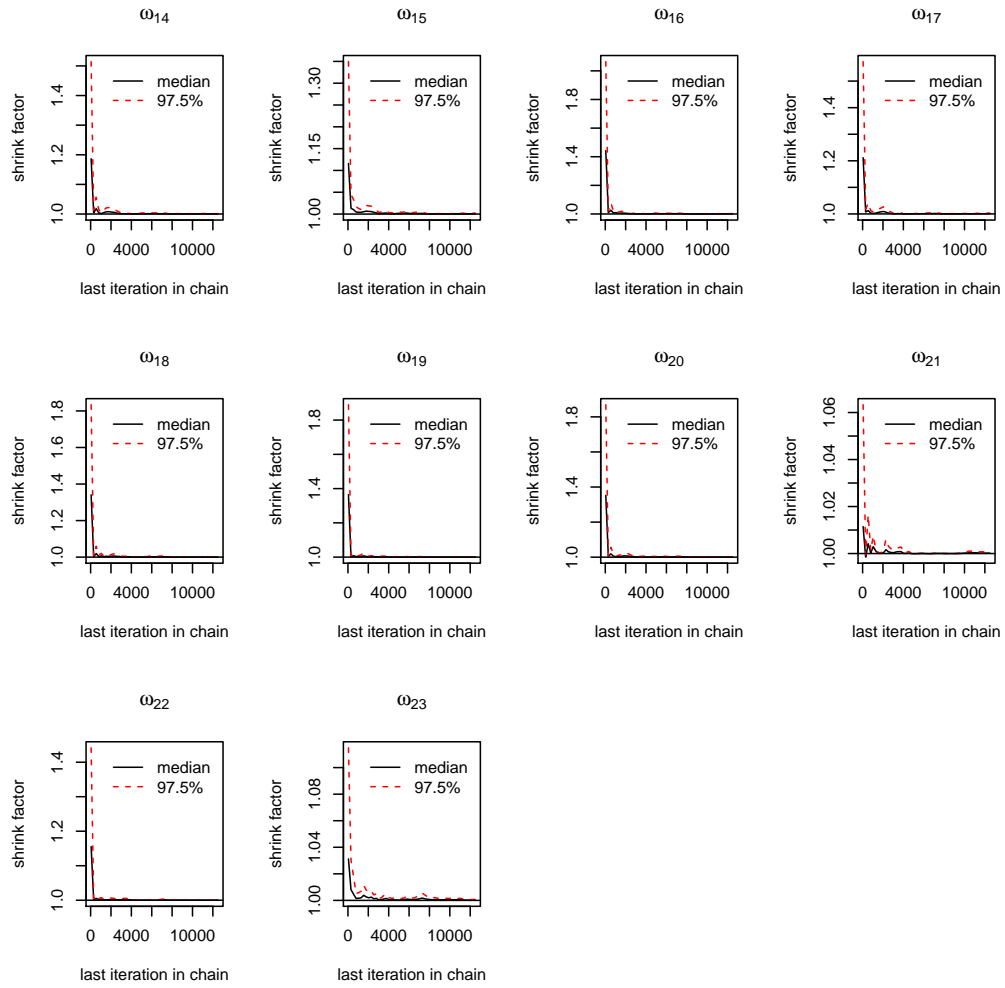
Figure D.8: Gelman-Rubin plots for the parameters in Model 3.
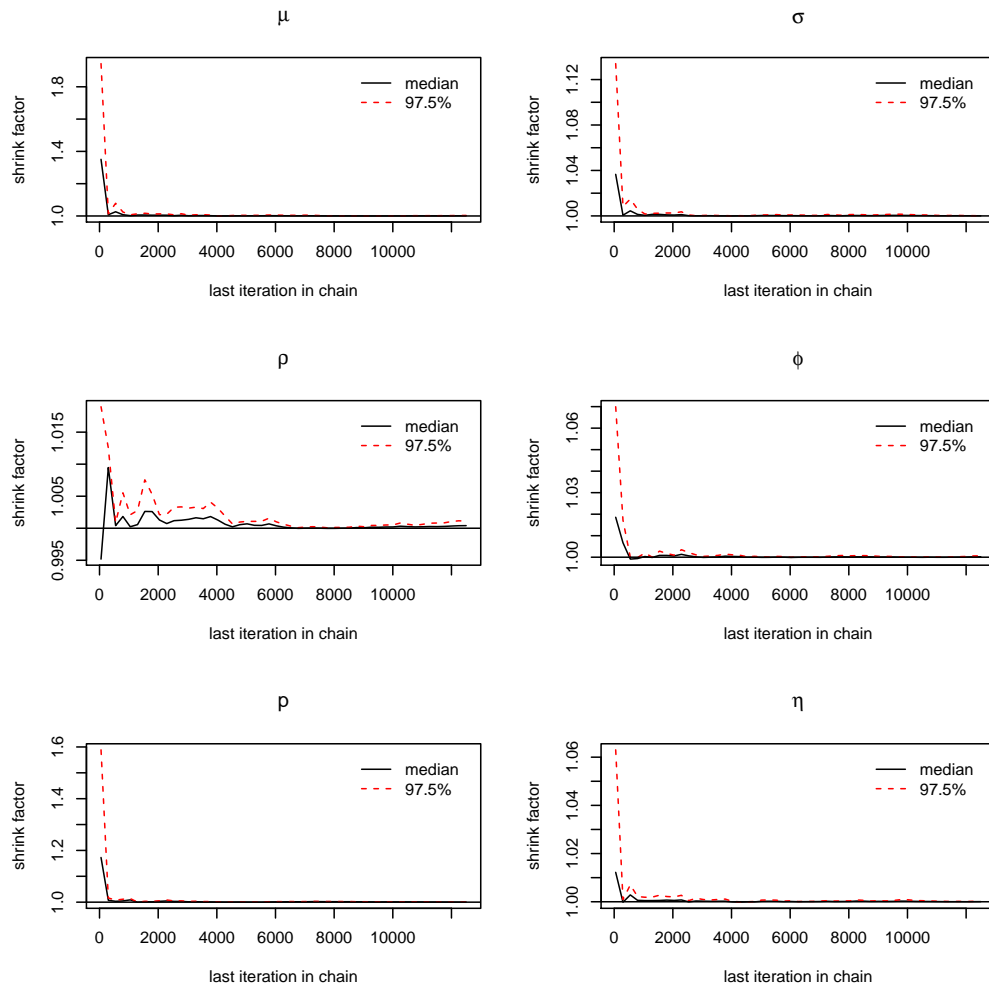
Figure D.9: Gelman-Rubin plots for the parameters in Model 3.

D.4 Model 4: Point Referenced CPUE Model

Figure D.10: Gelman-Rubin plots for the parameters in Model 4.

Figure D.11: Gelman-Rubin plots for the parameters in Model 4.

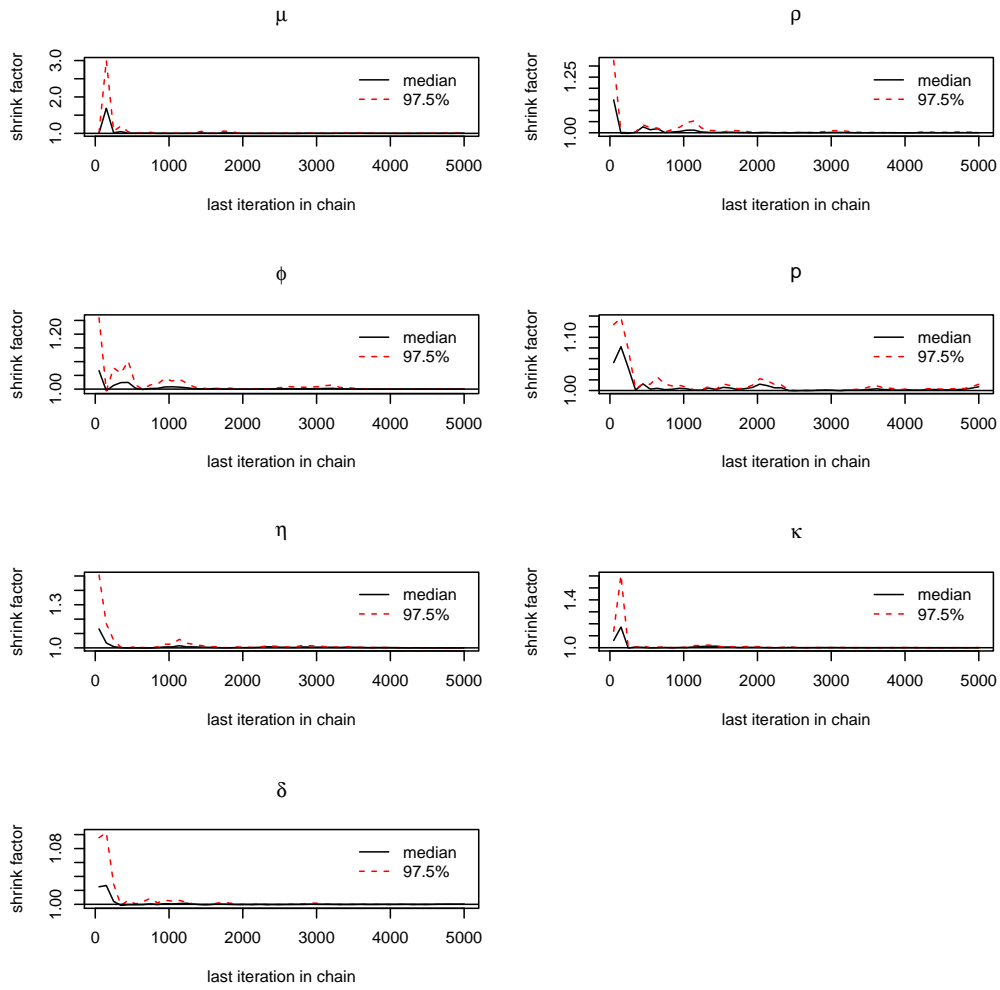Figure D.12: Gelman-Rubin plots for the parameters in Model 4.

Figure D.13: Gelman-Rubin plots for the parameters in Model 4.

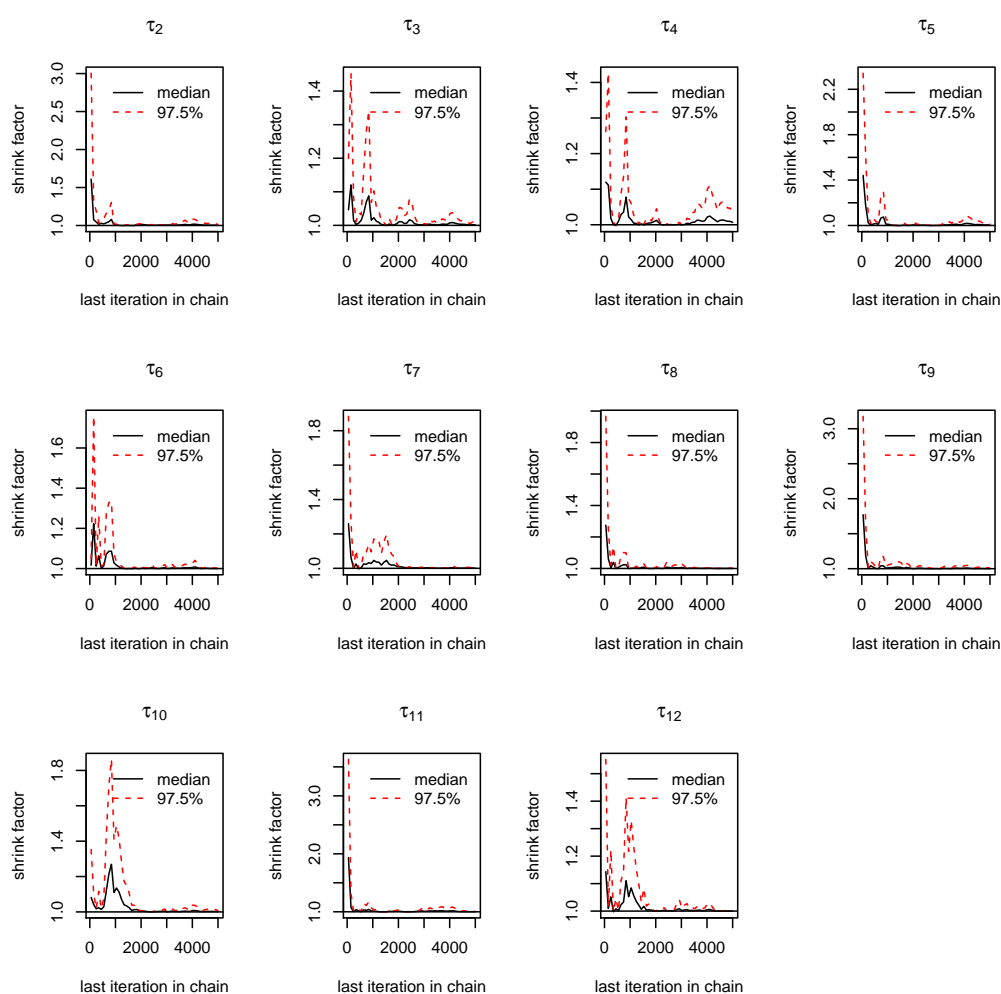## D.5   Model 5: Blocked Point Referenced CPUE Model

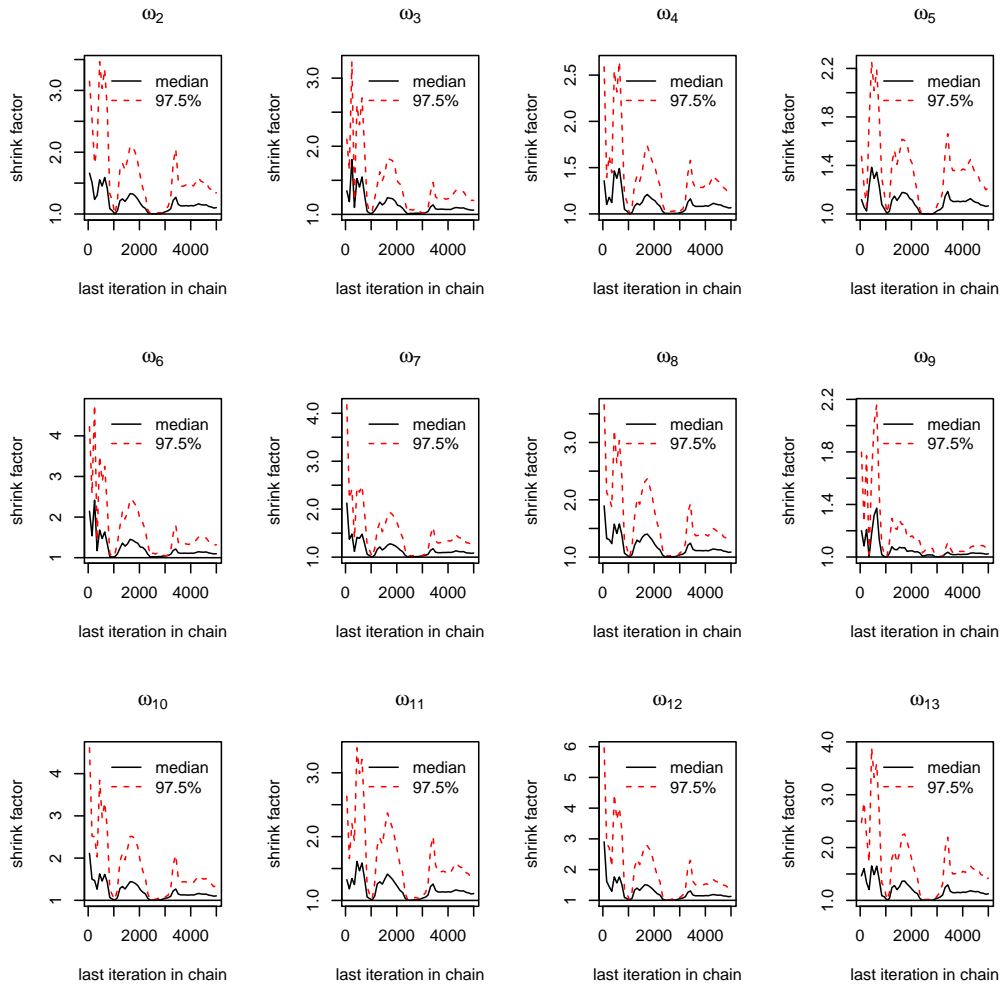Figure D.14: Gelman-Rubin plots for the parameters in Model 5.

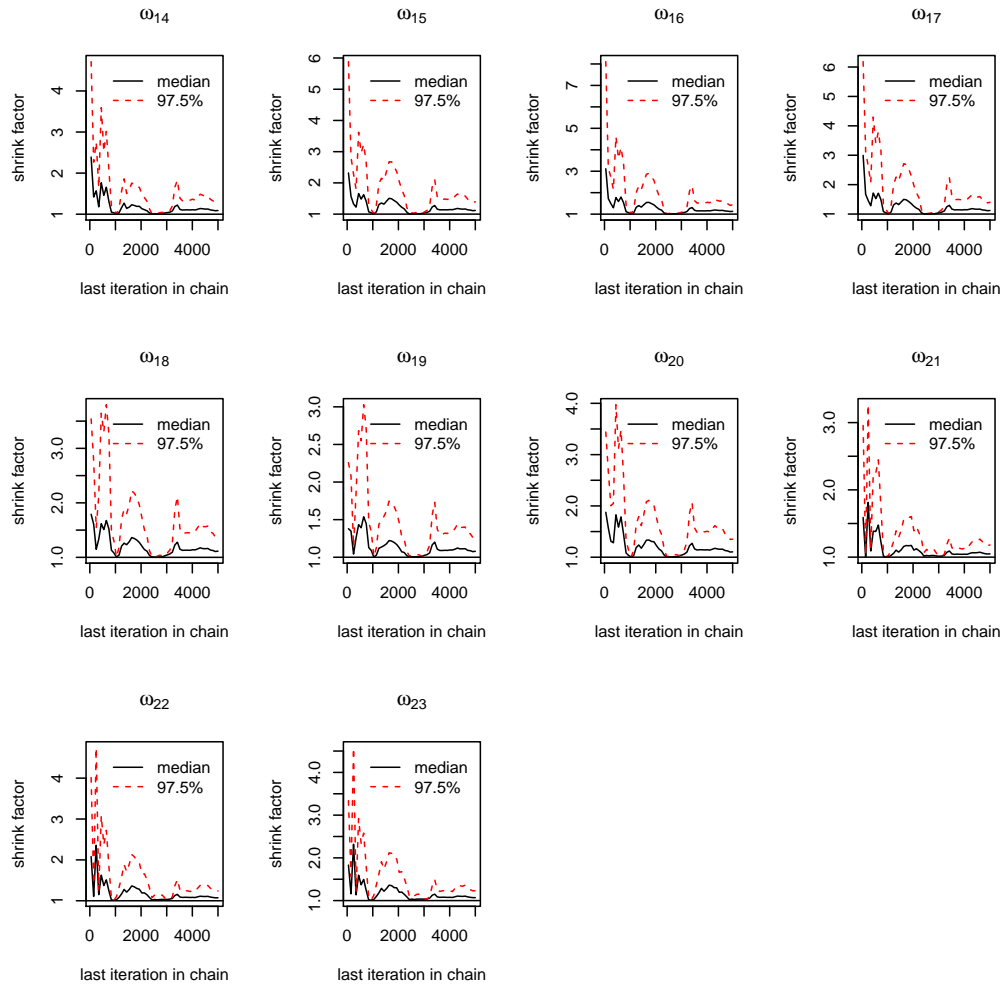Figure D.15: Gelman-Rubin plots for the parameters in Model 5.

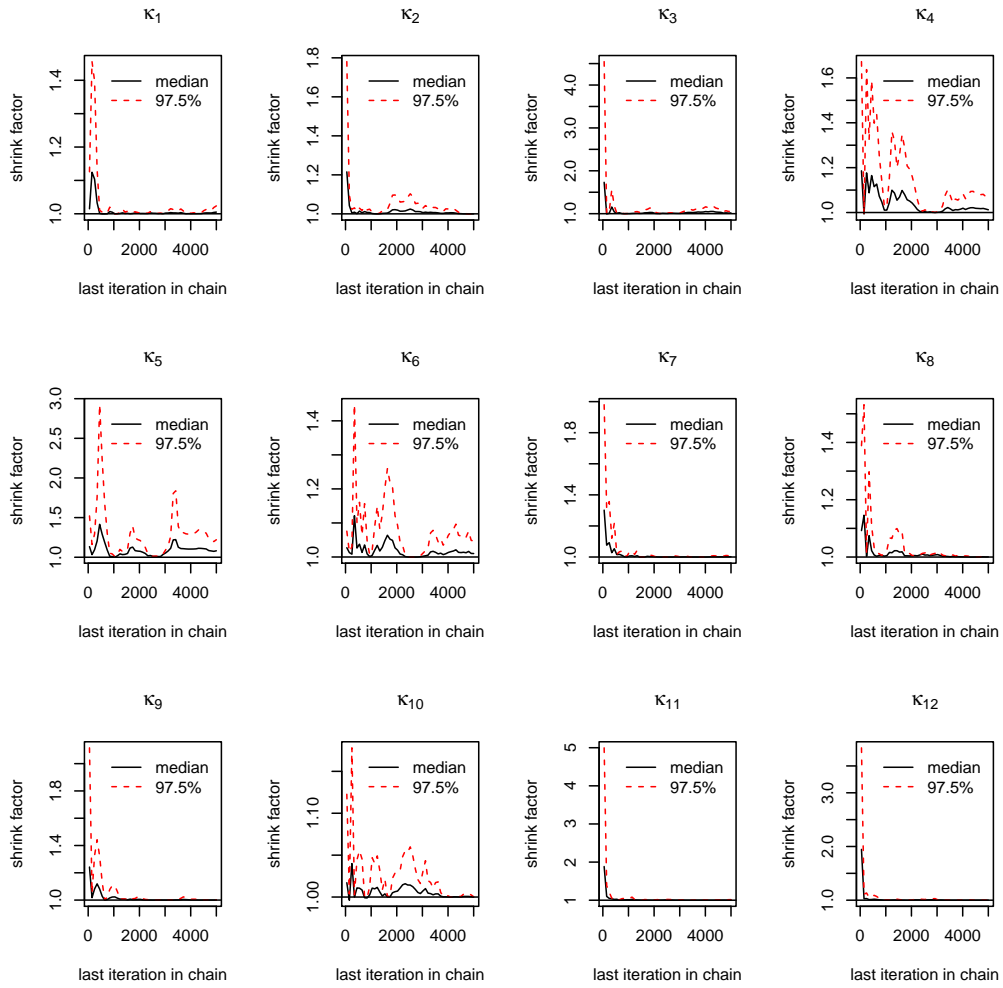Figure D.16: Gelman-Rubin plots for the parameters in Model 5.

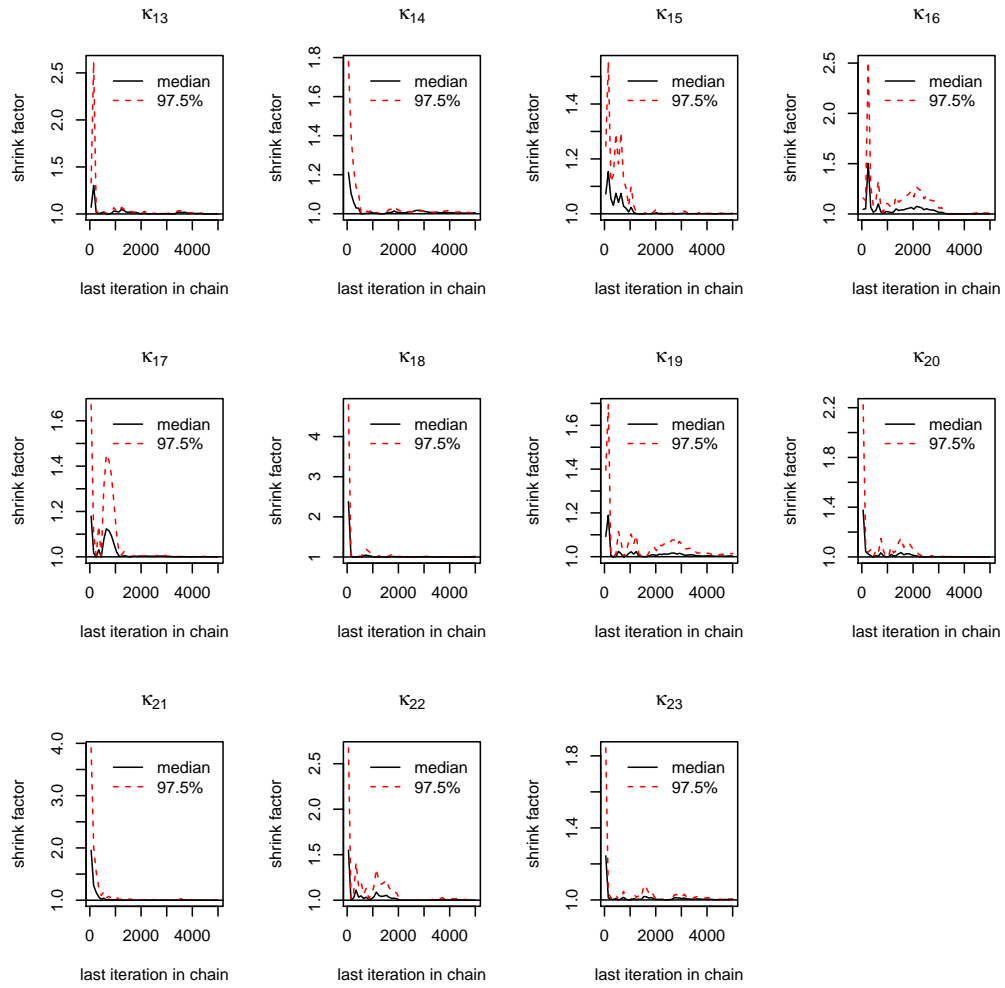Figure D.17: Gelman-Rubin plots for the parameters in Model 5.

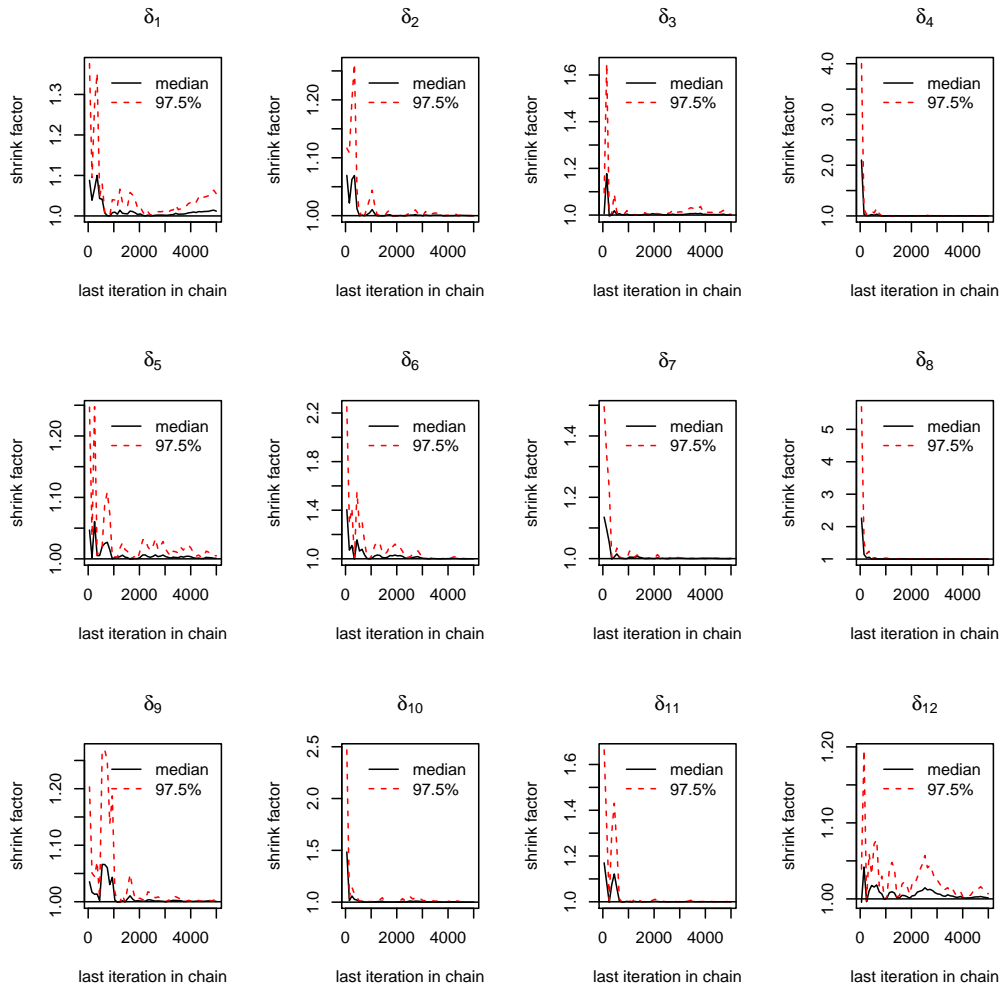Figure D.18: Gelman-Rubin plots for the parameters in Model 5.

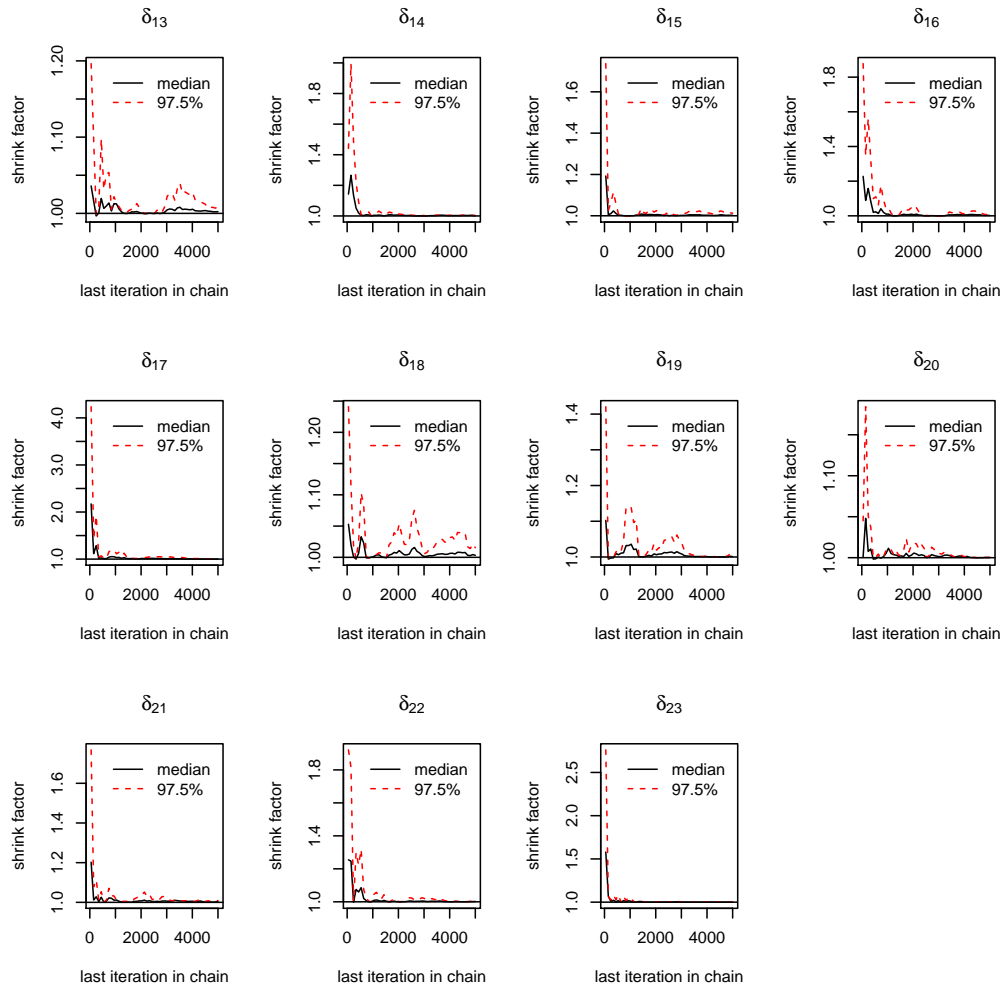Figure D.19: Gelman-Rubin plots for the parameters in Model 5.

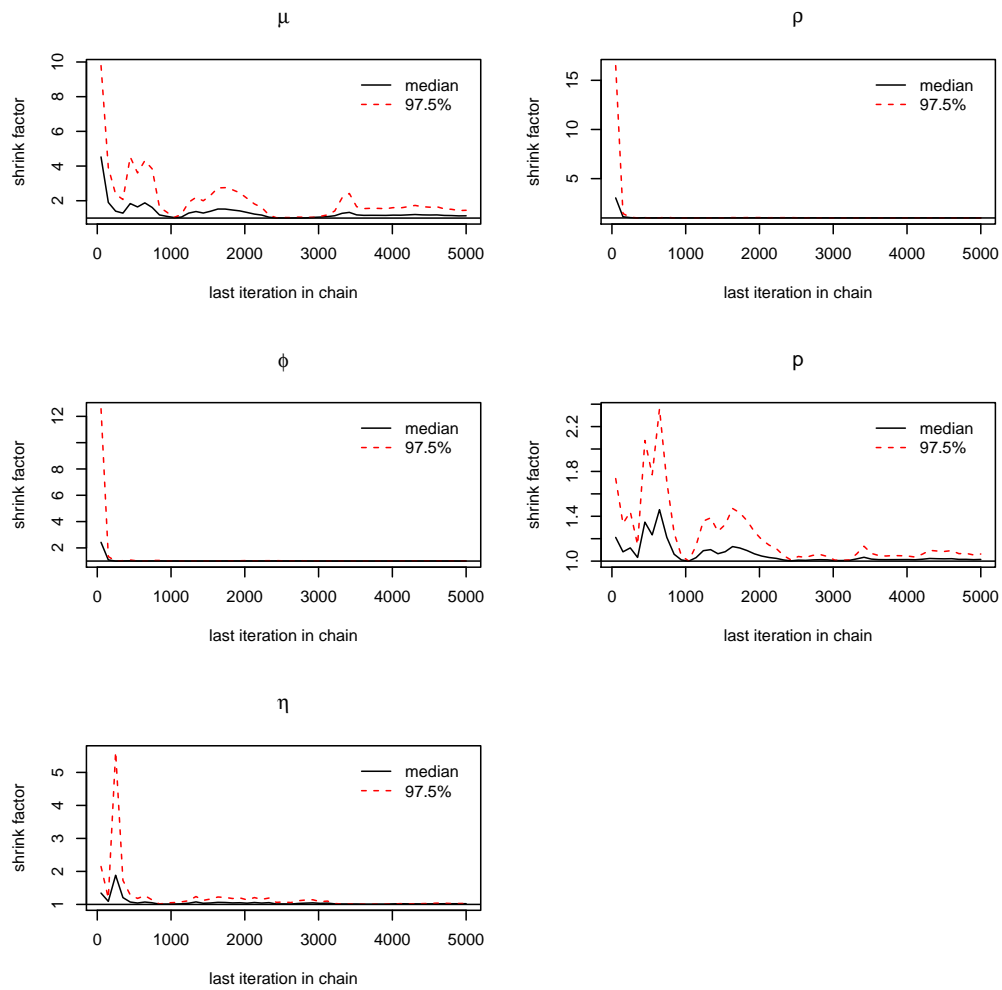Figure D.20: Gelman-Rubin plots for the parameters in Model 5.

Figure D.21: Gelman-Rubin plots for the parameters in Model 5.

Bibliography

Abensperg-Traun, M. & Boer, E. S. (1990). Species abundance and habitat differences in biomass of subterranean termites (isoptera) in the wheatbelt of western australia. *Australian Journal of Ecology*, **15**(2), 219–226.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Proceedings of the Second International Symposium on Information Theory* (pp. 267–281).

Allcroft, D. J. & Glasbey, C. A. (2003). A latent gaussian markov random-field model for spatiotemporal rainfall disaggregation. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **52**(4), 487–498.

Alqallaf, F. & Gustafson, P. (2001). On cross-validation of bayesian models. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique*, , 333–340.

Anselin, L. (1995). Local indicators of spatial associationlisa. *Geographical analysis*, **27**(2), 93–115.

Bagley, N. W., Ballara, S. L., ODriscoll, R. L., Fu, D., & Lyon, W. (2013). A review of hoki and middle depth summer trawl surveys of the subantarctic, november december 1991–1993 and 2000–2009. *New Zealand Fisheries Assessment Report*, **41**, 63.

Baranov, F. I. (1918). *On the question of the biological basis of fisheries*.

Beare, D., Batten, S., Edwards, M., McKenzie, E., Reid, P., & Reid, D. (2003). Summarising spatial and temporal information in cpr data. *Progress in Oceanography*, **58**(2), 217–233.

Bertram, D. F., Drever, M. C., McAllister, M. K., Schroeder, B. K., Lindsay, D. J., & Faust, D. A. (2015). Estimation of coast-wide population trends of marbled murrelets in canada using a bayesian hierarchical model. *PloS one*, **10**(8).

Besag, J., York, J., & Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the institute of statistical mathematics*, **43**(1), 1–20.

Beverton, R. J. (1957). On the dynamics of exploited fish populations vol-19. .

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in ecology & evolution*, **24**(3), 127–135.

Briggs, R. (2010). Global measures of spatial autocorrelation. *Henan University*, .

Campbell, R. A. (2004). Cpue standardisation and the construction of indices of stock abundance in a spatially varying fishery using general linear models. *Fisheries Research*, **70**(2), 209–227.

Chen, Y. (2012). On the four types of weight functions for spatial contiguity matrix. *Letters in Spatial and Resource Sciences*, **5**(2), 65–72.

Chen, Y. (2013). New approaches for calculating morans index of spatial autocorrelation. *PloS one*, **8**(7).

Cheng, T., Haworth, J., & Wang, J. (2012). Spatio-temporal autocorrelation of road network data. *Journal of Geographical Systems*, **14**(4), 389–413.

Cooper, A. B., Rosenberg, A. A., Stefánsson, G., & Mangel, M. (2004). Examining the importance of consistency in multi-vessel trawl survey design based on the us west coast groundfish bottom trawl survey. *Fisheries research*, **70**(2), 239–250.

Dakos, V., Scheffer, M., van Nes, E. H., Brovkin, V., Petoukhov, V., & Held, H. (2008). Slowing down as an early warning signal for abrupt climate change. *Proceedings of the National Academy of Sciences*, **105**(38), 14308–14312.

Efron, B. (1981). Nonparametric estimates of standard error: the jackknife, the bootstrap and other methods. *Biometrika*, , 589–599.

Ferreira, M. A. & De Oliveira, V. (2007). Bayesian reference analysis for gaussian markov random fields. *Journal of Multivariate Analysis*, **98**(4), 789–812.

Francis, M. P., Hurst, R. J., McArdle, B. H., Bagley, N. W., & Anderson, O. F. (2002). New zealand demersal fish assemblages. *Environmental Biology of Fishes*, **65**(2), 215–234.

Francis, R. (1984). An adaptive strategy for stratified random trawl surveys. *New Zealand Journal of Marine and Freshwater Research*, **18**(1), 59–71.

Fujita, Y. (2016). Clustering and classification in fisheries. .

Gavaris, S. (1980). Use of a multiplicative model to estimate catch rate and effort from commercial data. *Canadian Journal of Fisheries and Aquatic Sciences*, **37**(12), 2272–2275.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014a). *Bayesian data analysis*, volume 2. Chapman & Hall/CRC Boca Raton, FL, USA.

Gelman, A., Hwang, J., & Vehtari, A. (2014b). Understanding predictive information criteria for bayesian models. *Statistics and Computing*, **24**(6), 997–1016.

Gelman, A., Lee, D., & Guo, J. (2015). Stan a probabilistic programming language for bayesian inference and optimization. *Journal of Educational and Behavioral Statistics*, , 1076998615606113.

Gibbs, J. W. (1885). *On the fundamental formula of statistical mechanics, with applications to astronomy and thermodynamics*. Salem Press.

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, **106**(494), 746–762.

Gratwicke, B. & Speight, M. (2005). The relationship between fish species richness, abundance and habitat complexity in a range of shallow tropical marine habitats. *Journal of fish biology*, **66**(3), 650–667.

Haddon, M. (2010). *Modelling and quantitative methods in fisheries*. CRC press.

Haran, M. (2011). Gaussian random field models for spatial data. *Handbook of Markov Chain Monte Carlo*, , 449–478.

Hastie, T. & Tibshirani, R. (1986). Generalized additive models. *Statistical science*, , 297–310.

Helser, T. E., Punt, A. E., & Methot, R. D. (2004). A generalized linear mixed model analysis of a multi-vessel fishery resource survey. *Fisheries Research*, **70**(2), 251–264.

Hilborn, R. & Walters, C. J. (2013). *Quantitative fisheries stock assessment: choice, dynamics and uncertainty*. Springer Science & Business Media.

Hinton, M. G. & Maunder, M. N. (2004). Methods for standardizing cpue and how to select among them. *Col. Vol. Sci. Pap. ICCAT*, **56**(1), 169–177.

Hodges, J. S., Carlin, B. P., & Fan, Q. (2003). On the precision of the conditionally autoregressive prior in spatial models. *Biometrics*, **59**(2), 317–322.

Hoffman, M. D. & Gelman, A. (2014). The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *The Journal of Machine Learning Research*, **15**(1), 1593–1623.

Hrafnkelsson, B. & Cressie, N. (2003). Hierarchical modeling of count data with application to nuclear fall-out. *Environmental and Ecological Statistics*, **10**(2), 179–200.

Huxley, T. H. (1883). *Inaugural meeting of the fishery congress*. W. Clowes and sons, limited.

Jin, X., Carlin, B. P., & Banerjee, S. (2005). Generalized hierarchical multivariate car models for areal data. *Biometrics*, **61**(4), 950–961.

Katsanevakis, S. & Maravelias, C. D. (2009). Bathymetric distribution of demersal fish in the aegean and ionian seas based on generalized additive modeling. *Fisheries Science*, **75**(1), 13.

Knorr-Held, L. (1999). Bayesian modelling of inseparable space-time variation in disease risk. .

Lee, D. (2011). A comparison of conditional autoregressive models used in bayesian disease mapping. *Spatial and Spatio-temporal Epidemiology*, **2**(2), 79–89.

Lee, Y. & Nelder, J. A. (2001). Hierarchical generalised linear models: a synthesis of generalised linear models, random-effect models and structured dispersions. *Biometrika*, **88**(4), 987–1006.

Lindgren, F., Rue, H., & Lindström, J. (2011). An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **73**(4), 423–498.

Lindström, J. & Lindgren, F. (2008). A gaussian markov random field model for total yearly precipitation over the african sahel. *Preprints in Mathematical Sciences*, **2008**(8).

Link, W. A. & Eaton, M. J. (2012). On thinning of chains in mcmc. *Methods in Ecology and Evolution*, **3**(1), 112–115.

Lynch, P. D., Shertzer, K. W., & Latour, R. J. (2012). Performance of methods used to estimate indices of abundance for highly migratory species. *Fisheries Research*, **125**, 27–39.

MacNab, Y. C. (2010). On gaussian markov random fields and bayesian disease mapping. *Statistical Methods in Medical Research*, .

Magurran, A. E. (2007). Species abundance distributions over time. *Ecology letters*, **10**(5), 347–354.

Maunder, M. & Starr, P. (2003). Fitting fisheries models to standardised cpue abundance indices. *Fisheries Research*, **63**(1), 43–50.

Maunder, M. N. & Punt, A. E. (2004). Standardizing catch and effort data: a review of recent approaches. *Fisheries Research*, **70**(2), 141–159.

Maunder, M. N., Sibert, J. R., Fonteneau, A., Hampton, J., Kleiber, P., & Harley, S. J. (2006). Interpreting catch per unit effort data to assess the status of individual stocks and communities. *ICES Journal of Marine Science: Journal du Conseil*, **63**(8), 1373–1385.

McCullagh, P. & Nelder, J. A. (1989). *Generalized linear models*, volume 37. CRC press.

Mikkonen, S., Rahikainen, M., Virtanen, J., Lehtonen, R., Kuikka, S., & Ahvonen, A. (2008). A linear mixed model with temporal covariance structures in modelling catch per unit effort of baltic herring. *ICES Journal of Marine Science: Journal du Conseil*, **65**(9), 1645–1654.

Moran, P. A. (1950). Notes on continuous stochastic phenomena. *Biometrika*, **37**(1/2), 17–23.

Morgan, D., Sanz, C., Onononga, J. R., & Strindberg, S. (2006). Ape abundance and habitat use in the goualougo triangle, republic of congo. *International Journal of Primatology*, **27**(1), 147–179.

MPI (2016). Fisheries assessment plenary may 2016: Stock assessments and stock status. .

Murray, L. G., Hinz, H., Hold, N., & Kaiser, M. J. (2013). The effectiveness of using cpue data derived from vessel monitoring systems and fisheries logbooks to estimate scallop biomass. *ICES Journal of Marine Science: Journal du Conseil*, **70**(7), 1330–1340.

Neal, R. M. et al. (2011). Mcmc using hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, **2**, 113–162.

Nelder, J. A. & Baker, R. J. (1972). Generalized linear models. *Encyclopedia of statistical sciences*, .

OSB, O. S. B. (1998). Improving fish stock assessments. .

Paciorek, C. J. et al. (2013). Spatial models for point and areal data using markov random fields on a fine grid. *Electronic Journal of Statistics*, **7**, 946–972.

Paradis, E. (2016). Morans autocorrelation coefficient in comparative methods. .

Pereira, J. C., Leandro, R. A., Petrere, M., & Nishida, T. (2012). Comparison between univariate and bivariate geostatistical models for estimating catch per unit of effort (cpue): A simulation study. *Fisheries Research*, **121**, 115–125.

Preston, F. W. (1960). Time and space and the variation of species. *Ecology*, **41**(4), 611–627.

Raghavan, R. K., Goodin, D. G., Neises, D., Anderson, G. A., & Ganta, R. R. (2016). Hierarchical bayesian spatio–temporal analysis of climatic and socio–economic determinants of rocky mountain spotted fever. *PloS one*, **11**(3).

Ricker, W. E. (1954). Stock and recruitment. *Journal of the Fisheries Board of Canada*, **11**(5), 559–623.

Rue, H. & Held, L. (2005). *Gaussian Markov random fields: theory and applications*. CRC Press.

Rue, H. & Martino, S. (2007). Approximate bayesian inference for hierarchical gaussian markov random field models. *Journal of statistical planning and inference*, **137**(10), 3177–3192.

Russell, E. (1931). Some theoretical considerations on the overfishing problem. *Journal du conseil*, **6**(1), 3–20.

Schaefer, M. B. (1957). Some considerations of population dynamics and economics in relation to the management of the commercial marine fisheries. *Journal of the Fisheries Board of Canada*, **14**(5), 669–681.

Schwarz, G. et al. (1978). Estimating the dimension of a model. *The annals of statistics*, **6**(2), 461–464.

Shelton, A. O., Thorson, J. T., Ward, E. J., Feist, B. E., & Cooper, A. (2014). Spatial semiparametric models improve estimates of species abundance and distribution. *Canadian Journal of Fisheries and Aquatic Sciences*, **71**(11), 1655–1666.

Shen, C., Li, C., & Si, Y. (2016). Spatio-temporal autocorrelation measures for nonstationary series: A new temporally detrended spatio-temporal moran's index. *Physics Letters A*, **380**(1), 106–116.

Sørbye, S. H. & Rue, H. (2014). Scaling intrinsic gaussian markov random field priors in spatial modelling. *Spatial Statistics*, **8**, 39–51.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **64**(4), 583–639.

Steinsland, I. & Jensen, H. (2010). Utilizing gaussian markov random field properties of bayesian animal models. *Biometrics*, **66**(3), 763–771.

Swartzman, G., Huang, C., & Kaluzny, S. (1992). Spatial analysis of bering sea groundfish survey data using generalized additive models. *Canadian Journal of Fisheries and Aquatic Sciences*, **49**(7), 1366–1378.

Tasaki, H. (1998). From quantum dynamics to the canonical distribution: general picture and a rigorous example. *Physical review letters*, **80**(7), 1373.

Team, R. (2015). Rstudio: integrated development for r. *RStudio, Inc., Boston, MA URL http://www. rstudio. com*, .

Thorson, J. T., Shelton, A. O., Ward, E. J., & Skaug, H. J. (2015). Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for west coast groundfishes. *ICES Journal of Marine Science: Journal du Conseil*, .

Thorson, J. T. & Ward, E. J. (2013). Accounting for space–time interactions in index standardization models. *Fisheries Research*, **147**, 426–433.

Thorson, J. T. & Ward, E. J. (2014). Accounting for vessel effects when standardizing catch rates from co–operative surveys. *Fisheries Research*, **155**, 168–176.

Tian, S., Chen, X., Chen, Y., Xu, L., & Dai, X. (2009). Standardizing cpue of ommastrephes bartramii for chinese squid-jigging fishery in northwest pacific ocean. *Chinese Journal of Oceanology and Limnology*, **27**, 729–739.

Vehtari, A. & Gelman, A. (2014). Waic and cross-validation in stan. *Unpublished*, .

Vehtari, A., Gelman, A., & Gabry, J. (2016). Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Unpublished*, .

Venables, W. N. & Dichmont, C. M. (2004). Glms, gams and glmms: an overview of theory for applications in fisheries research. *Fisheries research*, **70**(2), 319–337.

Wang, Y. & So, M. K. (2016). A bayesian hierarchical model for spatial extremes with multiple durations. *Computational Statistics & Data Analysis*, **95**, 39–56.

Watanabe, S. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, **11**(Dec), 3571–3594.

West, W. M., Winker, H., & Kerwath, S. E. (2014). Standardization of the catch per unit effort for albacore (thunnus alalunga) for the south african tuna-pole (baitboat) fleet for the time series 1999–2011. *Collect. Vol. Sci. Pap. ICCAT*, **70**(3), 1247–1255.

White, G. & Ghosh, S. K. (2009). A stochastic neighborhood conditional autoregressive model for spatial data. *Computational statistics & data analysis*, **53**(8), 3033–3046.

Wikle, C. K. & Hooten, M. B. (2006). Hierarchical bayesian spatio-temporal models for population spread. *Applications of computational statistics in the environmental sciences: hierarchical Bayes and MCMC methods*, , 145–169.

Yuan, Y. & Johnson, V. E. (2012). Goodness-of-fit diagnostics for bayesian hierarchical models. *Biometrics*, **68**(1), 156–164.

Yue, Y. R. & Wang, X.-F. (2014). Spatial gaussian markov random fields: Modelling, applications and efficient computations. *Journal of Biometrics & Biostatistics*, **2014**.

Yule, G. U. (1921). On the time-correlation problem, with especial reference to the variate-difference correlation method. *Journal of the Royal Statistical Society*, **84**(4), 497–537.