Spatial-Longitudinal Bent-Cable Model with an Application to Atmospheric CFC Data

A Thesis Submitted to the

College of Graduate Studies and Research

in Partial Fulfillment of the Requirements

for the degree of Master of Science

in the Department of Mathematics and Statistics

University of Saskatchewan

Saskatoon

By

Md. Masud Rana

Permission to Use

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Mathematics and Statistics

Room 142 McLean Hall

106 Wiggins Road

University of Saskatchewan

Saskatoon, Saskatchewan

Canada

S7N 5E6

ABSTRACT

Spatial data (also called georeferenced data) arise in a wide range of scientific studies, including geography, agriculture, criminology, geology, urban and regional economics. The underlying spatial effects – the measurement error caused by any spatial pattern embedded in data – may affect both the validity and robustness of traditional descriptive and inferential techniques. Therefore, it is of paramount importance to take into account spatial effects when analysing spatially dependent data. In particular, addressing the spatial association among attribute values observed at different locations and the systematic variation of phenomena by locations are the two major aspects of modelling spatial data.

The bent-cable is a parametric regression model to study data that exhibits a trend change over time. It comprises two linear segments to describe the incoming and outgoing phases, joined by a quadratic bend to model the transition period. For spatial longitudinal data, measurements taken over time are nested within spatially dependent locations. In this thesis, we extend the existing longitudinal bent-cable regression model to handle spatial effects. We do so in a hierarchical Bayesian framework by allowing the error terms to be correlated across space. We illustrate our methodology with an application to atmospheric chlorofluorocarbon (CFC) data. We also present a simulation study to demonstrate the performance of our proposed methodology.

Although we have tailored our work for the CFC data, our modelling framework may be applicable to a wide variety of other situations across the range of the econometrics, transportation, social, health and medical sciences. In addition, our methodology can be further extended by taking into account interaction between temporal and spatial effects. With the current model, this could be done with a spatial correlation structure that changes as a function of time.

ACKNOWLEDGEMENTS

First and foremost, I wish to express my utmost gratitude and appreciation to my supervisors Shahedul A. Khan and Longhai Li for their excellent guidance, helpful supervision, constant support and caring patience. They have been my inspiration as I hurdle all the obstacles in the completion of this study. Research life became smooth and rewarding for me for their accessibility and willingness to help all the time.

I want to extend my appreciation to the NOAA/ESRL and GAGE/AGAGE global network program for making their CFC-11 and CFC-12 data publicly available. I am grateful to Geoffry Dutton, NOAA Earth System Research Laboratory, Global Monitoring Division for permitting me to use their data and to Derek Cunnold, a member of the AGAGE team, for clarification of their data used in this thesis.

I am also grateful to the Department of Mathematics and Statistics, University of Saskatchewan, for the computer facilities, and financial support in terms of graduate assistantships. The research assistantships from Longhai Li are highly appreciated.

I would like to thank all my friends for their encouragement and support.

I express my gratitude to my entire family. My parents have been a constant source of support – emotional, moral and of course financial – throughout my life, and this thesis would certainly not existed without them. My sisters and brother were always supportive me and encouraging me with their best wishes.

Finally, I would like to thank everybody who was important to the successful completion of this thesis, as well as expressing my apology that I could not mention personally one by one.

DEDICATION

I would like to dedicate this dissertation to my parents, who have offered me unconditional love and support all the way since the beginning of my studies. There is no doubt in my mind that without their continued encouragement and counsel I could not have completed this process.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CFC	Chlorofluorocarbon
ODS	Ozone Depleting Substances
CTP	Critical Time Point
DIC	Deviance Information Criteria
MCMC	Markov Chain Monte Carlo
NOAA	National Oceanic and Atmospheric Administration
ESRL	Earth System Research Laboratory
GAGE	Global Atmospheric Gases Experiment
AGAGE	Advanced Global Atmospheric Gases Experiment

Chapter 1

Introduction

Spatial or georeferenced data are tagged by locations, where the locations can be defined using geographical information such as well defined borders, latitude and longitude, and so on. This type of data can arise in a wide range of scientific areas, including agriculture, criminology, demography, geology, geography, international relations, natural resources, regional science, sociology, statistics, urban planning and urban and regional economics. Spatial data are also common in longitudinal studies, where measurements taken over time are nested within observational units drawn from some population of interests. For convenience purpose, in this thesis we will use the term "individual" to refer to the observational unit in the longitudinal study, including inanimate objects such as a measurement station.

Analysis of spatial data requires special attention, partly because of the underlying spatial configuration, featuring (1) spatial pattern of locations, (2) spatial association between attribute values observed at different locations (i.e. spatial dependency), and (3) systematic variation of phenomena by locations (i.e. spatial heterogeneity). Moreover, spatial effects – the measurement error caused by any spatial pattern embedded in data – affect both the validity and robustness of traditional descriptive and inferential techniques (Griffith & Layne, 1999). Therefore, it is of paramount importance to take into account spatial effects when analysing spatially dependent data.

An example of spatial data is the atmospheric concentration of chlorofluorocarbons (CFCs) monitored from different stations across the globe (see Chapter 3 for detailed description of the data). Individual CFC molecules are labelled with a unique numbering system. For example, the CFC number of 11 indicates the number of atoms of carbon, hydrogen, fluorine, and chlorine (e.g. CCl_3F as CFC-11). The most common CFCs are CFC-11, CFC-12, CFC-113, CFC-114, and CFC-115. The ozone depletion potential (the ratio of the impact
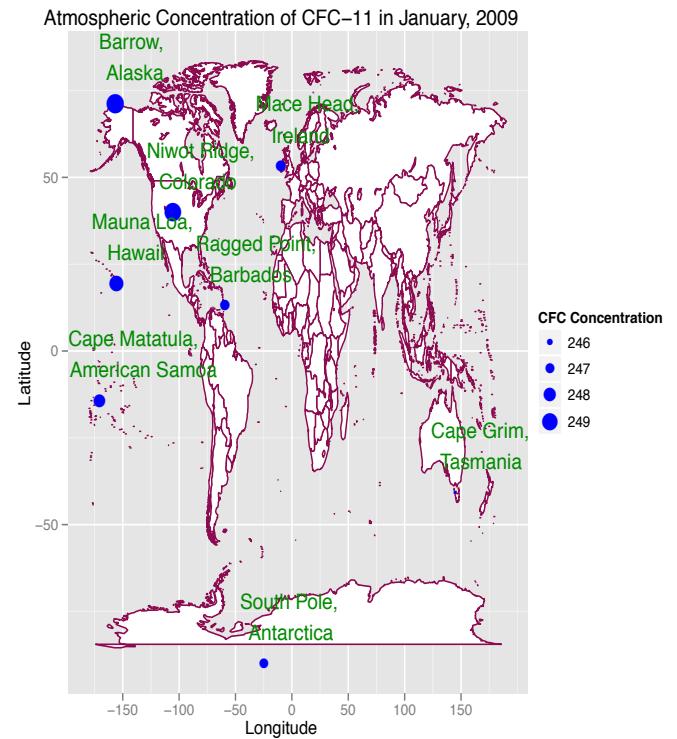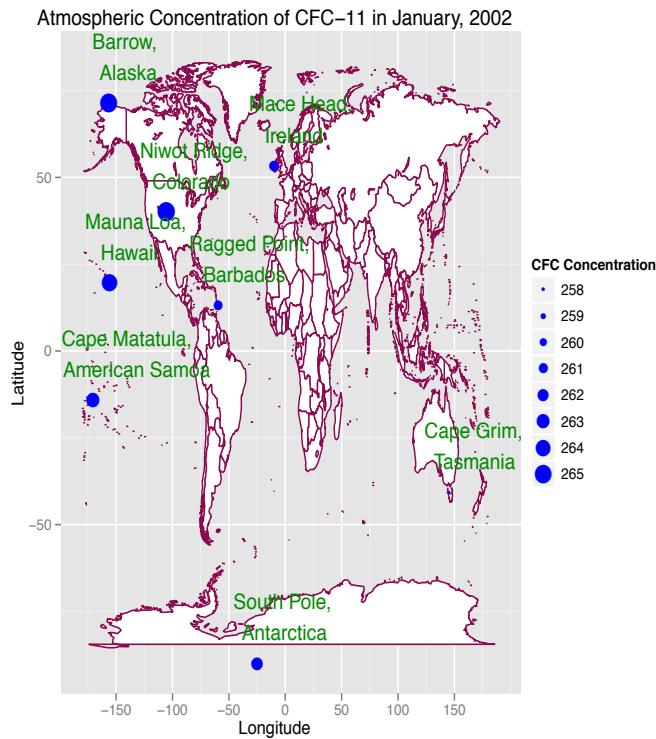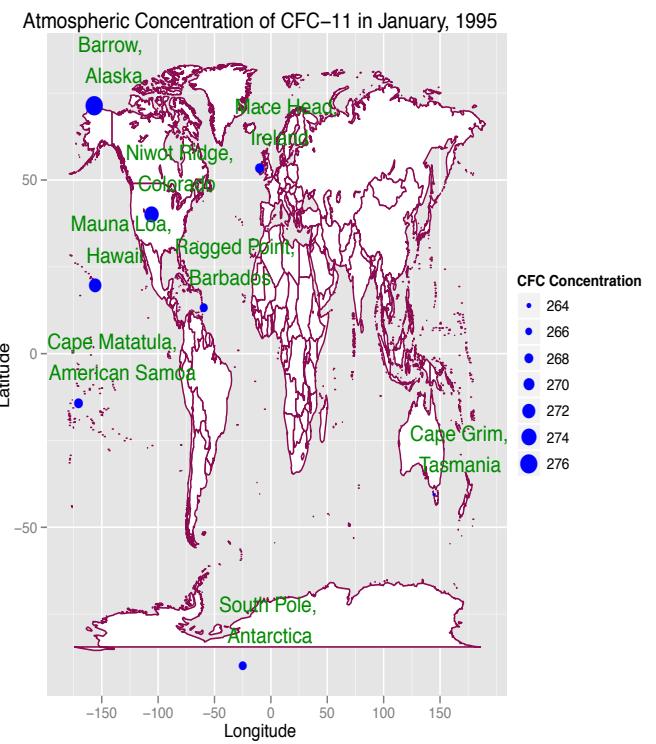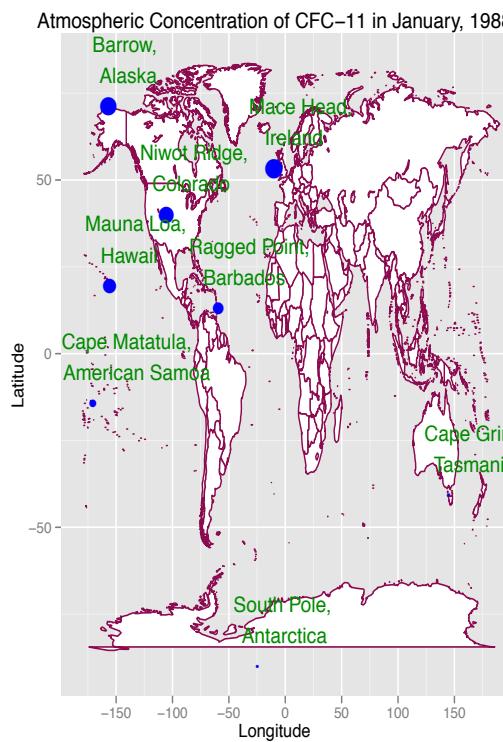
Figure 1.1: Monthly mean concentrations of CFC-11 in parts-per-trillion (ppt) for eight stations for January of 1988, 1995, 2002, and 2009.
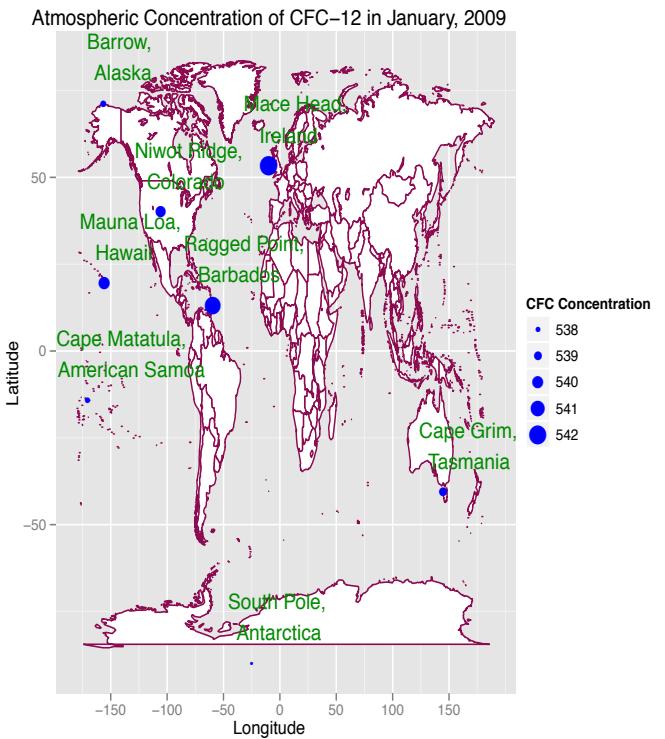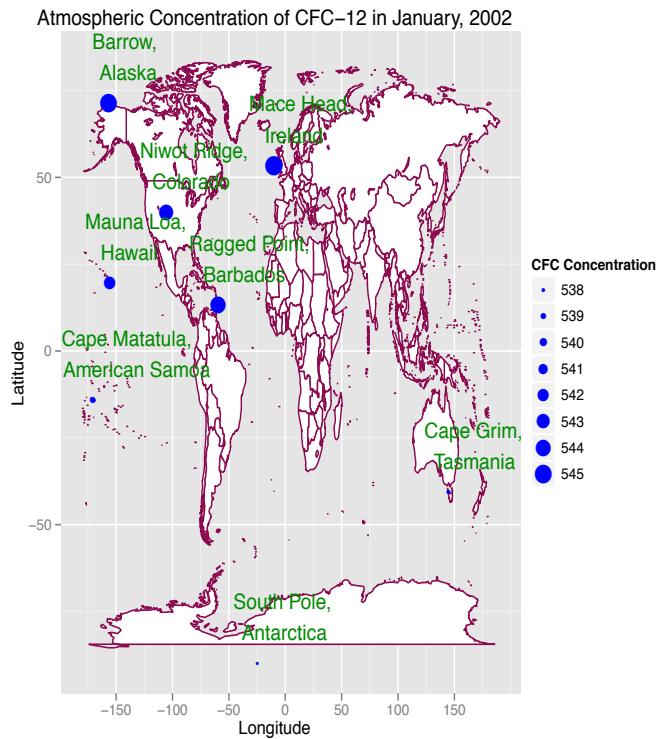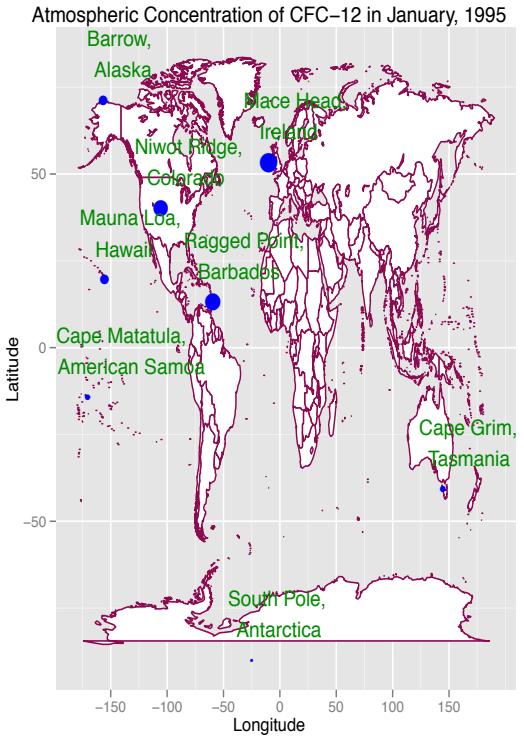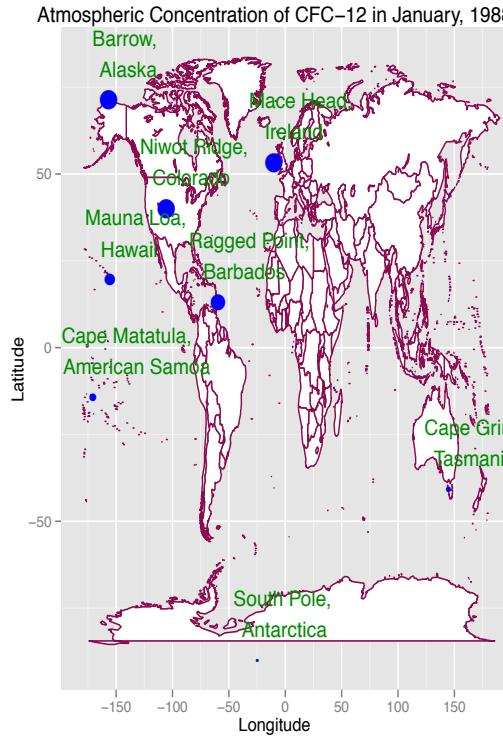
Figure 1.2: Monthly mean concentrations of CFC-12 in parts-per-trillion (ppt) for eight stations for January of 1988, 1995, 2002, and 2009.

on ozone of a chemical compared to the impact of a similar mass of CFC-11) for these CFCs are 1, 1, 0.8, 1, and 0.6, respectively (`http://www.epa.gov/ozone/defn.html`). We focus on CFC-11 and CFC-12 in this thesis, because they are considered two of the most dangerous CFCs to the atmosphere. Due to their extended lifetimes, CFCs persist long enough in the atmosphere, and consequently are believed to have spread across the world. Therefore, CFCs monitored from one station may depend on those from another station, giving rise to a presumed spatially dependent longitudinal process. To illustrate the spatial dependency, we plot the monthly mean concentrations of CFC-11 and CFC-12 in Figures 1.1 and 1.2, respectively, for eight monitoring stations for January of 1988, 1995, 2002, and 2009 with reference to the latitude and longitude of each station. The horizontal and vertical axes represent longitude and latitude of stations. The blue circle represents the relative concentration of CFCs, with a bigger/smaller circle indicates higher/lower concentration level. In general, the CFC measurements appeared to be closer for the nearer stations. For example, the level of CFC-11 at Barrow (252.765 ppt) is closer to that of Mauna Loa (246.664 ppt) than Cape Matatula (235 ppt). Such similarities roughly indicate spatial dependencies among the CFC measurements around the world.

We are particularly interested about CFCs for their notorious effect on depleting ozone layer. Stratospheric ozone, which accounts for around 90% of total atmospheric ozone, is vital for human health and environments on Earth, as it prevents harmful ultraviolet (UV) radiation from Earth's surface. Note that increased UV radiation poses significant risk on living organisms, including skin cancer, cataracts, irreversible damage to plants and decreases drifting organisms (e.g. animals, plants, archaea, bacteria) in the ocean's photic-zone (Struijs et al., 2010; Khan et al., 2009). CFCs constitute about 90% of the Ozone Depleting Substances (ODS) and 80% of which is accounted by CFC-11 and CFC-12, most abundant CFCs in the atmosphere (Moulijn et al., 2000). These CFCs, being sources of chlorine in the atmosphere, catalytically destroy stratospheric ozone (Malina & Rowland, 1974). Each chlorine atom has potential of breaking down an average of 100,000 ozone molecules during its one to two years of atmospheric lifetime. Moreover, they are not only strong infrared absorbers but also very potent greenhouse gas (referred in Zhang et al. (2010)). Recognizing the significant threat of the CFCs to the ozone layer, the *Montreal Protocol on Substances That Deplete Ozone Layer*

(herein Montreal protocol) provided a mechanism to reduce and phase-out global production, use, and consumption of CFCs with other ODSs.

For successful regulations of the Montreal protocol, the emission and production of CFCs along with other ODSs have been decreased globally (Velders et al., 2007; WMO, 2010). As a result, ozone science has entered into the second stage of recovery, where ozone is expected to increase for continuing decrease in ODSs (WMO, 2010). This recovery had been attributed mainly to the reductions in CFC-11 and CFC-12 productions (Velders et al., 2007). However, the quantity of CFCs in existing products and equipments, usually referred as 'banks', has the potential to make an important contribution to future emissions (Zhang et al., 2010). Hence, the analysis of CFC data is necessary not only to assess the global and regional emissions of CFC and to check their compliance with phase-out schedules (Zhang et al., 2010), but also to identify the progress in ozone recovery from ODSs and evaluate the effectiveness of climate and ozone protection policy options (WMO, 2010).
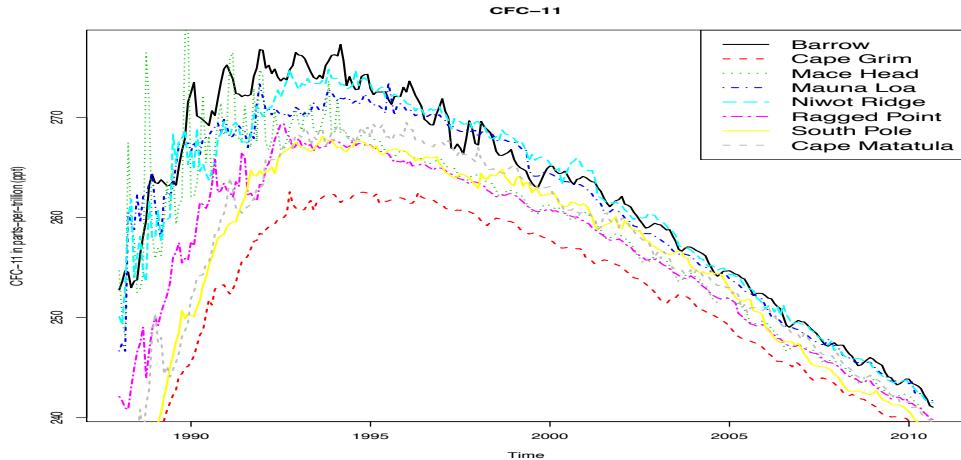
Figure 1.3: Monthly mean profile of 8 stations for CFC-11 over the period 1988 to 2010.

Another important feature of CFCs, as explored by Khan et al. (2009), is that their trends characterize a change due to the response of the Montreal protocol. Roughly, the trend can be characterized by three phases: linear incoming and outgoing, which are joined by a quadratic bend to characterize the transition (see Figure 1.3 for CFC-11 and Figure 1.4 for CFC-12 data). This type of characterization can be achieved using the single profile bent-cable model of Chiu et al. (2006). Later, Khan et al. (2009) extended the bent-cable
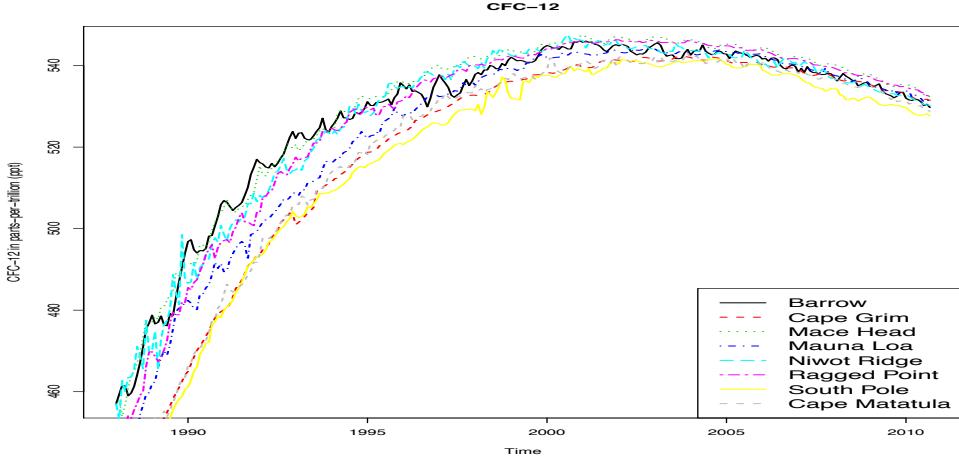
Figure 1.4: Monthly mean profile of 8 stations for CFC-12 over the period 1988 to 2010.

methodology for longitudinal data (i.e. for multiple profiles). Accounting for the spatial feature, the objective of this thesis is to extend the longitudinal bent-cable model of Khan et al. (2009) to accommodate spatial dependency embedded in data. Using our proposed methodology, we will address the following questions statistically:

Q1 What was the time point at which the CFC concentration took a downward turn from an increasing trend?

Q2 Has the concentration of CFC been decreasing significantly after the transition?

Q3 Are there any differences and/or similarities in the atmospheric concentration of CFCs in different parts of the world?

Q4 What is the spatial distribution of the CFC concentration around the globe?

The bent-cable model (Figure 1.5) comprises of two linear segments to describe the incoming and outgoing phases, joined by a quadratic bend to model the transition period. The model is appealing mainly for two reasons. First, it is parsimonious in the sense that it captures the change of direction of a single profile using only five regression coefficients: three linear parameters to describe the rates of changes in the incoming and outgoing phases, and two transition parameters to characterize the centre and the half-width of the bend (see Figure 1.5). Note that the critical time point (CTP), as defined by Chiu and Lockhart (2010),

6
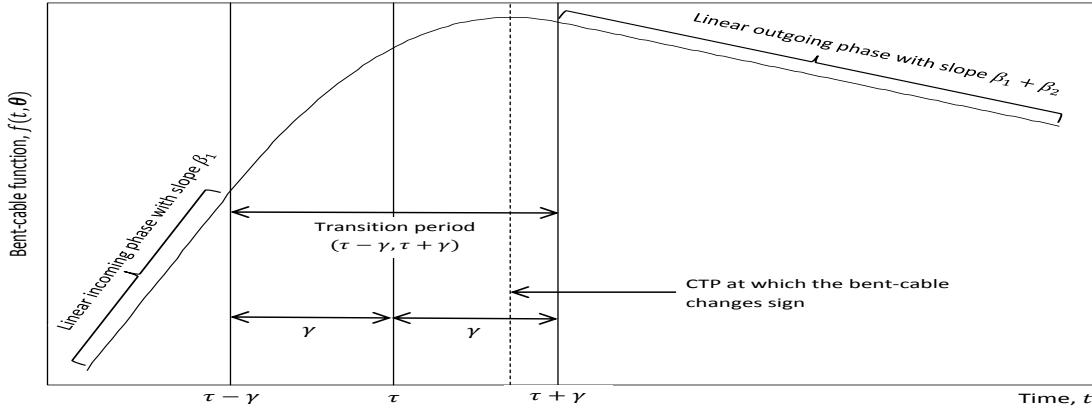
Figure 1.5: The schematic view of bent-cable regression model, modified from Khan et al. (2009).

is the time at which the slope of the bent-cable changes signs, and can be defined using the bent-cable parameters. Secondly, the parameters can be interpreted in the same way as usually done in regression analysis. Khan et al. (2009) extended the bent-cable model for longitudinal data, and developed a Bayesian modelling framework for statistical inference.

Our modelling approach will provide a straightforward methodology to model spatial dependency embedded in data. We plot the observed curve and corresponding fitted curve for our spatial-longitudinal bent-cable model for a representative monitoring station, Cape Matatula, American Samoa of CFC-11 and CFC-12 in Figure 1.6 and 1.7, respectively, to demonstrate the adequacy of our modelling methodology.
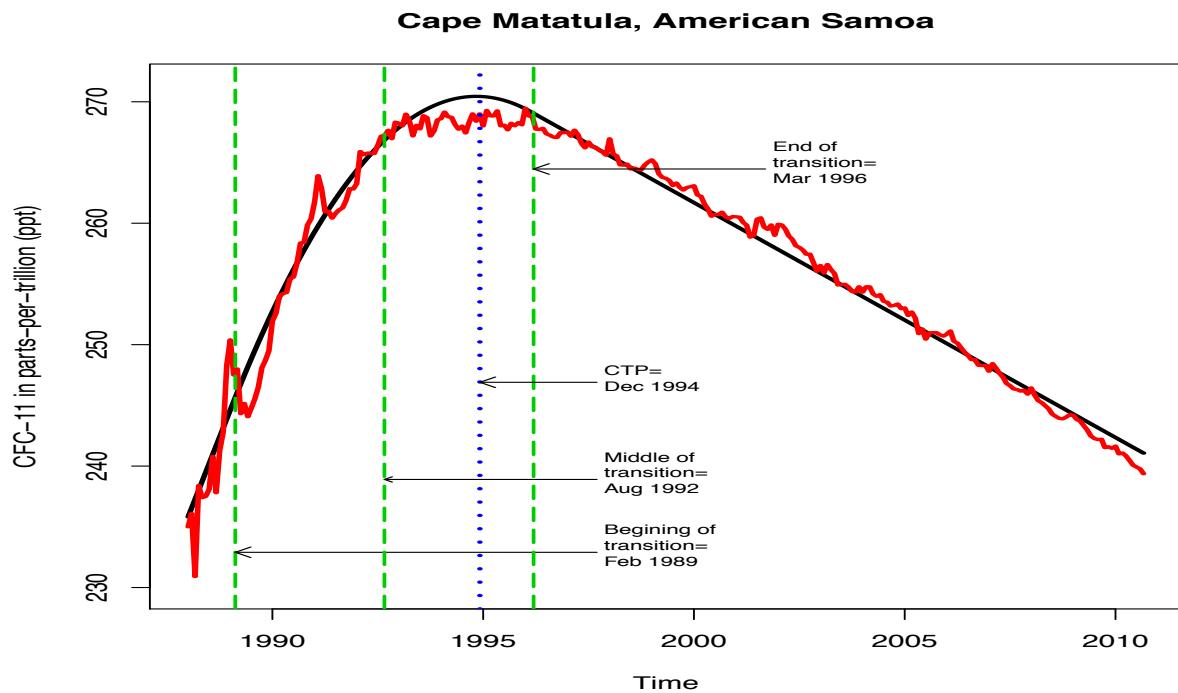
Figure 1.6: Observed data (Red line) and corresponding station-specific fitted curve of CFC-11 for Cape Matatula, American Samoa,
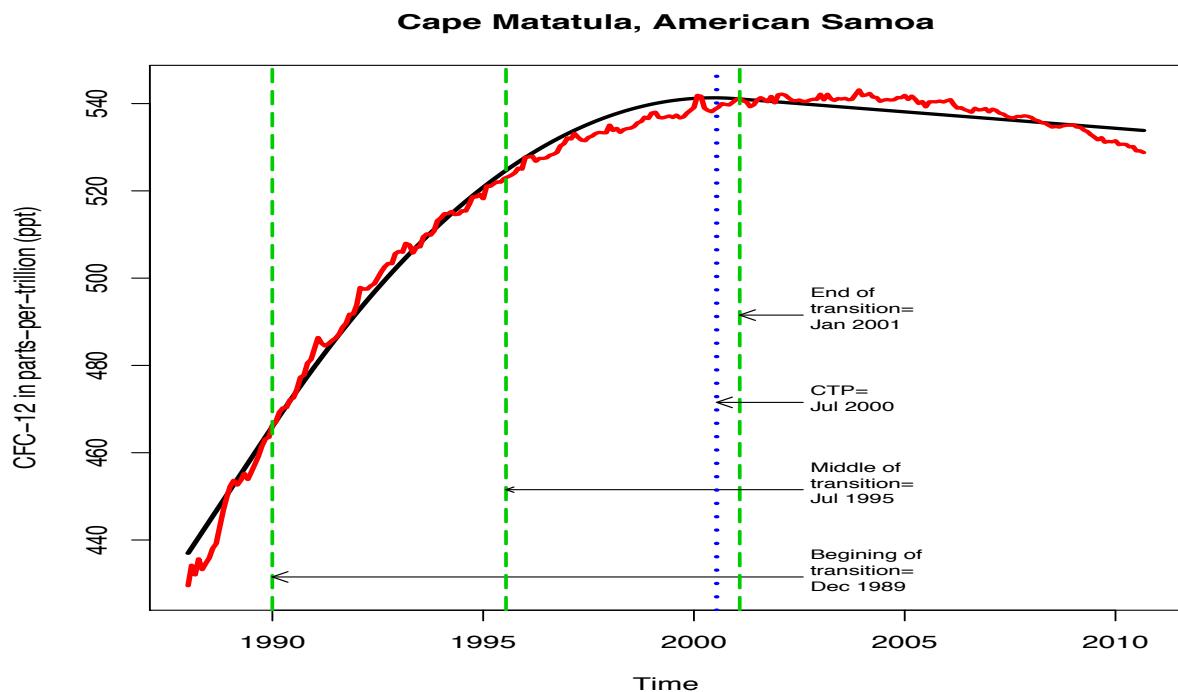
Figure 1.7: Observed data (Red line) and corresponding station-specific fitted curve of CFC-12 for Cape Matatula, American Samoa

8

Chapter 2

The Spatial-Longitudinal Bent-Cable Model

In this chapter, we present the details of our spatial-longitudinal bent-cable model. In Section 2.1, we describe the formulation of the model as an extension of longitudinal bent-cable model of Khan et al. (2009). We describe the inference method of our spatial-longitudinal bent-cable model in Section 2.2.

2.1 Hierarchical Formulation of the Model

Our spatial-longitudinal bent-cable model is based on Bayesian paradigm in the form of hierarchical model for making inference about the parameters of interest. Bayesian analysis provides a cohesive framework for combining complex data structure and external knowledge through modelling the observed data and unknown parameters as random variables. It offers sound foundation, unified approach to data analysis, and ability to formally incorporate prior information or external empirical evidence into the results by means of prior distribution of parameters (Banerjee et al., 2003). Uncertainty of a process is modelled by introducing latent variables at different stages, where the stages might describe conceptual and unobserved latent processes. Thus, Bayesian hierarchical modelling offers a flexible approach for complex data (or response with complex forms) through defining latent, intermediate between the observed data and the underlying parameters driving the process (Congdon, 2010), and incorporating additional knowledge via prior specification for the parameters.

Wikle (2003) defined a hierarchical model in which the dependent variable characterizes the data stage, latent effects quantify the process stage, and a set of hyper-parameters defines the parameter stage to quantify prior information about the random quantities. In this three stage hierarchical model, the first stage is defined as a conditional likelihood, given both

the process and parameters; the second stage describes the latent process conditional on higher stage parameters; and the third stage models the population parameters of interest (Congdon, 2010). For longitudinal data, the latent process is characterized by modelling each individual's trajectory over time. This stage addresses the questions: (1) how does the response change over time individually, and (2) do different individuals experience different pattern of change? In the next few sections, we describe the hierarchical formulation of our spatial-longitudinal bent-cable model through the three stages described above.

2.1.1 The Spatial-Longitudinal Bent-Cable Model

We index the individuals by $i = 1, 2, \ldots, m$ and the times by $j = 1, 2, \ldots, n$. We model the response for the ith individual at time t_{ij}, denoted by y_{ij}, by the bent-cable model

$$y_{ij} = f(t_{ij}, \boldsymbol{\theta}_i) + e_{ij} \tag{2.1}$$

where the bent-cable function $f(t_{ij}, \boldsymbol{\theta}_i)$ is defined by

$$f(t_{ij}, \boldsymbol{\theta}_i) = \beta_{0i} + \beta_{1i}t_{ij} + \beta_{2i}q(t_{ij}, \boldsymbol{\alpha}_i) \tag{2.2}$$

where

$$q(t_{ij}, \boldsymbol{\alpha}_i) = \frac{(t_{ij} - \tau_i + \gamma_i)^2}{4\gamma_i} 1\{|t_{ij} - \tau_i| \leq \gamma_i\} + (t_{ij} - \tau_i)1\{t_{ij} - \tau_i > \gamma_i\} \tag{2.3}$$

with $\boldsymbol{\theta}_i = (\boldsymbol{\beta}'_i, \boldsymbol{\alpha}'_i)'$, $\boldsymbol{\beta}_i = (\beta_{0i}, \beta_{1i}, \beta_{2i})'$ and $\boldsymbol{\alpha}_i = (\gamma_i, \tau_i)'$. As illustrated in Figure 1.5 of Chapter 1, β_{0i} and β_{1i} are, respectively, the intercept and slope of the incoming phase; $\beta_{1i} + \beta_{2i}$ is the slope of the outgoing phase; and γ_i and τ_i, the half-width and centre of the bend, respectively, characterize the transition period. Note that the beginning and end of a transition can be represented by $\tau_i - \gamma_i$ and $\tau_i + \gamma_i$, respectively. Chiu and Lockhart (2010) also defined the Critical Time Point (CTP), at which the bent cable changes direction (i.e., takes a downturn from an increasing trend, and vice versa) by $\tau_i - \gamma_i - 2\beta_{1i}\gamma_i/\beta_{2i}$. Henceforth, we will denote $f(t_{ij}, \boldsymbol{\theta}_i)$ and $q(t_{ij}, \boldsymbol{\alpha}_i)$ by f_{ij} and q_{ij}, respectively, for notational simplicity.

We take into account the spatial configuration through the error term, e_{ij}, by decomposing it into two components: one describing the spatial configuration and other one for white noise. That is, we model the error component as follows:

$$e_{ij} = \delta \sum_{\substack{k=1 \\ k \neq i}}^{m} w_{ik} e_{kj} + \epsilon_{ij} \tag{2.4}$$

where the parameter, δ is the spatial autocorrelation coefficient that measures the strength of spatial association, $\mathbb{W} = (w_{ik})_{m \times m}$ is the spatial weight matrix that describes the spatial arrangement of the locations (stations), and ϵ_{ij} is the white noise. Here, w_{ik} is assumed known representing the a priori assumption about the spatial relationship between stations i and k, with all the diagonal elements w_{ii} conventionally set to zeros. The spatial weight matrix $\mathbb{W}$ is typically normalized in an attempt to constrain δ in the interval $(-1, +1)$ (Elhorst, 2010). One such normalization technique is to divide the elements of spatial weight matrix, $\mathbb{W}$ by its largest characteristic root. Note that if the off-diagonal elements of $\mathbb{W}$ are set to zeros, the first term in the right hand side of (2.4)vanishes. This leads to the assumption of no spatial effects, and constitutes the longitudinal bent-cable model of Khan et al. (2009).

Equations (2.1)-(2.4) constitute our spatial-longitudinal bent-cable model. Next, we describe the stages of the Bayesian hierarchical formulation of our model.

2.1.2   The Hierarchy

The pragmatic view of Bayesian hierarchical modelling is to formulate the model in terms of three entities – data, process and parameters – all of which may have stochastic elements.

Stage 1: Data Modelling

The first stage (Stage 1 or within individual variation) of our hierarchical model deals with the observational process, which describes the distribution of data given the process of interest and parameters of the data model. This level characterizes an individual trajectory by taking into account correlation and variation among the repeated measurements over time. We accomplished this objective through (2.4), and assuming normality for ϵ_{ij}:

$$\left[\epsilon_{ij} | \sigma_i^2 \right] \sim \mathcal{N}(0, \sigma_i^2) \text{ for all } i, j \tag{2.5}$$

where σ_i^2 is the innovation variance.

Let $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{in})$, $\mathbf{y} = (\mathbf{y}_1', \mathbf{y}_2', \ldots, \mathbf{y}_m')'$, and $\mathbf{\Theta}$ be the vector of all model parameters collectively. Then, combining equation (2.1), (2.4) and (2.5), the first level of hierarchy

for our spatial-longitudinal bent-cable model can be expressed as:

$$\pi\left[\mathbf{y}|\boldsymbol{\Theta}\right] = (2\pi)^{-\frac{mn}{2}} \mid \mathbb{I}_m - \delta\mathbb{W} \mid^n \left\{ \prod_{i=1}^{m} \left(\sigma_i^{-2}\right)^{\frac{n}{2}} \right\}$$

$$\exp\left\{ -\frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} \left[\left(y_{ij} - \delta \sum_{k\neq i}^{m} w_{ik}y_{kj} \right) - \left(f_{ij} - \delta \sum_{k\neq i}^{m} w_{ik}f_{kj} \right) \right]^2 \right\} \quad (2.6)$$

where $\mid \mathbb{I}_m - \delta\mathbb{W} \mid^n$ in (2.6) represents the Jacobian term taking into account the endogeneity of $\sum_{k=1}^{m} w_{ik}y_{kj}$ (Anselin, 1988).

Stage 2: Process Modelling

The second stage models the underlying process given the parameters. At this stage, our objective is to capture the variation observed between different individuals, known as between individual variation in longitudinal studies. We achieve this goal by specifying models for individual specific regression coefficients $\boldsymbol{\theta}_i's$. We assume multivariate normal distribution for the linear parameters $\boldsymbol{\beta}_i's$:

$$\left[\boldsymbol{\beta}_i|\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta\right] \sim \mathcal{N}_3(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta) \quad (2.7)$$

where $\mathcal{N}_3$ represents a trivariate normal distribution. Here, $\boldsymbol{\mu}_\beta = (\mu_0, \mu_1, \mu_2)'$ and Σ_β are, respectively, the mean and covariance of $\boldsymbol{\beta}_i$, which characterize the global (population) trends in the incoming and outgoing phases: μ_0 and μ_1 are, respectively, the population intercept and slope for the incoming phase, $\mu_1 + \mu_2$ is the population slope for the outgoing phase, and Σ_β provides information about the variability of the linear parameters $\boldsymbol{\beta}_i$. Note that the assumption of multivariate normal is appealing due to two reasons. First, it can represent the dependence structure completely via its covariance matrix. Secondly, multivariate normal distribution is preferred from theoretical and computational point of view.

Since both the transition parameter γ_i and τ_i are non-negative, we assume a bivariate lognormal distribution for the transition parameter vector $\boldsymbol{\alpha}_i = (\gamma_i, \tau_i)'$:

$$[\boldsymbol{\alpha}_i|\boldsymbol{\mu}_\alpha, \boldsymbol{\Sigma}_\alpha] \sim \mathcal{LN}_2\left(\boldsymbol{\mu}_\alpha, \boldsymbol{\Sigma}_\alpha\right) \quad (2.8)$$

where $\mathcal{LN}_2$ stands for bivariate lognormal distribution, and $\boldsymbol{\mu}_\alpha = (\mu_\gamma, \mu_\tau)'$ characterizes the global transition. Since our assumption for $\boldsymbol{\alpha}_i$ involves a lognormal distribution, the medians $\exp\{\mu_\gamma\}$ and $\exp\{\mu_\tau\}$ can be regarded as the half-width and the centre of the bend, respectively, for the population. We can also use the standard deviations of γ_i and τ_i to describe the between-station variability for the transition parameters.

Stage 3: Parameter Modelling

The Bayesian approach to statistical analysis provides a cohesive framework for combining complex data models and external knowledge or expert opinion as prior information. Prior information, supplied in the form of prior distribution, quantifies uncertainty about a parameter or a vector of parameters. We summarize our prior knowledge about the parameters through assigning a particular distributional form:

$$\left.\begin{array}{c} \left[\boldsymbol{\mu}_\beta | \mathbf{h}_\beta, \mathbb{H}_\beta\right] \sim \mathcal{N}_3(\mathbf{h}_\beta, \mathbb{H}_\beta), \quad \left[\boldsymbol{\mu}_\alpha | \mathbf{h}_\alpha, \mathbb{H}_\alpha\right] \sim \mathcal{N}_2(\mathbf{h}_\alpha, \mathbb{H}_\alpha) \\[2mm] \left[\Sigma_\beta^{-1} | \nu_\beta, \mathbb{A}_\beta\right] \sim \mathcal{W}(\nu_\beta, (\nu_\beta \mathbb{A}_\beta)^{-1}), \quad \left[\Sigma_\alpha^{-1} | \nu_\alpha, \mathbb{A}_\alpha\right] \sim \mathcal{W}(\nu_\alpha, (\nu_\alpha \mathbb{A}_\alpha)^{-1}) \\[2mm] \left[\sigma_i^{-2} | \mathbf{a}_0, \mathbf{a}_1\right] \sim \mathcal{G}\left(\dfrac{\mathbf{a}_0}{2}, \dfrac{\mathbf{a}_1}{2}\right), \quad \left[\delta | \mathbf{b}_0, \mathbf{b}_1\right] \sim \mathcal{U}(\mathbf{b}_0, \mathbf{b}_1) \end{array}\right\} \qquad (2.9)$$

where $\mathcal{W}$, $\mathcal{G}$, and $\mathcal{U}$ denote Wishart, gamma, and uniform distribution, respectively. The gamma parametrization is in terms of shape and rate parameters. The hyper-parameters $\mathbf{h}_\beta$, $\mathbb{H}_\beta$, $\mathbf{h}_\alpha$, $\mathbb{H}_\alpha$, ν_β, $\mathbb{A}_\beta$, ν_α, $\mathbb{A}_\alpha$, $\mathbf{a}_0$, $\mathbf{a}_1$, $\mathbf{b}_0$, and $\mathbf{b}_1$ are all assumed to be known (see below).

2.1.3   Choice of Prior

Specification of the prior distribution for parameters in applied Bayesian analysis is still central to model identifiability, and robustness in inference (Congdon, 2003). Typically, prior can be chosen based on accumulated information from past studies or from the expert opinion of subject related area. An alternative is to endow such distribution with little information context, so that data from the present study will dominate the posterior. In practice, prior distributions are chosen based on two considerations: conjugacy and informativeness.

Bayesian inference is computationally extensive and solely based on posterior distributions of the relevant parameters. High dimensional integrals may involve in deriving a posterior

distribution, and it may not be possible to evaluate such integrals analytically. Therefore, such integrals are often numerically evaluated to approximate a posterior density. Monte Carlo integration is a common, convenient, and straightforward technique to approximate a posterior density for non-linear hierarchical models. Conjugacy often facilitates to implement the Monte Carlo technique.

Conjugate prior refers to a prior distribution whose selection leads to posterior belonging to the same distributional family as the prior. The prior is conditionally conjugate if the conditional distribution (i.e. the distribution of a parameter given the data and all other remaining parameters) is also from the same family (Gelman, 2006). The Monte Carlo technique works through generating random sample from full conditional distributions. As such, we prefer conditional conjugacy, because it leads to closed-form full conditional from which drawing random sample is computationally easier, and this conjugacy can be preserved when the model is expanded hierarchically (Gelman, 2006).

An informative prior expresses our specific, definite information about a parameter of interest. Often, reliable prior information does not exist or solely data based inference is desired. In such situations, we prefer non-informative prior in the sense that it does not favour one particular value over remaining values and expresses vague or general information about the parameter. Such prior leads to a posterior based on data only.

Our choices of prior for $\boldsymbol{\mu}_\beta$ and $\boldsymbol{\mu}_\alpha$ (see equation (2.9)) lead to conditional conjugacy. Bayesian estimates do not vary much for different choice of the mean vectors ($\mathbf{h}_\beta$ and $\mathbf{h}_\alpha$) as long as the diagonal elements of $\mathbb{H}_\beta$ and $\mathbb{H}_\alpha$ are taken to be large. Usually, large variances lead to a flat prior. Therefore, in practice, the mean vector is taken as zero and the covariance matrix, say $\mathbb{H}_\beta$, is defined in such a way that $\mathbb{H}_\beta^{-1} \approx \mathbb{O}$, where $\mathbb{O}$ is a matrix with all its elements zero (as referred in Khan (2010)). We choose gamma prior for σ_i^{-2} for the sake of conditional conjugacy (Gelman, 2006).

The choice of inverse Wishart distribution as prior for a covariance matrix, or equivalently, Wishart distribution for the inverse of a covariance matrix leads to conjugate prior. So, we choose inverse Wishart distribution as prior for $\boldsymbol{\Sigma}_\beta$ and $\boldsymbol{\Sigma}_\alpha$ and define the parameters of Wishart distribution in terms of its degrees of freedom and expectation following Carlin (1995); Wakefield et al. (1994). For example, $\left[\boldsymbol{\Sigma}_\beta^{-1}|\nu_\beta, \mathbb{A}_\beta\right] \sim \mathcal{W}\left(\nu_\beta, (\nu_\beta \mathbb{A}_\beta)^{-1}\right)$ has degrees

of freedom ν_β and expectation $\mathbb{A}_\beta^{-1}$. We make the Wishart prior nearly flat through setting its degrees of freedom equal to the order of the scale matrix (e.g. 3 for the prior of $\boldsymbol{\Sigma}_\beta^{-1}$) and choose $\mathbb{A}_\beta$ as an approximate prior estimate of $\boldsymbol{\Sigma}_\beta$. Finally, we consider the uniform distribution as prior for δ, as the value of δ is restricted to lie within the range $(0, 1)$.

2.1.4   Summary of the Hierarchy

In summary, we define the three stages of our spatial-longitudinal bent-cable model as follows:

$$
\pi\left[\mathbf{y} | \boldsymbol{\Theta}\right] = (2\pi)^{-\frac{mn}{2}} \mid \mathbb{I}_m - \delta\mathbb{W} \mid^n \left\{ \prod_{i=1}^{m} (\sigma_i^{-2})^{\frac{n}{2}} \right\}
$$

$$
\exp\left\{ -\frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} \left[\left(y_{ij} - \delta \sum_{k \neq i}^{m} w_{ik} y_{kj} \right) - \left(f_{ij} - \delta \sum_{k \neq i}^{m} w_{ik} f_{kj} \right) \right]^2 \right\} \tag{2.10}
$$

$$
\left.\begin{aligned}
\left[\boldsymbol{\beta}_i | \boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta\right] &\sim \mathcal{N}_3(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta) \\
\left[\boldsymbol{\alpha}_i | \boldsymbol{\mu}_\alpha, \boldsymbol{\Sigma}_\alpha\right] &\sim \mathcal{LN}_2\left(\boldsymbol{\mu}_\alpha, \boldsymbol{\Sigma}_\alpha\right)
\end{aligned}\right\} \tag{2.11}
$$

$$
\left.\begin{aligned}
\left[\boldsymbol{\mu}_\beta | \mathbf{h}_\beta, \mathbb{H}_\beta\right] &\sim \mathcal{N}_3(\mathbf{h}_\beta, \mathbb{H}_\beta), \quad \left[\boldsymbol{\mu}_\alpha | \mathbf{h}_\alpha, \mathbb{H}_\alpha\right] \sim \mathcal{N}_2(\mathbf{h}_\alpha, \mathbb{H}_\alpha) \\
\left[\Sigma_\beta^{-1} | \nu_\beta, \mathbb{A}_\beta\right] &\sim \mathcal{W}(\nu_\beta, (\nu_\beta \mathbb{A}_\beta)^{-1}), \quad \left[\Sigma_\alpha^{-1} | \nu_\alpha, \mathbb{A}_\alpha\right] \sim \mathcal{W}(\nu_\alpha, (\nu_\alpha \mathbb{A}_\alpha)^{-1}) \\
\left[\sigma_i^{-2} | \mathbf{a}_0, \mathbf{a}_1\right] &\sim \mathcal{G}\left(\frac{\mathbf{a}_0}{2}, \frac{\mathbf{a}_1}{2}\right), \quad \left[\delta | \mathbf{b}_0, \mathbf{b}_1\right] \sim \mathcal{U}(\mathbf{b}_0, \mathbf{b}_1)
\end{aligned}\right\} \tag{2.12}
$$

Equations (2.10), (2.11) and (2.12) represent the data model (Stage 1), process model (Stage 2) and parameter model (Stage 3), respectively, with hyper-parameters $\mathbf{h}_\beta$, $\mathbb{H}_\beta$, $\mathbf{h}_\alpha$, $\mathbb{H}_\alpha$, ν_β, $\mathbb{A}_\beta$, ν_α, $\mathbb{A}_\alpha$, $\mathbf{a}_0$, $\mathbf{a}_1$, $\mathbf{b}_0$, and $\mathbf{b}_1$, all assumed to be known.

2.2   Bayesian Inference

Statistical inference for our spatial-longitudinal bent-cable model (see Section 2.1.1 and 2.1.2) is carried out via Bayesian technique. As mentioned above, the most difficult part of Bayesian analysis is the evaluation of integrals required to draw inference from posterior density. We preferred Makov Chain Monte Carlo (MCMC) method, because it is straightforward to implement, and provides a unifying framework for approximating integrals. In this section, we describe the method of drawing inference from our spatial-longitudinal bent-cable model.

2.2.1 Posterior Density

Let $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2, \ldots\ldots\ldots, \boldsymbol{\beta}'_m)'$, $\boldsymbol{\alpha} = (\boldsymbol{\alpha}'_1, \boldsymbol{\alpha}'_2, \ldots\ldots\ldots, \boldsymbol{\alpha}'_m)'$ and $\boldsymbol{\sigma}^{-2} = (\sigma_1^{-2}, \sigma_2^{-2}, \ldots\ldots\ldots, \sigma_m^{-2})'$. We denote all the model parameters collectively by $\boldsymbol{\Theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\mu}_\beta, \boldsymbol{\mu}_\alpha, \boldsymbol{\Sigma}_\beta^{-1}, \boldsymbol{\Sigma}_\alpha^{-1}, \boldsymbol{\sigma}^{-2}, \delta)$. The joint distribution of the model parameters and the data (see Section A.1 in Appendix for detailed derivation) can be written as:

$$
\left.
\begin{aligned}
\pi\left(\boldsymbol{\Theta}, \mathbf{y}\right) &= \pi\left(\mathbf{y}|\boldsymbol{\Theta}\right)\pi(\boldsymbol{\Theta}) \\
&= \pi\left(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\mu}_\beta, \boldsymbol{\mu}_\alpha, \boldsymbol{\Sigma}_\beta^{-1}, \boldsymbol{\Sigma}_\alpha^{-1}, \boldsymbol{\sigma}^{-2}, \delta\right) \\
&\quad \times \pi\left(\boldsymbol{\beta}|\boldsymbol{\alpha}, \boldsymbol{\mu}_\beta, \boldsymbol{\mu}_\alpha, \boldsymbol{\Sigma}_\beta^{-1}, \boldsymbol{\Sigma}_\alpha^{-1}, \boldsymbol{\sigma}^{-2}, \delta\right) \\
&\quad \times \pi\left(\boldsymbol{\alpha}|\boldsymbol{\mu}_\beta, \boldsymbol{\mu}_\alpha, \boldsymbol{\Sigma}_\beta^{-1}, \boldsymbol{\Sigma}_\alpha^{-1}, \boldsymbol{\sigma}^{-2}, \delta\right) \\
&\quad \times \pi\left(\boldsymbol{\mu}_\beta, \boldsymbol{\mu}_\alpha, \boldsymbol{\Sigma}_\beta^{-1}, \boldsymbol{\Sigma}_\alpha^{-1}, \boldsymbol{\sigma}^{-2}, \delta\right)
\end{aligned}
\right\}
\tag{2.13}
$$

where the last term on the right-hand side of equation (2.13) represents the prior specification at Stage 3 of our spatial-longitudinal bent-cable model. Now, using the Bayes' theorem, we can write the posterior density for our model parameters as:

$$
\pi\left(\boldsymbol{\Theta}|\mathbf{y}\right) = \frac{\pi(\boldsymbol{\Theta}, \mathbf{y})}{\pi(\mathbf{y})} = \frac{\pi(\mathbf{y}|\boldsymbol{\Theta})\pi(\boldsymbol{\Theta})}{\pi(\mathbf{y})}
\tag{2.14}
$$

where $\pi\left(\mathbf{y}|\boldsymbol{\Theta}\right)$ is already defined in equation (2.13) and $\pi\left(\mathbf{y}\right) = \int \int \ldots\ldots \int \pi\left(\mathbf{y}|\boldsymbol{\Theta}\right) \pi\left(\boldsymbol{\Theta}\right) d\boldsymbol{\Theta}$ is the usual normalizing factor. Since $\boldsymbol{\Theta}$ denotes all model parameters collectively, $\pi\left(\mathbf{y}\right)$ requires evaluation of multi-dimensional integrals.

In Bayesian analysis, inference about a particular parameter (a vector of parameters) is drawn based on its marginal posterior distribution. For example, inference about $\boldsymbol{\mu}_\beta$ is based on its marginal posterior distribution and worked out as:

$$
\pi\left(\boldsymbol{\mu}_\beta|\mathbf{y}\right) = \frac{\int \int \ldots\ldots \int \pi\left(\mathbf{y}|\boldsymbol{\Theta}\right) \pi\left(\boldsymbol{\Theta}\right) d\boldsymbol{\beta}\, d\boldsymbol{\alpha}\, d\boldsymbol{\mu}_\alpha\, d\boldsymbol{\Sigma}_\beta^{-1}\, d\boldsymbol{\Sigma}_\alpha^{-1}\, d\boldsymbol{\sigma}^{-2}\, d\delta}{\int \int \ldots\ldots \int \pi\left(\mathbf{y}|\boldsymbol{\Theta}\right) \pi\left(\boldsymbol{\Theta}\right) d\boldsymbol{\beta}\, d\boldsymbol{\alpha}\, d\boldsymbol{\mu}_\beta\, d\boldsymbol{\mu}_\alpha\, d\boldsymbol{\Sigma}_\beta^{-1}\, d\boldsymbol{\Sigma}_\alpha^{-1}\, d\boldsymbol{\sigma}^{-2}\, d\delta}
\tag{2.15}
$$

The main obstacle in Bayesian inference is the evaluation of multi-dimensional integrals as in equation (2.15). While it is occasionally possible to do such evaluations analytically in simple cases, it is almost impossible for complex models. An alternative technique is to use

Markov Chain Monte Carlo (MCMC) sampling to generate samples from the posterior. We discuss the MCMC methods in brief and the way of implementing it in next few sections.

2.2.2 Bayesian Inference and MCMC Methods

Determination of posterior distribution involves the evaluation of complex, multi-dimensional integrals. Such typical integration problems arise in Bayesian inference to compute normalizing constant and marginal posterior distribution for a particular parameter or a vector of parameters of interest from the posterior. In addition, we are interested about posterior summary, which involves computing moments or quantiles and leads to more integration. For example, consider the population slope parameter $\boldsymbol{\mu}_\beta$. Its marginal posterior density involves the evaluation of normalizing constant (denominator of equation (2.15)) and the numerator. We are interested to summarize the distribution of a function, say $h\left(\boldsymbol{\mu}_\beta\right)$ of $\boldsymbol{\mu}_\beta$, based on its posterior mean. Then we have to check out the integrals of the form:

$$E\left[h\left(\boldsymbol{\mu}_\beta\right)|\mathbf{y}\right] = \int h\left(\boldsymbol{\mu}_\beta\right)\pi\left(\boldsymbol{\mu}_\beta|\mathbf{y}\right)d\boldsymbol{\mu}_\beta$$

where, for illustration, $h\left(\boldsymbol{\mu}_\beta\right) = \mu_1$ (population incoming slope), or $h\left(\boldsymbol{\mu}_\beta\right) = \mu_1 + \mu_2$ (outgoing slope), or $h\left(\boldsymbol{\mu}_\beta\right) = \{\mu_1 - E[\mu_1|\mathbf{y}]\}^2$ (posterior variance of incoming slope).

Such typical, almost intractable, or even intractable integration problems in Bayesian inference can be solved using the MCMC technique (Andrieu et al., 2003). Once we have a Monte Carlo samples of the posterior, the statistical inference can be carried out in a straightforward manner by averaging over MCMC iterations.

2.2.3 MCMC Methods

Monte Carlo integration provides a way of evaluating a function, which is usually multi-dimensional and complex in nature, by drawing samples from its desired distribution, say $\pi(.)$ (see Gilks et al. (1995) for details). The mean of this function is then approximated by sample mean, and the approximation can be made as accurate as desired by increasing the number of draws due to the law of large number, provided drawing of samples are independent of each

other. In practice, drawing samples independently from $\pi(.)$ is not always straightforward, specially if $\pi(.)$ is a non-standard distribution.

Markov chain is a process of generating a sequence of random variables, where the value of any variable at a given state depends only on the value of its previous state. This type of chain is usually constructed in such a way that it can draw samples throughout the support of its stationary distribution in the correct proportions. Here, stationary distribution indicates the limiting distribution of a Markov chain. Now, we can override the independence assumption of sample from $\pi(.)$ using a Markov chain with $\pi(.)$ as its stationary distribution. Then the whole process is called Markov chain Monte Carlo (MCMC).

Let $\left\{\Theta^{(s)}, s = 1, 2, \ldots, T, \ldots\right\}$ denotes a realization from a Markov chain constructed in an appropriate manner. We say the distribution of $\left\{\Theta^{(s)}\right\}$ converges to a stationary distribution, say $\pi(.)$, if the chain satisfies three important properties (Gilks et al., 1995). First, the chain should be irreducible. The irreducible property states that the Markov chain can reach a non-empty set with positive probability from any starting point in some number of iterations. Second, the chain should be aperiodic in the sense that it cannot oscillate between different sets of states in a regular periodic movement. Finally, the chain must be positive recurrent. The third property ensures that if the starting value $\Theta^{(0)}$ is sampled from $\pi(.)$, then all subsequent iterations will be generated from $\pi(.)$. If the regularity conditions hold, then

$$\Theta^{(s)} \xrightarrow{d} \Theta \sim \pi\left(\Theta | \mathbf{y}\right) \quad \text{and}$$

$$\frac{1}{T} \sum_{s=1}^{T} h\left(\Theta^{(s)}\right) \xrightarrow{a.s.} E\left[h\left(\Theta\right) | \mathbf{y}\right] \quad \text{as} \quad T \longrightarrow \infty$$

the later result is known as ergodic theorem (Roberts, 1995). Under these regularity conditions, we burn-in a sufficient number of iterations, say S to make the chain independent of initial state and the chain eventually converges to its stationary distribution. We can, then, consider the sample mean or sample median from MCMC output as a point estimate of $h\left(\Theta\right)$ based on this ergodic theorem. Although theoretically a Markov chain converges under these conditions but, in practice, it is hard to verify the ergodicity conditions and usually we use convergence diagnostics to identify non-convergence.

In summary, once we have constructed a Markov chain that necessarily converges to its stationary distribution $\pi(.)$, we can summarize the marginal posterior distribution of a

parameter of interest through posterior mean, median, standard deviation and other summary statistics based on their sample equivalence in the MCMC output. We can also construct the $100\,(1-2a)\,\%$ credible interval (p_1, p_2) where p_1 and p_2 are the ath and $(1-a)^{th}$ quantiles of the marginal posterior density, respectively. For example, assuming $h\,(\Theta) = \mu_1$, $h\,(\Theta) = \mu_1 + \mu_2$ and $h\,(\Theta) = \exp\{\mu_\tau\} - \exp\{\mu_\gamma\} - 2\mu_1 \frac{\exp\{\mu_\gamma\}}{\mu_2}$ and discarding the burn-in samples, we can approximate the population incoming and outgoing slopes and the CTP as:

$$\hat{\mu}_1 = \frac{1}{T - S} \sum_{s=S+1}^{T} \mu_1^{(s)}$$

$$\widehat{\mu_1 + \mu_2} = \frac{1}{T - S} \sum_{s=S+1}^{T} \left(\mu_1^{(s)} + \mu_2^{(s)}\right)$$

$$\widehat{CTP} = \frac{1}{T - S} \sum_{s=S+1}^{T} \left(\exp\{\mu_\tau^{(s)}\} - \exp\{\mu_\gamma^{(s)}\} - 2\mu_1^{(s)} \frac{\exp\{\mu_\gamma^{(s)}\}}{\mu_2^{(s)}}\right)$$

2.2.4   Construction of Markov Chain

We describe how to form a Markov chain for our spatial-longitudinal bent-cable model here. We prefer the Metropolis within Gibbs algorithm (Smith & Roberts, 1993), a convenient form of Markov chain for nonlinear hierarchical regression model (Davidian & Giltinan, 1995; Wakefield et al., 1994). We first describe the Metropolis-Hastings algorithm in a general context, and then the Gibbs sampler in the context of our spatial-longitudinal bent-cable model with the problems associated in its implementation and finally the Metropolis within Gibbs algorithm. It is to be noted that both Gibbs and Metropolis within Gibbs sampler are based on the full conditionals of each of the components of Θ. Here, a full conditional refers to the distribution of a particular parameter, say β_i, conditioned on all the remaining ones and the data and is denoted by $\pi\left(\beta_i | \beta_1, \beta_2, \ldots, \beta_{i-1}, \beta_{i+1} \ldots, \beta_m, \boldsymbol{\alpha}, \boldsymbol{\mu}_\beta, \boldsymbol{\mu}_\alpha, \boldsymbol{\Sigma}_\beta^{-1}, \boldsymbol{\Sigma}_\alpha^{-1}, \boldsymbol{\sigma}^{-2}, \delta, \mathbf{y}\right)$.

Metropolis-Hastings Algorithm

Let $\xi \sim \kappa(.)$, where $\kappa(.)$ denotes a target distribution. The evaluation of $\kappa(.)$ is possible but sampling from $\kappa(.)$ is not straightforward. This target distribution may be a full conditional distribution in implementing the Gibbs algorithm, which may or may not have any closed form expression. The Metropolis-Hastings algorithm is usually used to generate a draw from

a distribution approximating $\kappa(.)$. Readers are referred to Hastings (1970) and Metropolis et al. (1953) for detailed description of this algorithm.

We choose a proposal distribution to sample a candidate point. Suppose, $\xi^{(0)}$ and g denote an arbitrary starting point and a proposal distribution, respectively. Given $\xi^{(s)}$ at the sth iteration, $\xi^{(s+1)}$ is generated according to this algorithm as:

1 We draw a candidate point ξ^* from $g(.|\xi^{(s)})$.

2 We compute the Metropolis-Hastings ratio

$$R\left(\xi^{(s)}, \xi^*\right) = \frac{\kappa\left(\xi^*\right) g\left(\xi^{(s)}|\xi^*\right)}{\kappa\left(\xi^{(s)}\right) g\left(\xi^*|\xi^{(s)}\right)}$$

3 Let U denotes a random variable from $\mathcal{U}(0, 1)$. If $U \leq \min\left(1, R\left(\xi^{(s)}, \xi^*\right)\right)$, we accept ξ^* and set $\xi^{(s+1)} = \xi^{(*)}$, otherwise $\xi^{(s+1)} = \xi^{(s)}$.

4 We increment s and return to 1.

As $\xi^{(s+1)}$ depends on the history only through the previous state, $\xi^{(s)}$, the chain is certainly Markov in nature. A proposal distribution may be of any form. However, the choice of the distribution has large influence on the rate of convergence to the stationary distribution. The target distribution is usually well approximated by a good proposal distribution because of the followings:

• The candidate values can cover the full support of the stationary distribution through a reasonable number of iterations.

• The acceptance or rejection rate of candidate values are not too frequent (Chib & Greenberg, 1995). The acceptance rate in the range $[0.15, 0.50]$ is suggested as reasonable by Roberts (1995) for high dimensional target distribution. The acceptance rate is kept restricted within this range to ensure that the chain can draw samples from full support of target distribution.

Metropolis-Hastings algorithm can take different forms for different forms of proposal distribution. Metropolis et al. (1953) considered only symmetric proposal distribution of the form $g\left(\xi^*|\xi^{(s)}\right) = g\left(\xi^{(s)}|\xi^*\right)$. As a result, the Metropolis-Hastings ratio reduces to

$$R\left(\xi^{(s)}, \xi^*\right) = \frac{\kappa\left(\xi^*\right)}{\kappa\left(\xi^{(s)}\right)}.$$

In Random-walk Metropolis algorithm (Gilks et al., 1995; Givens & Hoeting, 2005), often referred to as Metropolis algorithm, the candidate point is generated from $\xi^* = \xi^{(s)} + u$ where u denotes a point and is drawn from a symmetric proposal density $g(.)$. In such case, $g\left(\xi^*|\xi^{(s)}\right) = g\left(|\ \xi^* - \xi^{(s)}\ |\right)$ (Gilks et al., 1995).

Another version of Metropolis-Hastings algorithm is proposed by Tierney (1994) and Gelman and Rubin (1992b). It has the benefit of sampling from approximate full conditionals whilst maintaining exactly the required stationary distribution of the Markov chain. According to this method, an approximate full conditional can be used as a proposal distribution $g(.)$ in any independent Metropolis-Hastings, for which $g\left(\xi^*|\xi^{(s)}\right) = g\left(\xi^*\right)$. The metropolis-Hastings ratio becomes

$$R\left(\xi^{(s)}, \xi^*\right) = \frac{\kappa\left(\xi^*\right) g\left(\xi^{(s)}\right)}{\kappa\left(\xi^{(s)}\right) g\left(\xi^*\right)}.$$

Gibbs Sampler

The single-component Metropolis-Hastings algorithm was proposed by Metropolis et al. (1953). According to this method, the model parameters are partitioned into different components of possibly differing dimensions for computational efficiency and then each of those components are updated one by one. The Gibbs sampler (Geman & Geman, 1984), a special case of single-component Metropolis-Hastings algorithm, provides a method of obtaining the marginals of interest from the set of full conditionals, where the probability of accepting a candidate in a Metropolis step is one (Gilks et al., 1995).

As mentioned earlier, Θ denotes the parameters collectively for our spatial-longitudinal bent-cable model. The breakdown of Θ into different components is problem specific and is usually done based on practical considerations. It may be convenient to partition Θ into its scalar components for some problems, while for others the breakdown of Θ into vectors or matrices is straightforward and computationally efficient. This type of breakdown is usually

done based on two considerations from practical point of view. First, the components are chosen in such a way that we have full conditionals and closed form expressions. The reason is that the MCMC method, in general, is computationally expensive, specially when samples are generated from a distribution that can be expressed only upto a proportionality constant. Second, the model parameters are partitioned based on correlation structure in the sequence of $\Theta^{(s)}$ at different iteration lags. For highly correlated components, the convergence of a chain to its equilibrium distribution may be painfully slow as a result of very little movement at each conditional random variate generation step, if the components are treated individually (Smith & Roberts, 1993). Such problems may be avoided by blocking correlated components together and performing the random draws from multivariate conditional distributions as pointed out in Smith and Roberts (1993).

The implementation of Gibbs algorithm is straightforward. We choose an arbitrary starting point, say $\Theta^{(0)}$. We, then, generate an instance from the full conditional of each of the components of Θ conditional on the current values of the remaining elements, which completes a transition from $\Theta^{(0)}$ to $\Theta^{(1)}$. We, thus, generate a sequence of data $\Theta^{(0)}$, $\Theta^{(1)}$, ..., $\Theta^{(T)}$ from the full conditionals, which is a Markov chain with stationary distribution $\pi(\Theta|\mathbf{y})$ under regulatory conditions (Davidian & Giltinan, 1995). We can, then, apply the Monte Carlo sampling to the converged Markov chain for Bayesian inference.

It is to be noted that Gibbs sampler is usually applied to full conditionals, which have closed form expression. We worked out all the full conditionals in Appendices (see Section A.3) for all model parameters and show that all of these conditionals have closed form except for $\boldsymbol{\alpha}_i$ and δ. This is why the usual context of Gibbs sampler does not hold for our model.

Metropolis within Gibbs Algorithm

Metropolis within Gibbs sampler refers to an algorithm, where different Metropolis-Hastings variants are applied at different steps of a Gibbs sampler. This is particularly useful when all of the full conditionals do not have closed form expression. For our present model, the full conditionals for $\boldsymbol{\alpha}_i$ and δ can only be expressed up to a proportionality constant. We, therefore, employ Gibbs steps for all model parameters except for $\boldsymbol{\alpha}_i$ and δ, for which Metropolis steps are considered.

Now, we summarize the steps involved in the construction of the Markov chain for our model. Since the full conditional $\pi(. \mid .)$ for a particular parameter is considered as a function of the terms depending on that parameter, we exclude other components of Θ from the conditional expression $\pi(. \mid .)$. For example, the full conditional of $\boldsymbol{\beta}_i$ depends only on $\boldsymbol{\alpha}_i$, σ_i^{-2}, $\boldsymbol{\mu}_\beta$, $\boldsymbol{\Sigma}_\beta^{-1}$, δ and the data $\mathbf{y}_i$. Hence, we use the notation $\pi\left(\boldsymbol{\beta}_i \mid \boldsymbol{\alpha}_i, \sigma_i^{-2}, \boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta^{-1}, \delta, \mathbf{y}_i\right)$ to denote the conditional expression. Let $\left\{\Theta^{(0)}\right\}$ and $\left\{\Theta^{(s)}\right\}$ denote an arbitrary starting point and its update at iteration s, respectively. Given $\left\{\Theta^{(s)}\right\}$, we update the components in the following order to achieve the new set $\left\{\Theta^{(s+1)}\right\}$ in one iteration:

1 Generate $\boldsymbol{\beta}_i^{(s+1)} \sim$ Normal for $i = 1$ via a Gibbs step.

2 Generate $\boldsymbol{\alpha}_i^{(s+1)} \sim \pi(\boldsymbol{\alpha}_i \mid .)$ for $i = 1$ via a Metropolis step. Here, $\pi(\boldsymbol{\alpha}_i \mid .)$ can be expressed only upto a proportionality constant.

3 Generate $\sigma_i^{-2(s+1)} \sim$ Gamma for $i = 1$ via a Gibbs step.

4 Repeat steps $1 - 3$ for $i = 1, 2, \ldots, m$, which completes the updating of individual-specific parameters. Now, we have $\boldsymbol{\beta}^{(s+1)}$, $\boldsymbol{\alpha}^{(s+1)}$ and $\boldsymbol{\sigma}^{-2(s+1)}$.

5 Generate $\boldsymbol{\Sigma}_\beta^{-1(s+1)} \sim$ Wishart via a Gibbs step.

6 Generate $\boldsymbol{\Sigma}_\alpha^{-1(s+1)} \sim$ Wishart via a Gibbs step.

7 Generate $\boldsymbol{\mu}_\beta^{(s+1)} \sim$ Normal via a Gibbs step.

8 Generate $\boldsymbol{\mu}_\alpha^{(s+1)} \sim$ Normal via a Gibbs step.

9 Generate $\delta^{(s+1)} \sim \pi(\delta \mid .)$ via a Metropolis step. Here, $\pi(\delta \mid .)$ can be expressed only upto a proportionality constant.

We can, now, apply the Monte Carlo sampling for doing Bayesian inference.

2.2.5   Mixing and Convergence

The efficiency of an MCMC algorithm depends on its mixing ability. The mixing property of a chain refers to two characteristics. First, how quickly a chain forgets its initial values. Second,

how quickly a chain can explore the full support and shape of target distribution. Apart from good mixing, another concern is the convergence of a chain. That is, how efficiently the chain has approximated its stationary distribution. As the discussion in empirical literature about the goals of diagnosing convergence to the stationary distribution and investigating the mixing property of a chain is overlapping (referred in Khan (2010)), we summarize the review in the context of our spatial-longitudinal bent-cable model from Khan (2010).

Number of Chains

The literature regarding the number of chains to be used in MCMC algorithm is conflicting. Some scholars (Gelman & Rubin, 1992a, 1992b) suggest several long chains, while others (Geyer, 1992) recommend one very long chain. Merits and demerits of each method are discussed in Gilks et al. (1995) and Givens and Hoeting (2005). Even for a very long chain, the chain can reach around the mode of target distribution and can stay there forever. In such case, the convergence diagnostic may indicate the convergence of chain, although the chain does not fully explore the support and shape of target distribution. On the other hand, running multiple chain can ensure that at least one of them will explore the features of target distribution and will wash out the influence of initial values. As a compromise for our present case, we consider two very long independent chains ($5,000,000$ MCMC iteration for each) to analyse both CFC-11 and CFC-12 data.

Burn-in and Stopping Time

The dependence of a Markov chain on its starting value may remain strong even after the chain has been run for a sufficiently long period of time. As a consequence, if the chosen initial values are far different from the posterior mode, this dependency may make the chain converge slowly. In practice, an initial S iterations are discarded as a burn-in period to make the chain independent of its starting values.

Typically, a chain should stopped at a certain time after running the chain sufficiently long period to obtain good mixing. In practice, it is difficult to decide about the stopping time, T, in advance. Gilks et al. (1995) suggested an informal way of deciding the stopping time is to run several long chains and to compare the estimates (posterior means/medians)

from each chain. If the estimates produced by different chains do not agree closely, then the run length, $L = T - S$, should be increased.

Gelman-Rubin R statistic (Gelman & Rubin, 1992a) is a popular technique to determine the convergence as well as burn-in and run length of an MCMC output, if at least two chains are run. The statistic is based on a comparison of within-chain and between-chain variances. Let W and B denote the within and between chain variances, respectively for an MCMC sample consisting of u chains, each of v iterations. Then the empirical variance from all chains combined is calculated as:

$$\hat{\sigma^2}_{mc} = \frac{(v-1)W}{v} + \frac{B}{v}$$

Now, assuming that the target distribution is normal, we work out the convergence diagnostic, R as:

$$R = \sqrt{\frac{(d+3)\hat{V}}{(d+1)W}}$$

where $\hat{V} = \hat{\sigma}^2_{mc} + \frac{B}{uv}$ is the sample variance of all chains and $d = \frac{2 * \hat{V}^2}{\text{var}(\hat{V})}$ is the degrees of freedom. If the computed value of R is close to 1, convergence of the chain is assumed. In any case, if the value of R is not acceptable, then S or L or both should be increased.

Geweke Z-score (Geweke, 1992) is another way of assessing convergence based on two non-overlapping parts (usually first 0.1 and last 0.5 proportions) of a chain from an MCMC output. It compares the means of both parts, using a difference of mean test to see if the two parts come from the same distribution. The test statistic is defined as:

$$Z = \frac{\bar{\theta}_1 - \bar{\theta}_2}{\sqrt{\frac{s_1}{n_1} + \frac{s_2}{n_2}}}$$

where $\bar{\theta}_l$, s_l, and n_l, $l = 1, 2$ denote the estimated mean, standard error adjusted for autocorrelation, and number of iterations for l^{th} part of the chain. This is a two sided test, and large absolute Z-scores indicate the failure of convergence.

Thinning

Poor mixing of a chain is generally exhibited by slow decay of autocorrelation. It is, therefore, good practice that the inference should be based on every h^{th} iteration of a chain, with h set to some value high enough that successive draws are approximately independent (Gelman, 1995). This strategy is known as thinning in the literature. Thinning is also useful from computational point of view, where the set of simulated MCMC values is so large that reducing the number of simulations by a factor of h gives important savings in storage. We choose $h = 200$ in analysing both of CFC-11 and CFC-12 data based on the examination of the autocorrelation plot.

Graphical Diagnostics

Convergence and mixing property of a chain can also be examined through three widely used graphical tools: trace plot, autocorrelation plot and density plot.

A trace plot plots the realization of the chain versus the iteration number. It shows how rapidly the chain is mixing. Lack of trend in the trace plot indicates that stationarity has been achieved. On the other hand, a clear trend suggests that a longer run is necessary. A well behaved chain will move away from its starting values quickly, no matter where it started, and the samples will wiggle about vigorously in the supported region by the posterior density (Givens & Hoeting, 2005). Burn-in and stopping time can be determined through visual inspection of a trace plot.

An autocorrelation plot shows the serial correlation in the chain at different lags of iteration. The Gibbs sampler is slow to explore the entire support of the posterior density for a highly autocorrelated chain. In general, the autocorrelation decreases as the lag increases. If the situation is different, thinning should be explored.

In the absence of high autocorrelation, convergence is not guaranteed if the distribution is multi-modal. This is true if the kernel density plot displays not only multiple modes but also lumpiness rather than a smooth curve. It is to be noted that such behaviour in the density plot may result for high autocorrelation within a converged chain. In such cases, severity can be reduced through running the chain for a longer time and/or with heavier thinning.

2.3 Software Implementation

Since our spatial-longitudinal bent-cable model is an extension of longitudinal bent-cable model of Khan et al. (2009), we collect the code from the authors and modify them according to our model. The codes are written using two languages: R (R Core Team, 2012) and C. MCMC samples are generated by interfacing R and C. We call the standard functions for random sample generation from univariate distributions such as uniform, normal, gamma of R from C for generating MCMC samples. We also modified other functions required to generate samples from the full conditionals such as calculation of residual sum of squares, likelihood function, spatial weight matrix, data generation from multivariate normal and Wishart distribution and so on based on C.

Once MCMC samples are generated, we use the package "CODA" (Plummer et al., 2006) of R to perform standard statistical analysis. It is to be noted that there are several built-in functions in "CODA" package to analyse MCMC sample i.e. calculation of posterior summary, credible interval, convergence diagnostic criteria (Gelman-Rubin R statistic (Gelman & Rubin, 1992a), Geweke Z-score (Geweke, 1992)), and for drawing graphs (trace plot, density plot, and autocorrelation plot).

2.4 Discussion

The formulation of spatial-longitudinal bent-cable model through decomposing the error component has certain advantages. First, the introduction of spatial effects as a random element enable us to provide surrogate measure for factors which are neither measured nor observed (Gelfand et al., 2006). Although a number of variables are usually included in a statistical model as predictors, there still might be some unmeasured variables containing spatial information. Ignorance of those variables containing spatial data may result inaccurate parameter estimation, poor prediction, and inadequate quantification of uncertainty (Hoef et al., 2001). We overcome this situation by partitioning the error and introducing spatial effects as a random effect. Secondly, individual specific regression coefficients ($\boldsymbol{\theta}_i$) give us the opportunity of incorporating spatial heterogeneity rather than spatial dependence (Anselin et al., 2008).

Spatial heterogeneity is a special case of observed and unobserved heterogeneity. On the other hand, spatial dependence is assumed to present whenever correlation across cross-sectional units is non-zero, and the non-zero correlation follows a certain spatial order. Finally, modelling spatial effects based on continuous longitudinal data and random coefficient model is fairly straightforward (Anselin et al., 2008).

Chapter 3

Data Analysis and Simulation Study

In this chapter, we present the simulation study and two applications of our methodology. In Chapter 1, we presented the background of this study. We describe about data in details in Section 3.1. The detailed application is presented in Section 3.2. We conclude this chapter showing the efficacy of our methodology through simulation in Section 3.3.

Most of the materials in this chapter have taken from our journal article entitled "A Statistical Investigation to Monitor and Understand Atmospheric CFC Decline with the Spatial-longitudinal Bent-cable Model" appeared in the International Journal of Statistics and Probability published by Canadian Center of Science and Education(Khan et al., 2012). We repeat it here for the sake of completeness of this thesis.

3.1 Data

CFCs are monitored from different stations all over the globe by the Global Monitoring Division of the National Oceanic and Atmospheric Administration (NOAA/ESRL Halocarbons Group, 2012), and Advanced Global Atmospheric Gases Experiment (*AGAGE*, 2012; Prinn et al., 2000, 2001; Cunnold et al., 1997) program. Both these programs currently use gas chromatographic (GC) technique (known as *in situ* program) to measure CFCs.

The *in situ* program of NOAA/ESRL started in 1987. The original program is called Radiatively Important Trace Species (RITS) program. The RITS GCs were deployed at five NOAA observatories: (1) Barrow, Alaska, (2) Cape Matatula, American Samoa, (3) Mauna Loa, Hawaii, (4) South Pole, Antarctica, and (5) Niwot Ridge, Colorado. However, widespread use of the replacement compounds to the CFCs prompted improvement of the RITS program. A new generation of GC called Chromatograph for Atmospheric Trace Species

(CATS) was developed in 1998. The CATS system replaced the RITS *in situ* instrumentation in 1998, and has been used to measure CFCs ever since.

The ALE/GAGE/AGAGE program (Prinn et al., 2000, 2001; Cunnold et al., 1997) consists of three phases corresponding to advances and upgrades in GC techniques: ALE (began in 1978), GAGE (began between 1981 and 1985), and AGAGE (began between 1993 and 1996). Under this program, CFCs are measured from stations located in (6) Mace Head, Ireland, (7) Cape Grim, Tasmania, (8) Ragged Point, Barbados, (9) Cape Matatula, American Samoa, and (10) Trinidad Head, California.

Table 3.1: Geographical locations (latitude, longitude, and elevation in metres above sea level (masl)) of the stations and the instrumentations used to record data.

	Latitude[1]	Longitude[1]	Elevation[1](in masl)	Instrumentation[1]
Barrow, Alaska	71.32	−156.60	11	RITS/CATS
Cape Matatula, American Samoa	−14.24	−170.57	42	RITS/CATS
Mauna Loa, Hawaii	19.54	−155.58	3397	RITS/CATS
South Pole, Antarctica	−89.98	−24.80	2,810	RITS/CATS
Niwot Ridge, Colorado	40.04	−105.54	3,018	RITS/CATS
Mace Head, Ireland	53.33	−9.90	5	GAGE/AGAGE
Cape Grim, Tasmania	−40.68	144.68	94	GAGE/AGAGE
Ragged Point, Barbados	13.17	−59.43	45	GAGE/AGAGE

[1] Source: NOAA/ESRL Halocarbons Group (2012) and *AGAGE* (2012)

We take into account monthly mean CFC data based upon two considerations: (a) the availability of complete data, and (b) the necessity of a sufficiently long study period to capture the trend change over time. For (a), we choose eight monitoring stations (stations 1-8) for which data are available until September of 2010. Although the CFC monitoring process began in the late 1970's for some stations, complete data for all stations are not available until January of 1988. Therefore, we choose a study period beginning from January of 1988 to September of 2010 (273 months), which is also long enough to observe the trend change over time. Table 3.1 summarizes the geographical locations of the stations and the instrumentations used to record data during the period of study.

We replotted Figure 1.3 and 1.4 in Figure 3.1 and 3.2, respectively for the convenience of reader. In general, each station shows an individual curve consisting of an increasing trend, a gradual transition and a decreasing trend after the transition. The differences among the curves may be attributed to actual level of CFCs during measurement, exposure to wind as
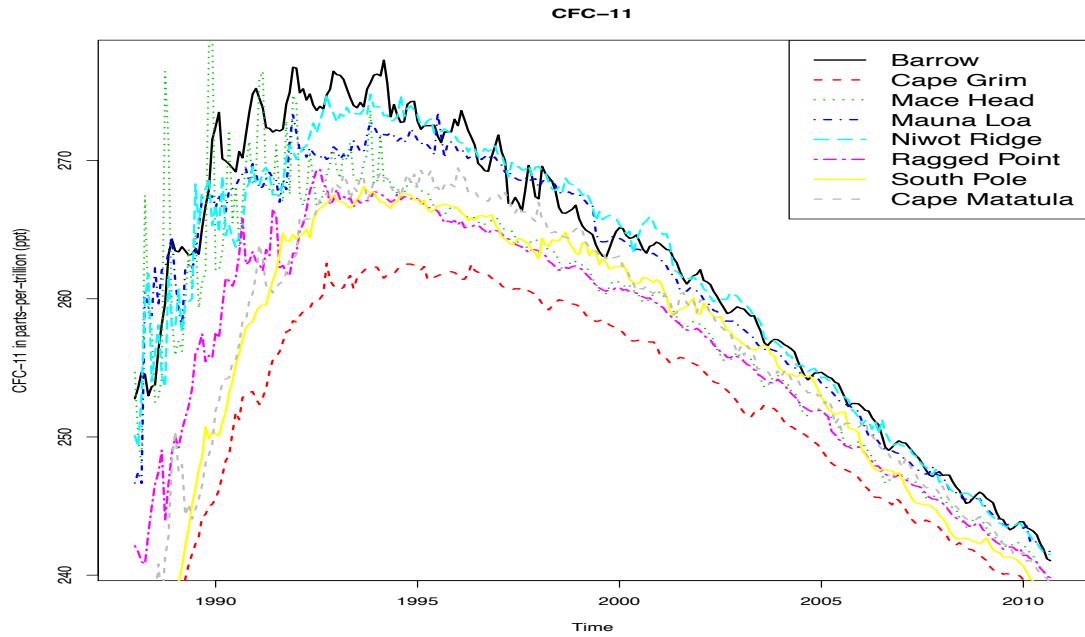
Figure 3.1: Monthly mean profile of 8 stations for CFC-11 over the period 1988 to 2010. We reproduced Figure 1.3 here.
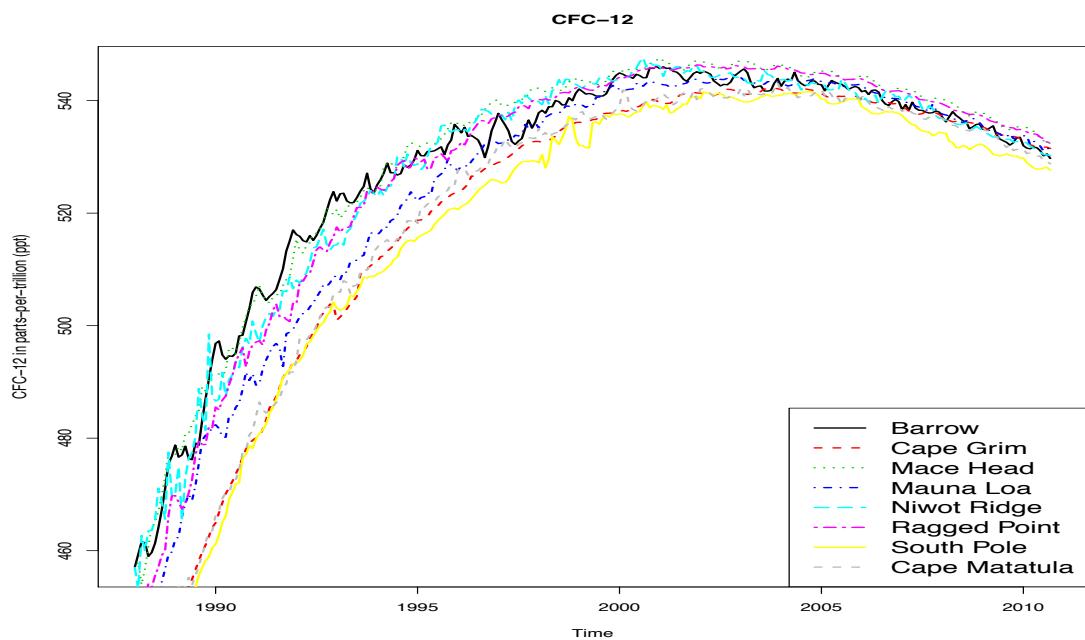
Figure 3.2: Monthly mean profile of 8 stations for CFC-12 over the period 1988 to 2010. We reproduced Figure 1.4 here.

well as other environmental variables, sampling techniques and so on. We are interested to assess the global and regional concentration of CFCs in the atmosphere.

We construct the t_{ij} following Khan (2010). For illustration, we assign $t_{1,1} = 1$ for January, 1988 and $t_{1,273} = 273$ for September, 2010 for the first station: Barrow, Alaska. In a similar way, we allocate $t_{2,1} = 1$ and $t_{2,273} = 273$ for second station for the first and last month with recorded data, respectively. Due to failure of instrument and other unavoidable reasons, some data are missing (only 3.94% for CFC-11 and 2.70% for CFC-12). Missing data are replaced by observation from another data set (e.g. flask data or CATS or AGAGE), if available, or by mean imputation based on 6 neighbouring time points: 3 from earlier and 3 from later time series. In any case, if 6 consecutive time points are not available, we impute them by available observations. The deleterious effect of mean substitution is minimal if only a few missing values are replaced by the mean (McKnight et al., 2007). Hence, our analysis is expected to be minimally affected for missing value replacement.

3.2 Data Analysis

We apply our spatial-longitudinal bent-cable model to CFC-11 and CFC-12 data. There are some similarities as well as dissimilarities between CFC-11 and CFC-12. The main dissimilarity is in their atmospheric lifetime. This disparity leads to quite different estimates of model parameters and other concerns. We report all parameter estimates based on their posterior mean or median, depending on the extent of asymmetry of the corresponding marginal posterior density in our application. We, then, produce a fitted curve as determined by the instances of the regression coefficients in the Markov chain.

3.2.1 The Spatial Matrix and Model Building

For the CFC data, since wind can spread CFCs all over the world, there is no clearly defined boundary by which the locations of the monitoring stations can be differentiated. Therefore, we consider several models by defining the spatial configurations based on geodetic distances among the stations. Specifically, letting d_{ik} the distance between stations i and k, we consider the following spatial configurations:

Model 1: $\mathbb{W} = \mathbb{I}$ (i.e., no spatial effects),

Model 2: $i \sim k$ if $d_{ik} \leq 6,000$ km,

Model 3: $i \sim k$ if $d_{ik} \leq 10,000$ km,

Model 4: $i \sim k$ if $d_{ik} \leq 20,000$ km.

Then the elements of the spatial matrix $\mathbb{W}$ is defined by

$$
w_{ik} = \begin{cases} 1, \text{if } i \neq k \text{ and } i \sim k \\ 0, \text{ otherwise} \end{cases} \tag{3.1}
$$

where $i \sim k$ denotes the contiguity of station i with station k. The specification of the spatial weights is often chosen on an ad hoc basis (Anselin et al., 2008). Under the specification of Model 4, each station is a neighbour of the remaining one, which hypothesizes that CFCs are distributed all around the globe. We normalize $\mathbb{W}$ by dividing all of its elements by its largest characteristic root. Finally, we choose one model as best for which the estimate of the Deviance Information Criteria (DIC) is minimum (Spiegelhalter et al., 2002).

Table 3.2: Model Comparison for the CFC Data Using DICs

Spatial	DIC	
Configuration	CFC-11	CFC-12
Model 1	$7,899$	$10,681$
Model 2	$7,586$	$9,428$
Model 3	$7,546$	$8,935$
Model 4	$7,379$	$8,291$

We summarize the DICs for CFC data in Table 3.2. The smallest DIC results from Model 4 for both CFC-11 and CFC-12. Hence, according to the DIC values, Model 4 is estimated to be the model that would best fit the data among the models under consideration. This also supports the notion of the global distribution of the CFCs (i.e., each station is a neighbour of the remaining ones). Now, we report the results for Model 4.

3.2.2 CFC-11
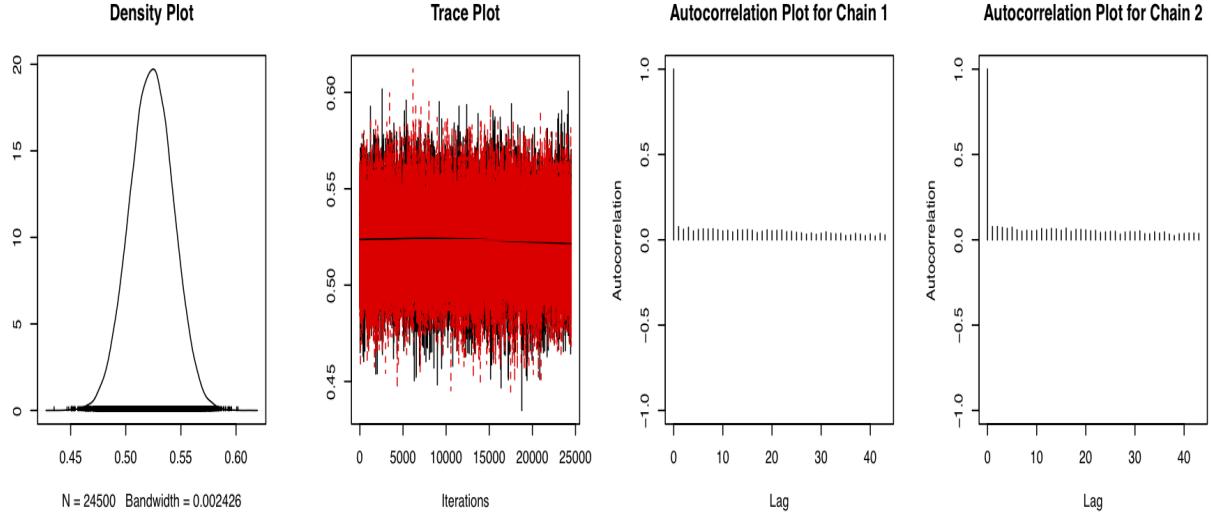
Figure 3.3: Density, trace and autocorrelation plots for the posteriors of the spatial autocorrelation parameters δ from two chains

The density, trace and autocorrelation plots of δ for CFC-11 are presented in Figure 3.3. The density plots display no signs of multimodality. The lack of any trend in the trace plot indicates good mixing. These plots show stationarity of the Markov chains. We also test the stationarity of the chains for δ by the Gelman-Rubin R statistic (Gelman & Rubin, 1992a) and the Geweke Z score (Geweke, 1992). Value of R is 1 and the $|Z|$ score is 1.28, which is less than the critical value 1.96 at 5% level of significance. So, these two diagnostic criteria also provide no evidence against stationarity of the chains for δ for CFC-11.

The posterior mean for δ is 0.52 with 95% credible interval (0.48, 0.58). This indicates significant spatial autocorrelation for CFC-11 as the credible interval does not include zero. Table 3.3 quantifies the posterior characteristics of the global concentrations for CFC-11. The global drop occurred between August, 1988 and September, 1995 approximately. The posterior mean for global incoming and outgoing slopes are 0.64/month and -0.16/month, respectively. Both of the slopes appeared as statistically significant as their 95% credible intervals do not include zero. Thus, on average, CFC-11 was increasing at a rate of 0.64 ppt per month during the incoming phase (January, 1988 to July, 1988) and decreasing at 0.16

Table 3.3: Posterior summaries of the global regression parameters for CFC-11

Parameter	Posterior Mean	Convergence Diagnostic	
	95% Credible Interval	Gelman-Rubin R	Geweke $|Z|$
μ_0	244.28	1.00	0.76
(Incoming Intercept)	(233.28, 255.62)		
μ_1	0.64	1.00	0.62
(Incoming Slope)	(0.45, 0.83)		
$\mu_1 + \mu_2$	-0.16	1.00	0.64
(Outgoing Slope)	$(-0.313, -0.002)$		
$\exp\{\mu_\tau\} - \exp\{\mu_\gamma\}2\mu_1\exp\{\mu_\gamma\}/\mu_2$	May, 1994	1.00	0.48
(CTP)	(Nov. 1992 to Oct. 1995)		
$\exp\{\mu_\tau\} \pm \exp\{\mu_\gamma\}$	Aug. 1988 to	1.00, 1.00	0.91, 0.68
(Transition Period)	Sep. 1995		

ppt per month during the outgoing phase (October, 1995 to September, 2010). The posterior mean for the global CTP is estimated as May, 1994 with 95% credible interval ranging from November, 1992 to October, 1995. So, overall CFC-11 went from increasing to decreasing state around May, 1994 across all stations. CFC-11 took about 85 months (August, 1988 to September, 1995) to complete its transition.
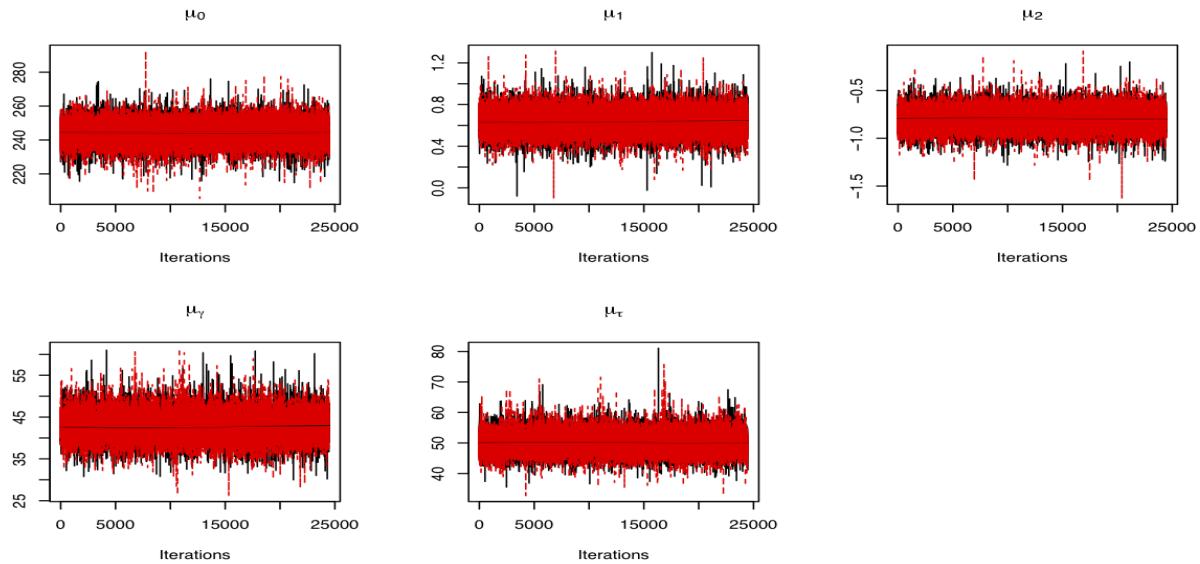
Figure 3.4: Trace plots for the posteriors of the population parameters μ_0, μ_1, μ_2, μ_γ and μ_τ from two chains

Figures 3.4, 3.5 and 3.6 shows the trace, density and autocorrelation plots, respectively, for each of the population coefficients μ_0, μ_1, μ_2, μ_γ and μ_τ. The lack of trends in the trace
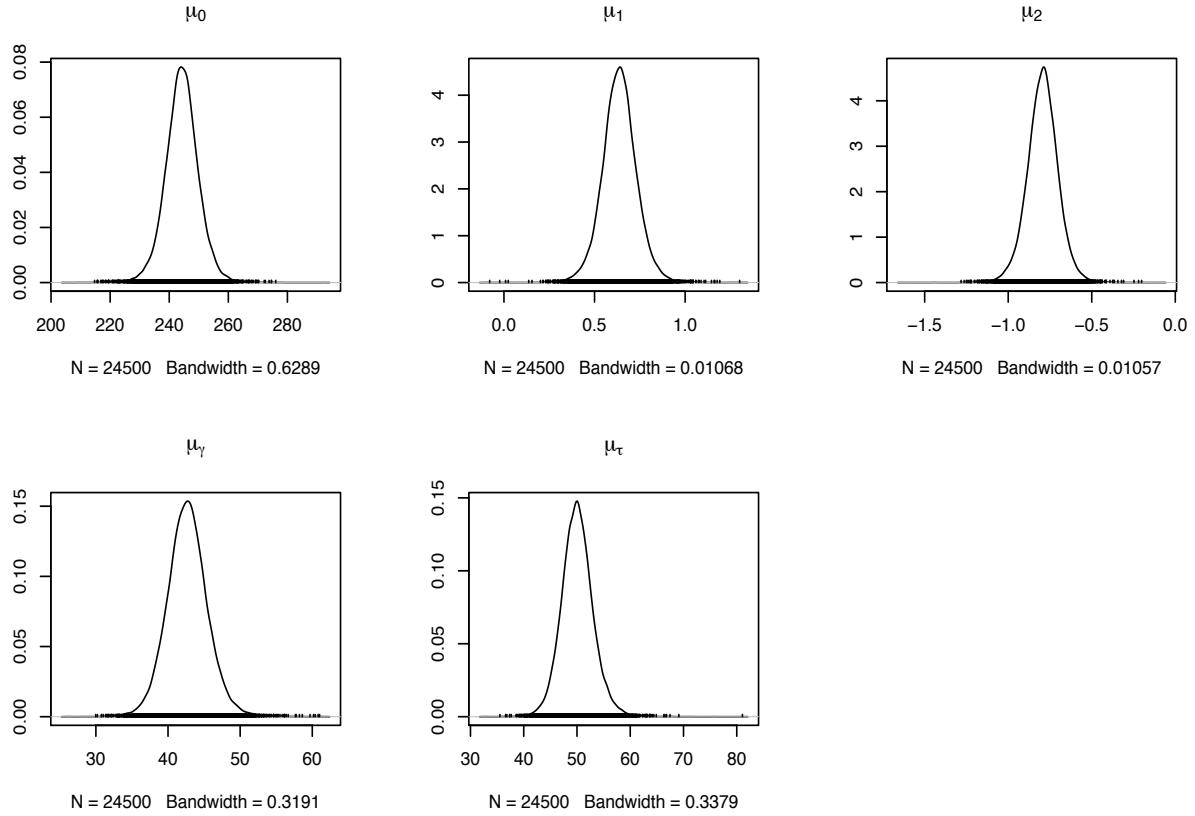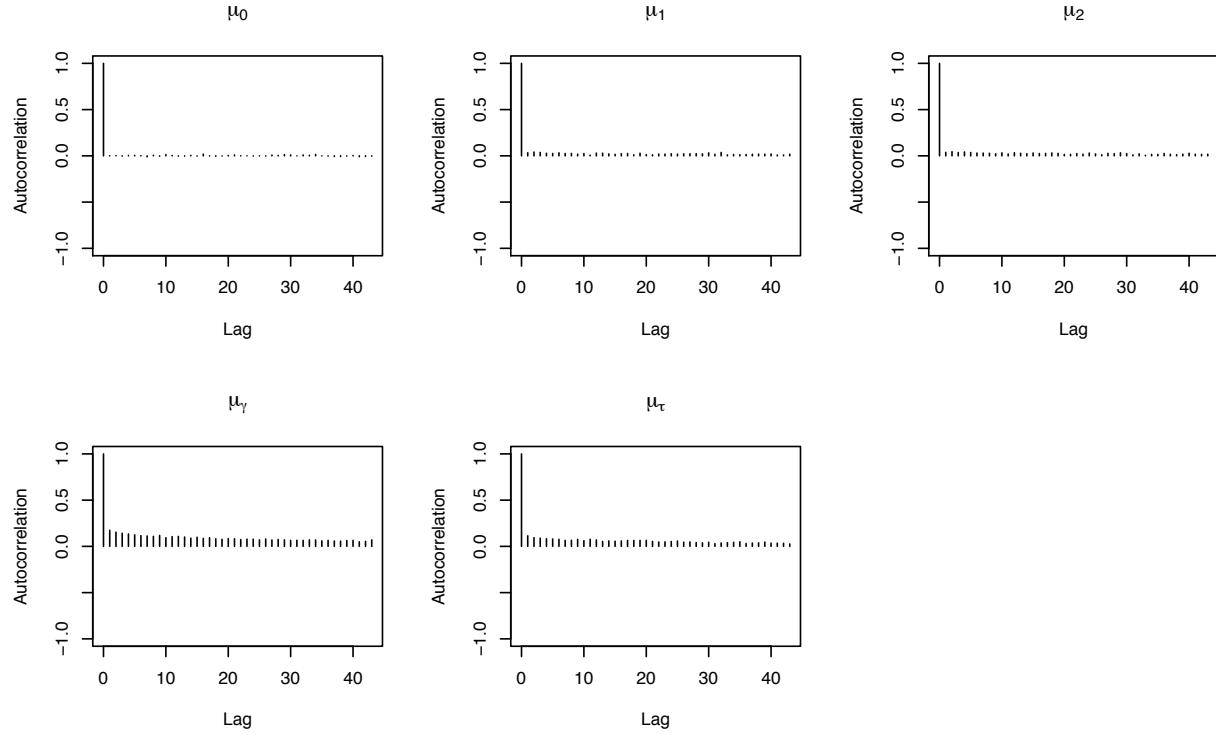
Figure 3.5: Kernel density estimate plots for the posteriors of the population parameters μ_0, μ_1, μ_2, μ_γ and μ_τ from two chains

plot and low autocorrelation in the two chains jointly indicate good mixing of Markov chain. There is no sign of multi-modality in the density plot. These plots show the stationarity of the chain. Furthermore, evidence of stationarity is also clear from the Gelman-Rubin R statistics and Geweke $|Z|$ scores reported in Table 3.3.

The global and station-specific fits are displayed in Figure 3.7 and 3.8, respectively. They reveal that our model fits the data quite well, with the observed data and corresponding fits agreeing closely. We summarize the station-specific numerical fits in Table 3.4 and the posterior characteristics of the standard deviations associated with Σ_β and Σ_α (prior for the station-specific random regression coefficients) in Table 3.5. From Table 3.4, some variations are observed in the estimates of the intercepts across the stations. The incoming and outgoing slopes appeared as statistically significant and agree closely for all stations. The large variation in estimated intercepts and small variation in estimated slopes are also supported

Chain 1

Chain 2
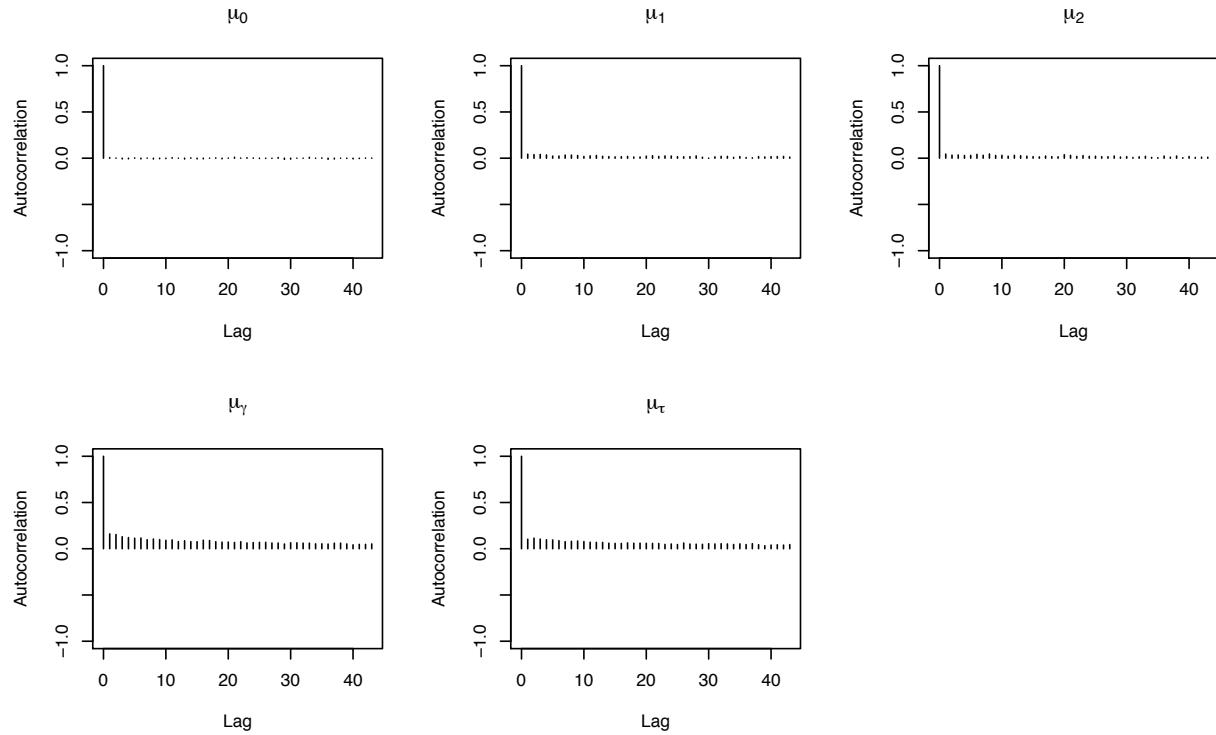
Figure 3.6: Autocorrelation Plots for the two Chains for Population Parameters μ_0, μ_1, μ_2, μ_γ and μ_τ
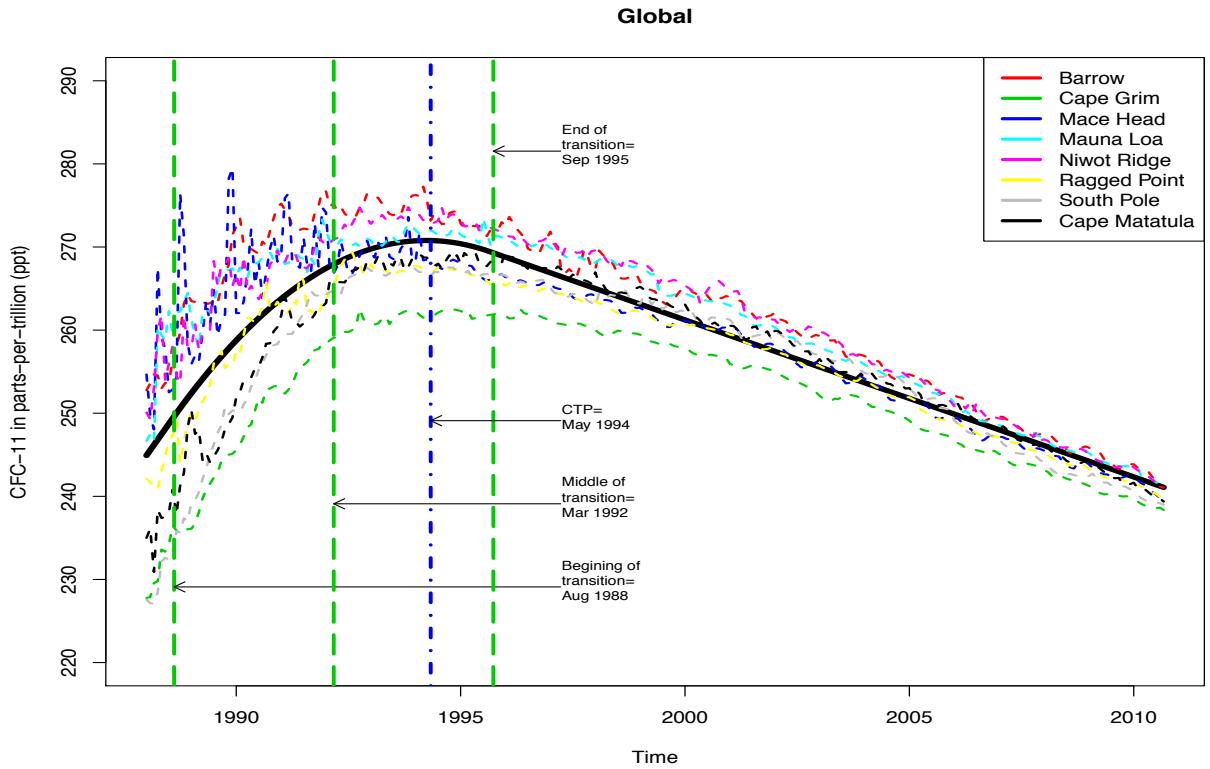
Figure 3.7: Observed data and corresponding population (Global) fit (solid black) for CFC-11
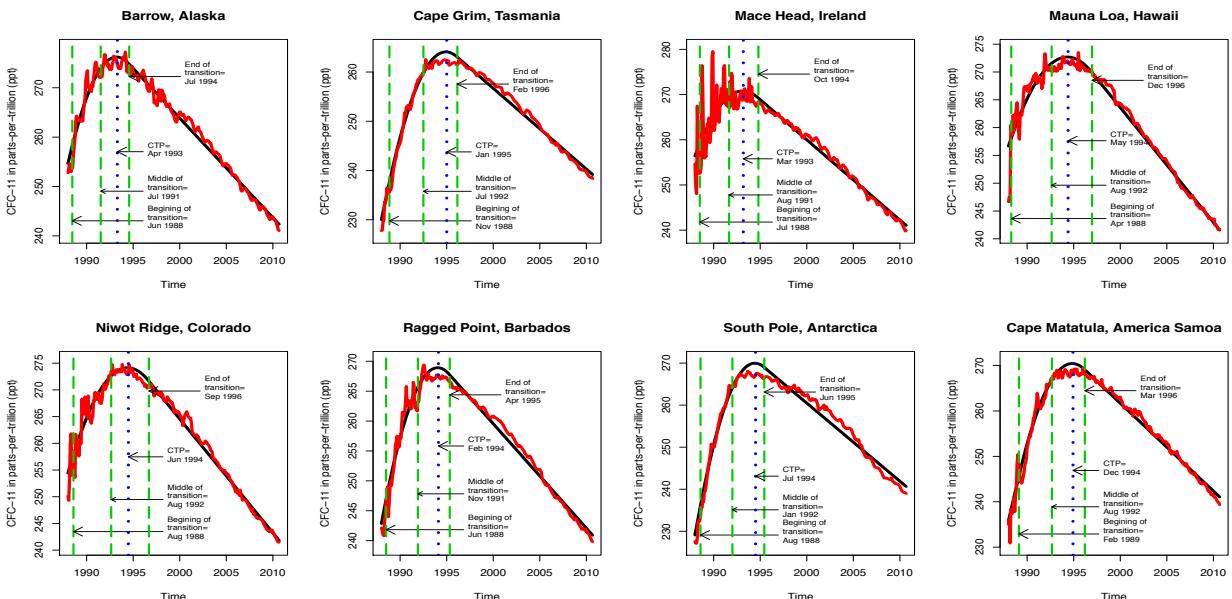
Figure 3.8: Observed data (red curve)and corresponding station-specific fitted (black) curve for CFC-11

Table 3.4: Posterior Summaries of the Station-specific Regression Parameters

Station	β_{0i} (95% Credible Interval)	β_{1i} (95% Credible Interval)	$\beta_{1i} + \beta_{2i}$ (95% Credible Interval)	Transition Period (Duration)	CTP (95% Credible Interval)
Barrow, Alaska $\hat{\sigma_1}=1.17$	254.03 $(253.23, 254.82)$	0.64 $(0.58, 0.71)$	-0.168 $(-0.171, -0.165)$	06/1988 to 07/1994 (73 months)	04/1993 (03/1993 to 06/1993)
Cape Grim, Tasmania $\hat{\sigma_2}=0.80$	229.29 $(228.61, 230.01)$	0.74 $(0.67, 0.82)$	-0.137 $(-0.139, -0.134)$	11/1988 to 02/1996 (87 months)	01/1995 (11/1994 to 02/1995)
Mace Head, Ireland $\hat{\sigma_3}= 2.47$	255.99 $(254.44, 257.54)$	0.43 $(0.34, 0.52)$	-0.148 $(-0.154, -0.142)$	07/1988 to 10/1994 (75 months)	03/1993 (11/1992 to 07/1993)
Mauna Loa, Hawaii $\hat{\sigma_4}=1.28$	256.26 $(255.54, 257.01)$	0.41 $(0.37, 0.45)$	-0.172 $(-0.176, -0.168)$	04/1988 to 12/1996 (104 months)	05/1994 (04/1994 to 07/1994)
Niwot Ridge, Colorado $\hat{\sigma_5}= 1.24$	253.96 $(253.20, 254.76)$	0.48 $(0.42, 0.53)$	-0.178 $(-0.181, -0.174)$	08/1988 to 09/1996 (85 months)	06/1994 (05/1994 to 08/1994)
Ragged Point, Barbados $\hat{\sigma_6}= 1.08$	242.19 $(241.45, 242.93)$	0.67 $(0.61, 0.74)$	-0.146 $(-0.149, -0.143)$	06/1988 to 04/1995 (82 months)	02/1994 (12/1993 to 03/1994)
South Pole, Antarctica $\hat{\sigma_7}=1.38$	228.13 $(227.17, 229.08)$	0.98 $(0.90, 1.07)$	-0.155 $(-0.159, -0.151)$	08/1988 to 06/1995 (82 months)	07/1994 (05/1994 to 08/1994)
Cape Matatula, American Samoa $\hat{\sigma_8}=1.18$	235.08 $(234.03, 236.17)$	0.73 $(0.64, 0.84)$	-0.161 $(-0.165, -0.158)$	02/1989 to 03/1996 (83 months)	12/1994 (10/1994 to 01/1995)

by the estimated Σ_β (Table 3.5). The estimate of standard deviations for β_{0i}, β_{1i} and β_{2i} are 15.18, 0.25 and 0.25, respectively.

Our findings support the notion of constant rate of increase and decrease before and after the enforcement of the Montreal protocol, respectively. The widespread adoption and implementation of the protocol across the globe is evident from the findings also. However, CFC-11 will remain in the atmosphere throughout the 22nd century, should current conditions prevail - suggested by its global decreasing rate (0.16 ppt per month, globally).

Table 3.5: Posterior Summaries of the Standard Deviations of the Random Regression Coefficients

Parameter	Posterior Mean	95% Credible Interval
$SD(\beta_{0i})$	15.18	(8.29, 23.89)
$SD(\beta_{1i})$	0.25	(0.14, 0.40)
$SD(\beta_{2i})$	0.25	(0.13, 0.39)
$SD(\gamma_i)$	6.45	(3.42, 11.59)
$SD(\tau_i)$	7.19	(3.83, 12.83)

The transition period and critical time point varied somewhat across stations (Table 3.4). Such variations is also reflected in the estimate of Σ_α given in Table 3.5. The standard deviations for γ_i and τ_i are estimated as 6.45 and 7.19, respectively. This variation may be attributed to different schedules of phase-out for different countries in the Montreal protocol - 1996 for developed countries and 2010 for developing countries. Thus, while developed countries had stopped CFC consumption, the developing countries continued to contribute CFC in the atmosphere. The transition began between April, 1988 and February, 1989 - a period of 11 months only. We observed almost same duration of transition period for all stations except Mauna Loa, Hawaii. So, CFC-11 took almost same time to start dropping linearly in different parts of the world.

We report the estimates of the innovation variances (σ_i^2's) in the first column of Table 3.4. The profile plot (Figure 3.1) reveals that Mace Head measurements are more variable, whereas Cape Grim show little variation over time. The estimated innovation variance (2.47 for Mace Head and 0.80 for Cape Grim) also support the fact.

In summary, CFC-11 started to decrease globally since the Montreal protocol came into effect. The spatial autocorrelation among CFC-11 data observed at different stations worldwide is of moderate level and statistically significant. This finding makes scientific sense due to the fact that CFC-11 has an extended life time of 45 years and during this period it is transported vertically and horizontally several times by wind and convection. The continuing decreases in the global CFC-11 trend suggests that the Montreal protocol can be regarded as successful to reduce the negative impact of CFC-11 on the ozone layer.

3.2.3 CFC-12

We plot the density, trace and autocorrelation plots of δ for CFC-12 in Figure 3.9. These plots show stationarity of the Markov chain. The value of Gelman-Rubin R statistic (Gelman & Rubin, 1992a) is again 1.00 and the Geweke Z score (Geweke, 1992) is 0.53, which is also less than the critical value at 5% level of significance. Hence, both diagnostic criteria reveal no evidence against stationarity of the chains for δ for CFC-12.

Like CFC-11, the posterior mean of δ for CFC-12 appeared as statistically significant (95% credible interval 0.76-0.80) and estimated as 0.78. The estimate indicates that CFC-12
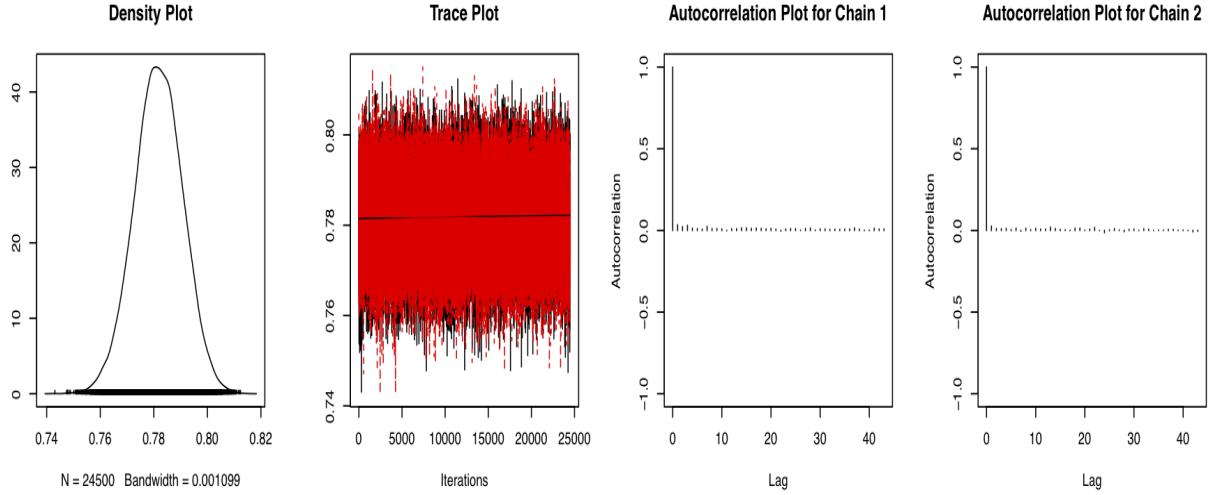
Figure 3.9: Density, trace and autocorrelation plots for the posteriors of the spatial autocorrelation parameters δ from two chains

is highly spatially autocorrelated across the globe. We summarize the posterior characteristics of the global concentration of CFC-12 in Table 3.6. The global drop took place between May, 1988 and April, 2001 approximately. CFC-12, on average, was increasing at a rate of 1.23 ppt per month (about double of CFC-11) during the incoming phase (January, 1988 to April, 1988) and decreasing at 0.08 ppt per month (about half of CFC-11) during the outgoing phase (May, 2001 to September, 2010). Unlike CFC-11, the outgoing slope appeared as insignificant. The higher incoming and lower outgoing slopes may be attributed to its super extended atmospheric lifetime, about 100 years.

Table 3.6: Posterior Summaries of the Global Regression Parameters for CFC-12

Parameter	Posterior Mean	Convergence Diagnostic	
	95% Credible Interval	Gelman-Rubin R	Geweke $|Z|$
μ_0	449.62	1.00	0.04
(Incoming Intercept)	(435.98, 463.47)		
μ_1	1.23	1.00	0.98
(Incoming Slope)	(1.04, 1.41)		
$\mu_1 + \mu_2$	-0.08	1.00	0.44
(Outgoing Slope)	$(-0.18, 0.02)$		
$\exp\{\mu_\tau\} - \exp\{\mu_\gamma\}2\mu_1\exp\{\mu_\gamma\}/\mu_2$	Jun, 2000	1.00	1.41
(CTP)	(Jun. 1999 to Jun. 2001)		
$\exp\{\mu_\tau\} \pm \exp\{\mu_\gamma\}$	May 1988 to	1.00, 1.01	0.99, 0.93
(Transition Period)	Apr. 2001		

41

The global CTP is estimated as June, 2000 with 95% credible interval ranging from June, 1999 to June, 2000. Thus, overall CFC-12 trend went from increasing to decreasing state around June, 2000 across all stations. Moreover, it took about 155 months (May, 1988 to April, 2001) to complete the transition.
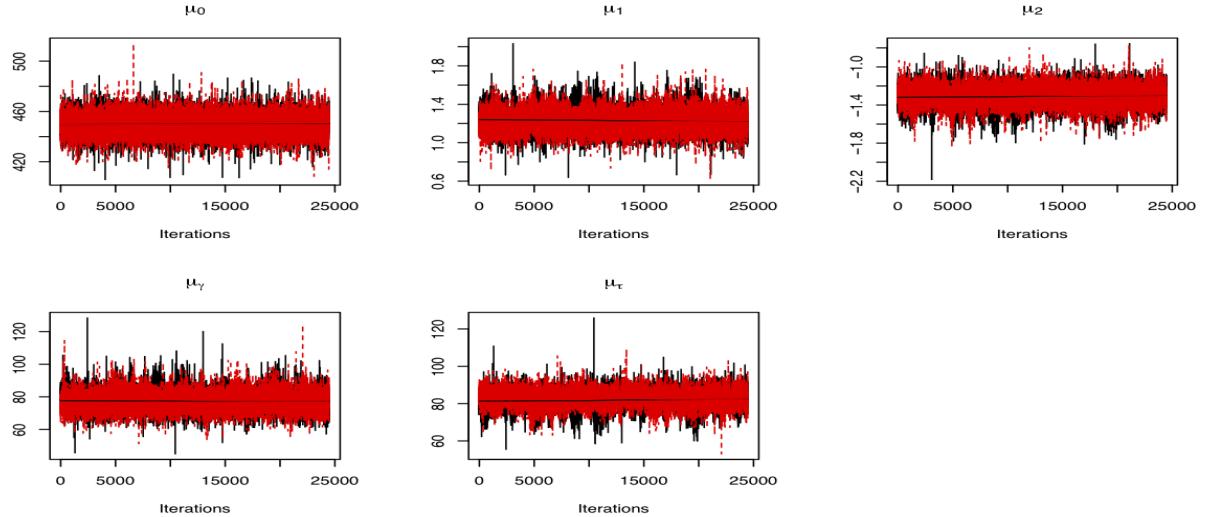
Figure 3.10: Trace Plots for the Posteriors of the Population Parameters μ_0, μ_1, μ_2, μ_γ and μ_τ from two Chains
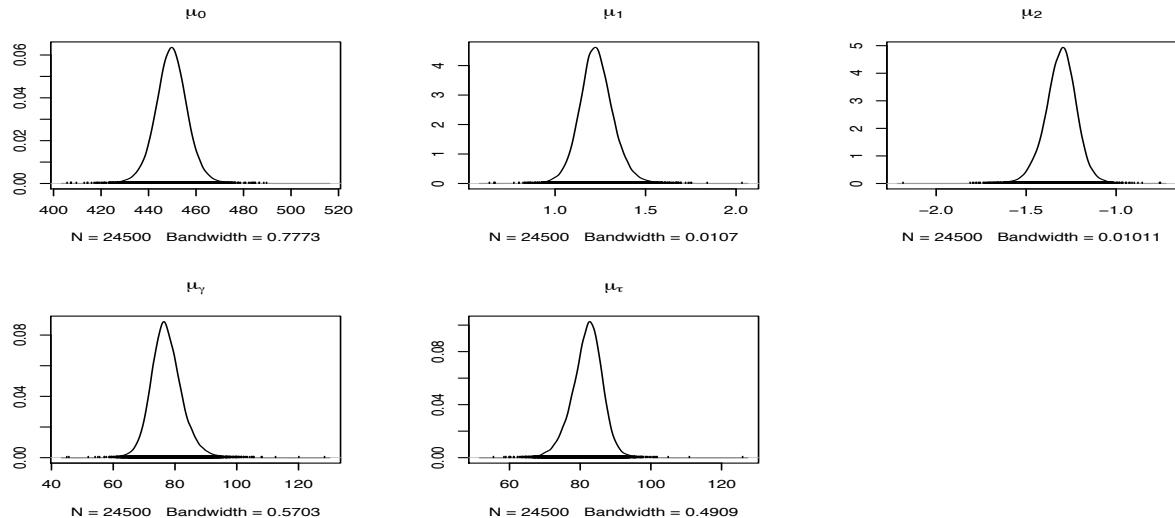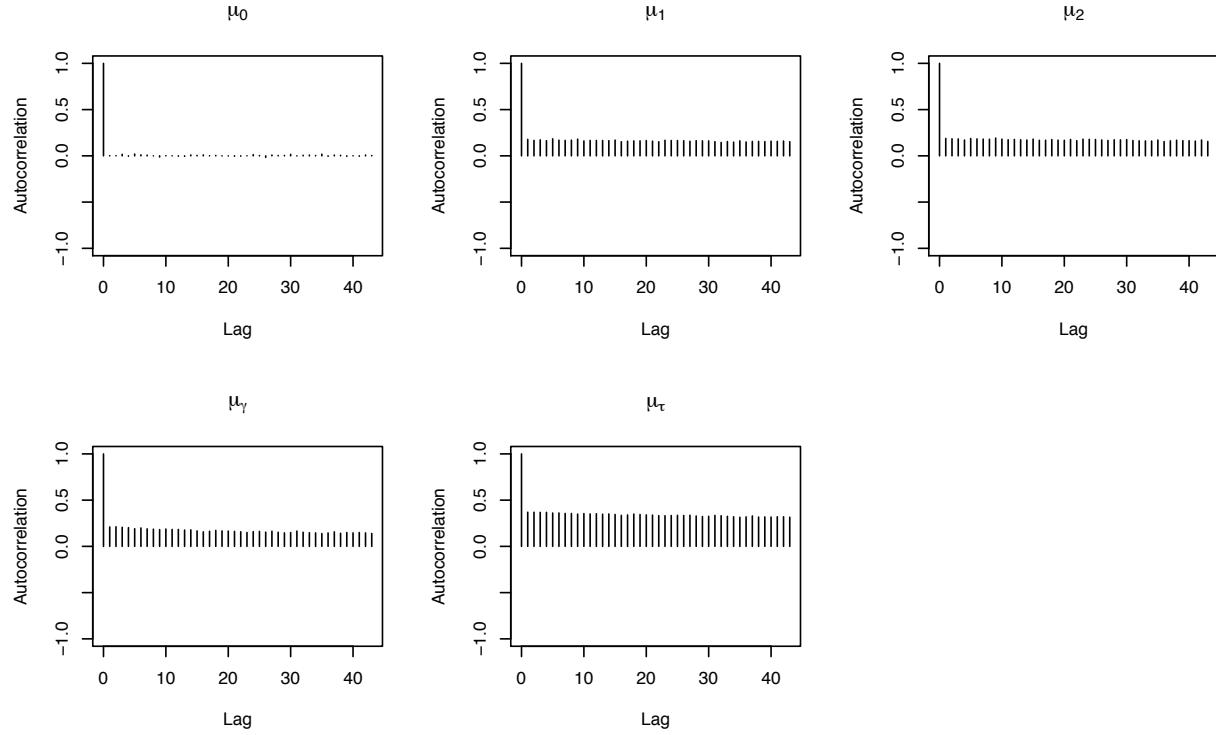
Figure 3.11: Kernel Density Estimate Plots for the Posteriors of the Population Parameters μ_0, μ_1, μ_2, μ_γ and μ_τ from two Chains

We report the trace, density and autocorrelation plots for each of the population coefficients μ_0, μ_1, μ_2, μ_γ and μ_τ in Figure 3.10, 3.11 and 3.12, respectively. Like former, the lack
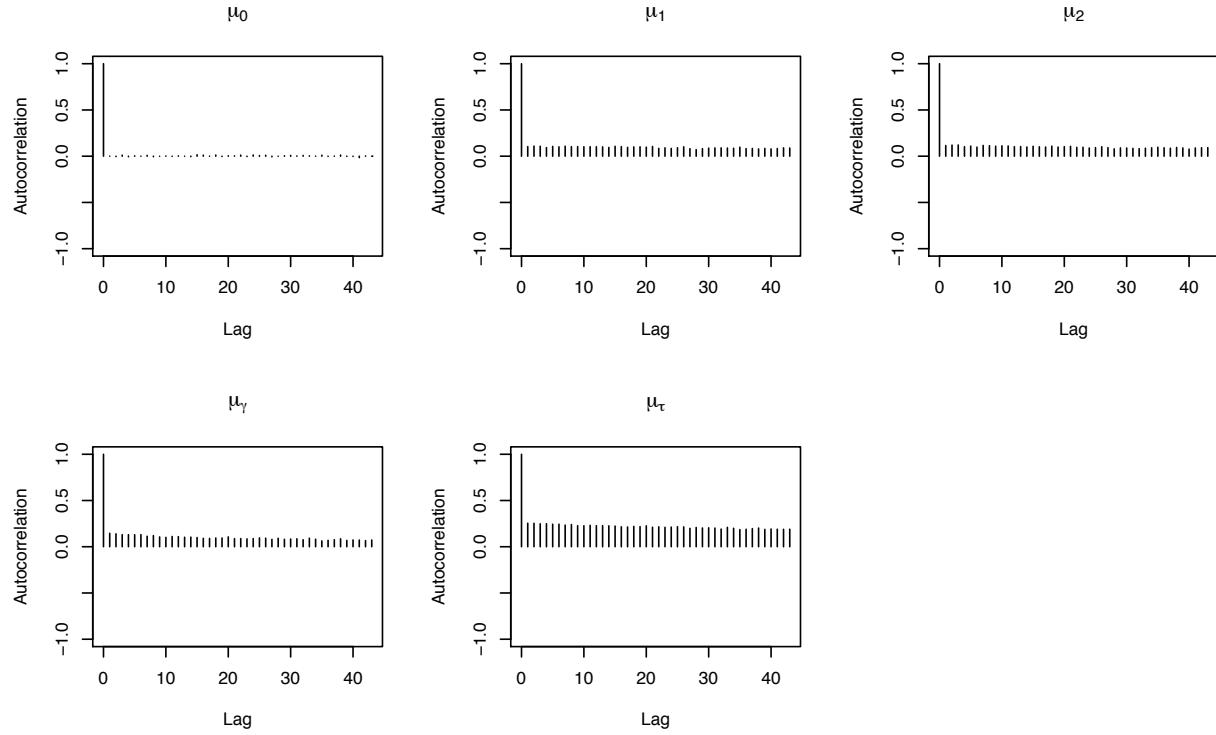
Figure 3.12: Autocorrelation Plots for the two Chains for Population Parameters μ_0, μ_1, μ_2, μ_γ and μ_τ

of trends in trace plot and low autocorrelation in the two chains jointly indicate good mixing of Markov chain. The density plot shows no sign of multi-modality. The plots as well as the Gelman-Rubin R statistic and Geweke $|Z|$ score (Table 3.6) indicate the stationarity of the chains for CFC-12 data.
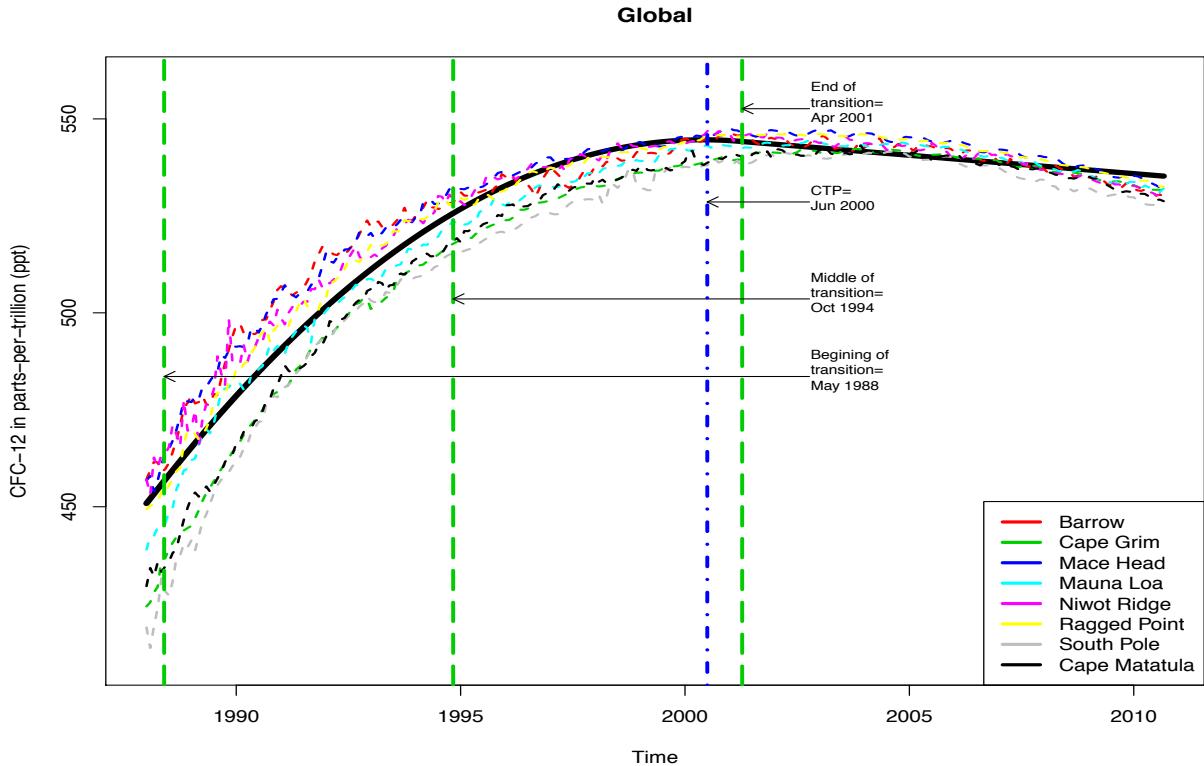
Figure 3.13: Observed data and corresponding population (Global) fit (solid black) for CFC-12

Figure 3.13 and 3.14 displayed the global and station-specific fits for CFC-12 data. Numerical fits are summarized in Table 3.7 for all stations whereas Table 3.8 reports the posterior characteristics of the standard deviations associated with Σ_β and Σ_α. In Table 3.7, some variations are observed in the estimates of the intercepts across all stations. The incoming and outgoing slopes appeared as statistically significant and agreed closely for all stations. Similar to CFC-11, large variation in the estimated intercepts and small variation in the estimated slopes are again supported by the estimates of Σ_β (Table 3.8). For CFC-12, the standard deviations for β_{0i}, β_{1i} and β_{2i} are estimated as 18.77, 0.24 and 0.22, respectively.
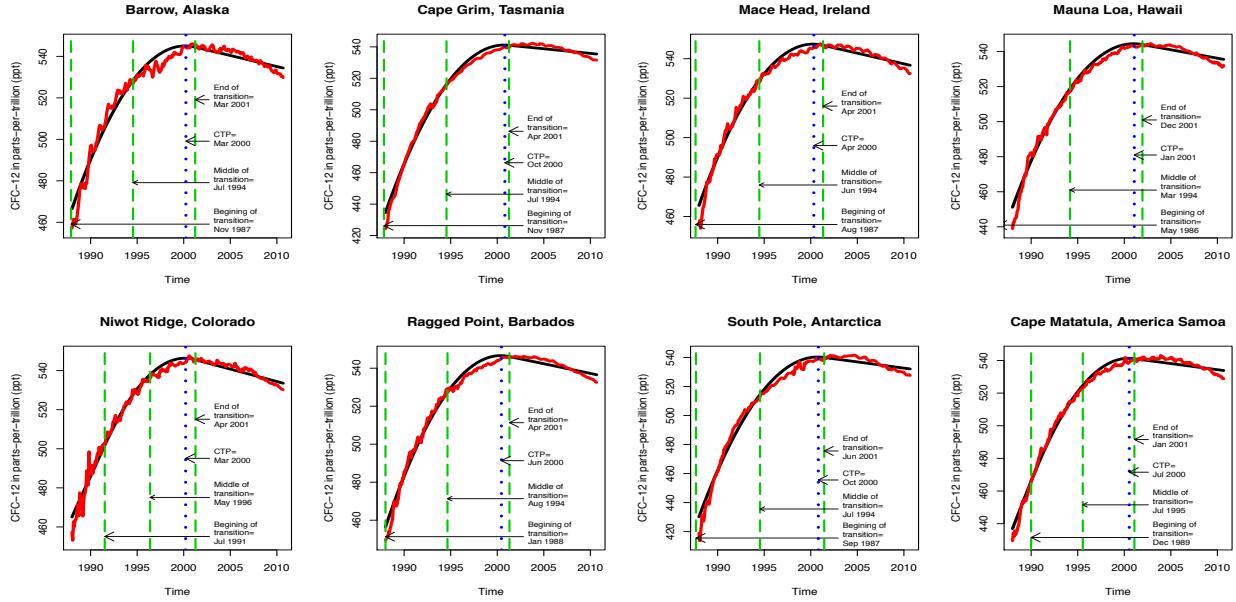
Figure 3.14: Observed data (red curve) and corresponding station-specific fitted (black) curve for CFC-12

Table 3.7: Posterior Summaries of the Station-specific Regression Parameters

Station	β_{0i} (95% Credible Interval)	β_{1i} (95% Credible Interval)	$\beta_{1i} + \beta_{2i}$ (95% Credible Interval)	Transition Period (Duration)	CTP (95% Credible Interval)
Barrow, Alaska $\hat{\sigma_1}$=2.17	465.42 (464.03, 466.87)	1.09 (0.98, 1.23)	−0.089 (−0.100, −0.77)	11/1987 to 03/2001 (160 months)	03/2000 (12/1999 to 06/2000)
Cape Grim, Tasmania $\hat{\sigma_2}$= 1.01	433.26 (432.33, 434.25)	1.42 (1.30, 1.55)	−0.049 (−0.057, −0.041)	11/1987 to 04/2001 (161 months)	10/2000 (08/2000 to 12/2000)
Mace Head, Ireland $\hat{\sigma_3}$= 1.10	464.57 (463.44, 466.07)	1.16 (1.06, 1.29)	−0.090 (−0.098, −0.082)	08/1987 to 04/2001 (165 months)	04/2000 (03/2000 to 06/2000)
Mauna Loa, Hawaii $\hat{\sigma_4}$= 1.43	451.37 (448.53, 456.24)	1.35 (1.17, 1.54)	−0.080 (−0.090, −0.070)	05/1986 to 12/2001 (187 months)	01/2001 (11/2000 to 03/2001)
Niwot Ridge, Colorado $\hat{\sigma_5}$= 1.75	464.23 (463.04, 465.41)	0.87 (0.83, 0.91)	−0.106 (−0.116, −0.096)	07/1991 to 04/2001 (117 months)	03/2000 (12/1999 to 05/2000)
Ragged Point, Barbados $\hat{\sigma_6}$= 1.20	455.16 (454.11, 456.28)	1.23 (1.11, 1.37)	−0.085 (−0.093, −0.077)	01/1988 to 04/2001 (159 months)	06/2000 (04/2000 to 08/2000)
South Pole, Antarctica $\hat{\sigma_7}$= 2.17	428.75 (427.07, 430.83)	1.49 (1.35, 1.66)	−0.071 (−0.085, −0.058)	09/1987 to 06/2001 (165 months)	10/2000 (08/2000 to 01/2001)
Cape Matatula, American Samoa $\hat{\sigma_8}$= 1.27	435.75 (434.76, 436.74)	1.21 (1.16, 1.26)	−0.063 (−0.072, −0.054)	12/1989 to 01/2001 (133 months)	07/2000 (05/2000 to 09/2000)

45

Table 3.8: Posterior Summaries of the Standard Deviations of the Random Regression Coefficients

Parameter	Posterior Mean	95% Credible Interval
$SD(\beta_{0i})$	18.77	(10.50, 29.75)
$SD(\beta_{1i})$	0.24	(0.12, 0.38)
$SD(\beta_{2i})$	0.22	(0.11, 0.36)
$SD(\gamma_i)$	11.63	(5.40, 21.58)
$SD(\tau_i)$	8.87	(4.36, 16.51)

3.2.4 Discussion and Conclusion

In summary, our analysis reveals the following points of interest:

- Strong spatial dependence among all stations suggests that both CFC-11 and CFC-12 have already been distributed globally.

- Both types of CFCs increased significantly before entering into a transition zone, though CFC-12 increased at a faster rate compared to the increase of CFC-11.

- On average, CFC-11 completed the transition between August, 1988 and September, 1995 whereas CFC-12 between May, 1988 and April, 2001.

- CFC-11 has been decreasing significantly after completing the transition, but the rate at which CFC-12 has been decreasing is not significant.

- Although the rates of increase and decrease are similar across stations for each type of CFC, there are considerable variations in the times to transition zone. Though the two types of CFCs may pose potentially differential health effects, the above findings also indicate a much more severe global concern for CFC-12 with respect to its concentration in the atmosphere.

3.3 Simulation Study

In this section, we present a simulation study to demonstrate the efficacy of our spatial-longitudinal bent-cable methodology. In particular, we supplement the motivation of our

methodology by illustrating the importance of properly taking into account the spatial component. To this end, we present a scenario, where, in reality, spatial dependency exists among the units. We then analyze the data using the true model, as well as models with misspecified spatial configuration.

Model parameter values in the simulations are chosen to approximately mimic the CFC-11 data. We take $m = 8$, $n = 273$ and $t_{ij} = j - 1$ for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$. We generate data assuming Model 4 with $\delta = 0.52$. We also take $\boldsymbol{\mu}_\beta = (244.28, 0.64, -0.79)'$, $\boldsymbol{\mu}_\alpha = (3.75, 3.91)'$,

$$
\Sigma_\beta = \begin{bmatrix} 204.66 & -2.46 & 2.35 \\ -2.46 & 0.06 & -0.04 \\ 2.35 & -0.04 & 0.05 \end{bmatrix}, \quad \Sigma_\alpha = \begin{bmatrix} 0.022 & 0.012 \\ 0.012 & 0.020 \end{bmatrix},
$$

and σ_i^2's are 1.36, 0.63, 6.08, 1.63, 1.54, 1.16, 1.89, and 1.39. Given the parameters, we generate $\boldsymbol{\beta}_i$'s and $\boldsymbol{\alpha}_i$'s from $N_3(\boldsymbol{\mu}_\beta, \Sigma_\beta)$ and $LN_2(\boldsymbol{\mu}_\alpha, \Sigma_\alpha)$, respectively. Then, y_{ij}'s are generated using Metropolis-Hastings algorithm from the following model:

$$
\pi(\mathbf{y}|\boldsymbol{\Theta}) =
$$

$$
(2\pi)^{-\frac{mn}{2}} \, |\, \mathbb{I}_m - \delta \mathbb{W} \,|^n \left\{ \prod_{i=1}^m (\sigma_i^{-2})^{\frac{n}{2}} \right\} \exp \left\{ -\frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n \left[\left(y_{ij} - \delta \sum_{k \neq i}^m w_{ik} y_{kj} \right) - \left(f_{ij} - \delta \sum_{k \neq i}^m w_{ik} f_{kj} \right) \right]^2 \right\}
$$

We analyze the data using Model 1 (no spatial dependency), Model 2 (misspecified spatial configuration), and Model 4 (true model) to evaluate the performance of our methodology and the effects of misspecified spatial configuration.

For each simulation, 500 data sets are generated, and $10,000$ MCMC iterations are used to approximate posterior distributions per set. DICs and posterior summaries for each parameter are then averaged over the 500 sets. In addition to the DICs, relative mean square error (RMSE) of the posterior summaries is used to evaluate the performance of the analytic procedures. RMSE takes into account both bias (measured by the relative bias, RB) and variability (gauged by the relative standard deviation, RSD), according to the formula RMSE $=$ $\text{RB}^2 + \text{RSD}^2$, where RB $=$ bias/true parameter value, RSD $=$ SD/true parameter value, and SD is the standard deviation of the posterior means/medians across the 500 simulated data sets. Note that the smaller the RMSE, the more accurate an estimator is.

Table 3.9: Absolute relative bias (RB), absolute relative standard deviation (RSD) and relative mean square error (RMSE) of the population regression coefficients and the spatial autocorrelation coefficient.

| | | Analysis assuming | | | | | | | | |
| | | Model 1 | | | Model 2 | | | Model 4 | | |
	True	|RB|	|RSD|	RMSE	|RB|	|RSD|	RMSE	|RB|	|RSD|	RMSE
μ_0	244.28	2.7×10^{-4}	0.001	4.9×10^{-7}	1.6×10^{-4}	0.001	4.7×10^{-7}	8.2×10^{-5}	0.001	5.4×10^{-7}
μ_1	0.64	0.064	0.031	0.00508	0.047	0.034	0.00336	0.031	0.039	0.00253
μ_2	-0.79	0.063	0.025	0.00461	0.051	0.027	0.00331	0.038	0.032	0.00245
μ_γ	3.75	0.018	0.006	0.00036	0.013	0.007	0.00023	0.011	0.008	0.00018
μ_τ	3.91	0.012	0.006	0.00018	0.010	0.007	0.00015	0.008	0.007	0.00011
δ	0.52	–	–	–	0.260	0.005	0.06760	9.4×10^{-4}	0.008	6.0×10^{-5}

Table 3.10: Absolute relative bias (RB), absolute relative standard deviation (RSD) and relative mean square error (RMSE) of σ_i^2's.

| | | Analysis assuming | | | | | | | | |
| | | Model 1 | | | Model 2 | | | Model 4 | | |
	True	|RB|	|RSD|	RMSE	|RB|	|RSD|	RMSE	|RB|	|RSD|	RMSE
σ_1^2	1.36	0.3281	0.0071	0.10770	0.1058	0.0060	0.01123	0.0063	0.0064	0.00008
σ_2^2	0.63	0.6909	0.0091	0.47747	0.0427	0.0081	0.00189	0.0063	0.0073	0.00009
σ_3^2	6.08	0.1234	0.0062	0.01528	0.0643	0.0064	0.00418	0.0031	0.0058	0.00004
σ_4^2	1.63	0.3711	0.0074	0.13777	0.1239	0.0092	0.01543	0.0092	0.0097	0.00018
σ_5^2	1.54	0.2809	0.0064	0.07896	0.0188	0.0059	0.00039	0.0003	0.0057	0.00003
σ_6^2	1.16	0.3465	0.0071	0.12014	0.2317	0.0067	0.05374	0.0013	0.0062	0.00004
σ_7^2	1.89	0.5129	0.0078	0.26313	0.2838	0.0074	0.08057	0.0054	0.0069	0.00008
σ_8^2	1.39	0.2114	0.0068	0.04473	0.0996	0.0056	0.00995	0.0080	0.0068	0.00011

Table 3.11: Absolute relative bias (RB), absolute relative standard deviation (RSD) and relative mean square error (RMSE) of the covariance matrices Σ_β and Σ_α.

| | | Analysis assuming | | | | | | | | |
| | | Model 1 | | | Model 2 | | | Model 4 | | |
	True	|RB|	|RSD|	RMSE	|RB|	|RSD|	RMSE	|RB|	|RSD|	RMSE
$(\Sigma_\beta)_{11}$	204.660	0.116	0.019	0.014	0.116	0.020	0.014	0.114	0.020	0.013
$(\Sigma_\beta)_{22}$	0.060	0.373	0.114	0.152	0.352	0.135	0.142	0.311	0.143	0.117
$(\Sigma_\beta)_{33}$	0.050	0.487	0.132	0.254	0.467	0.159	0.243	0.428	0.167	0.211
$(\Sigma_\beta)_{12}$	-2.460	0.099	0.086	0.017	0.095	0.099	0.019	0.050	0.115	0.016
$(\Sigma_\beta)_{13}$	2.350	0.099	0.089	0.018	0.095	0.103	0.020	0.049	0.121	0.017
$(\Sigma_\beta)_{23}$	-0.040	0.267	0.167	0.099	0.240	0.200	0.097	0.185	0.210	0.078
$(\Sigma_\alpha)_{11}$	0.022	0.621	0.359	0.515	0.378	0.342	0.260	0.255	0.306	0.159
$(\Sigma_\alpha)_{22}$	0.020	0.382	0.356	0.273	0.250	0.352	0.187	0.165	0.312	0.125
$(\Sigma_\alpha)_{12}$	0.012	0.647	0.277	0.495	0.165	0.289	0.111	0.252	0.312	0.161

Numerical results for the population regression coefficients and spatial autocorrelation parameter are summarized in Table 3.9. For μ_0, both $|RB|$ and RMSE are very close to zero regardless of the assumption about the spatial configuration. Smallest $|RB|$ and RMSE for each of the other parameters are observed when we analyze using the true model (i.e., Model 4), whereas the largest $|RB|$ and RMSE are observed if we completely ignore spatial dependency (i.e., Model 1). Similar results are obtained for σ_i^2's (Table 3.10), and Σ_β and Σ_α (Table 3.11), that is, the smallest $|RB|$ and RMSE occur when we analyze using the true model, and the largest $|RB|$ and RMSE if we ignore spatial dependency. The only exception occurs for $(\Sigma_\alpha)_{12}$ (Table 3.11), for which the smallest $|RB|$ and RMSE are observed for Model 2. The average DICs are 7893, 7720, and 7379 for Model 1, Model 2 and Model 4, respectively. This indicates that the best fit is achieved by Model 4 (smallest DIC), and a comparatively poor fit if we completely ignore the spatial dependency (largest DIC).

The above simulation results demonstrate the importance of modelling the spatial component when, in reality, spatial dependency exists among the units. Moreover, smaller bias and highly accurate (smaller RMSE) estimates indicate that our proposed methodology performs well to analyze data that resemble those from the CFC study.

Chapter 4

Concluding Remarks and Future Work

Under a spatially dependent process, each unit is typically affected by those of the neighbouring units. Therefore, the assumption of independence across the neighbouring units might be unrealistic, especially when they represent geographical locations. Moreover, addressing the spatial effects provide important insights about such a process. In this thesis, we propose an extension of the longitudinal bent-cable model (Khan et al., 2009) by taking into account spatial effects. We have tailored our work especially for the scientific context of the CFC data. CFCs persist long enough in the atmosphere due to their extended lifetimes: 45 and 100 years for CFC-11 and CFC-12, respectively. As a consequence, they are believed to have spread across the world. Therefore, CFCs monitored from one station may depend on those from another station, giving rise to a presumed spatially dependent longitudinal process. Our methodology provides understanding not only of the spatial distribution, but also of the global threat that CFCs may pose to all living organisms. It also reveals useful information regarding the atmospheric CFC decline throughout the globe.

Since the Montreal Protocol came into effect, a global decrease in the CFCs is monitored and confirmed by our analysis. Note that the Montreal Protocol contains an extended CFC phase-out schedules: 1996 for developed countries and 2010 for developing countries. Thus, many countries at various geographical locations continued to contribute CFCs to the atmosphere during the 273 months in our study over the period January, 1988 to September, 2010. In our analysis, this fact is reflected with a slow decrease in CFC concentrations from the atmosphere. In fact, our analysis does not reveal a significant decrease for CFC-12. This makes sense due to the extended lifetime for CFC-12 compared to CFC-11. Nevertheless, the Montreal Protocol can be regarded as a successful international agreement to reduce the use of the CFCs, in particular, thus far, for CFC-11.

There are other modelling approaches that can be used to characterize non-linear trend change over time. For example, a piecewise linear model with three knots could be fit, where the transition period for the bent cable model would instead be approximated by two straight lines with one internal knot. In case a piecewise linear model with three knots would not fit the data well, another model could be fit with, say, five knots (three inside the transition, instead of one). This process may lead to investigating many models, and, due to the smooth transition phase, at least for the CFC data, may lead to too many parameters, which might diminish interpretability as compared to the bent-cable model. Another modelling approach for such data is the functional mixed effects model (e.g., (Baladandayuthapani et al., 2008)). So, further investigation is necessary to compare our methodology with other available statistical techniques to model this type of longitudinal change-point data. Nevertheless, our proposed methodology is appealing for its many attractive features (great interpretability, flexibility and parsimony), and, it fits the CFC data well.

Although we have tailored our work for the CFC data, our modelling framework may be applicable to a wide variety of other situations across the range of the econometrics, social, health and medical sciences. We summarize some cautionary remarks and potential future research topics in Sections 4.1 and 4.2, respectively.

4.1 Cautionary Remarks

Our methodology could serve as a powerful statistical tool in analysing spatial data, that exhibits change of direction, for its flexibility and appealing features. Such data may arise in many areas such as biology, medical and environmental applications. However, careful attention is required for the following reasons:

1 The bent-cable methodology is defined for data exhibiting only one transition period over time. It is possible to apply the methodology for multiple transition periods but the data should be partitioned into several portions so that each part has only one transition period. Moreover, the process of combining information from individual portions for drawing inference is not clear.

2 Poor starting values may result in slow-mixing for a Markov chain. Hence, starting values should be chosen carefully to avoid lengthy runs.

4.2 Future Work

There is scope to extend spatial-longitudinal bent-cable model to different directions. Some of such possible directions are pointed below:

1 One possibility is to extend our framework to take into consideration of an interaction between temporal and spatial effects. With our current model, this could be done with a spatial correlation structure that changes as a function of time.

2 The covariance structure can be modelled as a continuous function of distance, that separates the stations from each other, instead of defining the spatial weight matrix.

3 The longitudinal bent-cable model can be extended through decomposing the mean structure to take into account the spatial effects.

As mentioned earlier, our article based on the application of spatial-longitudinal bent-cable model has been accepted in the *International Journal of Statistics and Probability* (Khan et al., 2012) published by Canadian Center of Science and Education.

References

AGAGE. (2012). Atmospheric Lifetime Experiment/Global Atmospheric Gases Experiment /Advanced Global Atmospheric Gases Experiment. Available from *http : //agage.eas.gatech.edu/data_archive*

Andrieu, C., Freitas, N. D., Doucet, A., & Jordan, M. I. (2003). An introduction to mcmc for machine learning. *Machine Learning*, *50*, 5-43.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic.

Anselin, L., Gallo, J. L., & Jayet, H. (2008). Spatial panel econometrics. In M. Laszlo & S. Patrick (Eds.), *The Econometrics of Panel Data* (3rd ed., p. 625-660). Springer.

Baladandayuthapani, V., Mallick, B. K., Hong, M. Y., Lupton, J. R., Turner, N. D., & Carroll, R. J. (2008). Bayesian hierarchical spatially correlated functional data analysis with application to colon carcinogenesis. *Biometrics*, *64*(1), 64-73.

Banerjee, S. (2005, June). On geodetic distance computations in spatial modeling. *Biometrics*, *61*, 617-625.

Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2003). *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton, Florida: Chapman and Hall.

Carlin, B. P. (1995). Hierarchical longitudinal modelling. In W. R. Gilks, S. Richardson, & D. J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice* (pp. 303–319). London: Chapman & Hall.

Chib, S., & Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *The American Statistician*, *49*, 327-335.

Chiu, G., Lockhart, R., & Routledge, R. (2006, June). Bent-cable regression theory and applications. *Journal of the American Statistical Association*, *101*(474), 542–553.

Chiu, G., & Lockhart, R. A. (2010, September). Bent-cable regression with autoregressive noise. *The Canadian Journal of Statistics*, *38*(3), 386-407.

Congdon, P. D. (2003). *Applied Bayesian Modelling* (First ed.). John Wiley and Sons.

Congdon, P. D. (2010). *Applied Bayesian Hierarchical Methods*. Boca Raton, Florida: Chapman & Hall.

Cunnold, D. M., Weiss, R. F., Prinn, R. G., Hartley, D. E., Simmonds, P. G., Fraser, P. J., et al. (1997). GAGE/AGAGE measurements indicating reduction in global emission of CCl_3F and CCl_2F_2 in 1992–1994. *Journal of Geophysical Research*, *102*, 1259-1269.

Davidian, M., & Giltinan, D. M. (1995). *Nonlinear Models for Repeated Measurement Data*. New York: Chapman & Hall.

Elhorst, J. P. (2010). Spatial panel data models. In M. M. Fischer & A. Getis (Eds.), *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications* (First ed., pp. 377–407). Berlin Heidelberg: Springer-Verlag.

Gelfand, A. E., Latimer, A., Wu, S., & Silander, J. A. (2006, May). Building statistical models

to analyze species distribution. In J. S. Clark & A. E. Gelfand (Eds.), *Hierarchical Modelling for the Environmental Sciences: Statistical Methods and Applications* (First ed., p. 77-97). New York, USA: Oxford University Press.

Gelman, A. (1995). Inference and monitoring convergence. In W. Gilks, S. Richardson, & D. J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice* (p. 131-143). London: Chapman and Hall.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, *1*, 1–19.

Gelman, A., & Rubin, D. B. (1992a). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457-472.

Gelman, A., & Rubin, D. B. (1992b). A single series from the gibbs sampler provides a false sense of security. In J. M. Bernardo, J. Berger, A. P. Dawid, & A. F. M. Smith (Eds.), *Bayesian Statistics* (Vol. 4, p. 625-631). Oxford: Oxford University Press.

Geman, S., & Geman, D. (1984). Stochastic relaxation, gibbs distributions and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *6*, 721-741.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In J. M. Bernardo, J. Berger, A. P. Dawid, & A. F. M. Smith (Eds.), *Bayesian Statistics* (Vol. 4, p. 169-193). Oxford, UK: Clarendon Press.

Geyer, C. J. (1992). Practical markov chain monte carlo. *Statistical Science*, *7*(4), 473-483.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1995). Introducing markov chain monte carlo. In W. Gilks, S. Richardson, & D. J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice* (p. 1-19). London: Chapman and Hall.

Givens, G. H., & Hoeting, J. A. (2005). *Computational Statistics.* New Jersey: John Wiley and Sons.

Griffith, D. A., & Layne, L. J. (1999). *A Casebook for Spatial Statistical Data Analysis.* USA: Oxford University Press.

Hastings, W. K. (1970, April). Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, *57*(1), 97-109.

Hoef, J. M. V., Cressie, N. A. C., Fisher, R. N., & Case, T. J. (2001, June). Uncertainty and spatial linear models for ecological data. In C. T. Hunsaker, M. F. Goodchild, M. A. Friedl, & T. J. Case (Eds.), *Spatial Uncertainty in Ecology: Implications for Remote Sensing and GIS Applications* (First ed., p. 214-237). New York: Springer-Verlag.

Khan, S. A. (2010). *Flexible bent-cable models for mixture longitudinal data.* Unpublished doctoral dissertation, University of Waterloo, Waterloo, Ontario, Canada.

Khan, S. A., Chiu, G., & Dubin, J. A. (2009). Atmospheric concentration of chlorofluorocarbons: Addressing the global concern with the longitudinal bent-cable model. *CHANCE*, *22*(3), 8–17.

Khan, S. A., Rana, M., Li, L., & Dubin, J. A. (2012, November). A statistical investigation to monitor and understand atmospheric cfc decline with the spatial-longitudinal bent-cable model. *International Journal of Statistics and Probability*, *1*(2), 56-68.

Malina, M., & Rowland, F. S. (1974). Stratospheric sink for chlorofluoromethanes: chlorine catalyzed destruction of ozone. *Nature*, *249*, 810-812.

McKnight, P. E., Figueredo, A. J., & Sidani, S. (2007). *Missing Data: A Gentle Introduction.*

New York: Guilford Press.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953, June). Equations of state calculations by fast computing machine. *Journal of Chemical Physics*, *21*(6), 1087-1091.

Moulijn, J., Makkee, M., Wiersma, A., & Sandt, E. van de. (2000, June). Selective hydrogenolysis of CCl_2F_2 into CH_2F_2 over palladium on activated carbon: Kinetic mechanism and process design. *Catalysis Today*, *59*(3-4), 221 - 230.

NOAA/ESRL Halocarbons Group. (2012). Global Monitoring Division, National Oceanic and Atmospheric Administration. Retrieved July 18, 2012, from `http://www.esrl.noaa.gov/gmd/hats/combined/CFC11.html;` `http://www.esrl.noaa.gov/gmd/hats/combined/CFC12.html`

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News*, *6*(1), 7–11.

Prinn, R. G., Huang, J., Weiss, R. F., Cunnold, D. M., Fraser, P. J., Simmonds, P. G., et al. (2001). Evidence for significant variations of atmospheric hydroxyl radicals in the last two decades. *Science*, *292*, 1882–1888.

Prinn, R. G., Weiss, R. F., Fraser, P. J., Simmonds, P. G., Cunnold, D. M., Alyea, F. N., et al. (2000, July). A history of chemically and radiatively important gases in air deduced from ale/gage/agage. *Journal of Geophysical Research*, *105*(D14), 17751–17792.

R Core Team. (2012). R: A Language and Environment for Statistical Computing [Computer software manual]. R Foundation for Statistical Computing, Vienna, Austria. Available from `http://www.R-project.org/`

Roberts, G. O. (1995). Markov chain concepts related to sampling algorithms. In W. Gilks, S. Richardson, & D. J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice* (p. 45-57). London: Chapman and Hall.

Smith, A. F. M., & Roberts, G. O. (1993). Bayesian computation via gibbs sampler and related markov chain monte carlo methods. *Journal of the Royal Statistical Society, Series B*, *55*, 3-23.

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

Struijs, J., Dijk, A. van, Slaper, H., Wijnen, H. J. van, Velders, G. J. M., Chaplin, G., et al. (2010). Spatial- and time-explicit human damage modeling of ozone depleting substances in life cycle impact assessment. *Environmental Science & Technology*, *44*(1), 204-209.

Tierney, L. (1994, December). Markov chains for exploring posterior distributions. *The Annals of Statistics*, *22*(4), 1701-1728.

Velders, G. J. M., Andersen, S. O., Daniel, J. S., Fahey, D. W., & McFarland, M. (2007, March). The importance of the montreal protocol in protecting climate. *Proceedings of the National Academy of Science of the United States of America*, *104*(12), 4814-4819.

Wakefield, J. C., Smith, A. F. M., Racine-Poon, A., & Gelfand, A. E. (1994). Bayesian analysis of linear and non-linear population models by using the gibbs sampler. *Applied Statistics*, *43*, 201-221.

Wikle, C. K. (2003, August). Hierarchical models in environmental science. *International Statistical Review*, *71*(2), 181-199.

WMO, W. M. O. (2010). *Scientific assessment of ozone depletion: 2010* (Tech. Rep.). Geneva, Switzerland: Global Ozone Research and Monitoring Project-Report No. 52.

Zhang, F., Zhou, L., Yao, B., Vollmer, M. K., Greally, B. R., Simmonds, P. G., et al. (2010). Analysis of 3-year observations of cfc-11, cfc-12 and cfc-113 from a semi-rural site in china. *Atmospheric Environment*, *44* (35), 4454 - 4462.

APPENDICES

In this chapter, we present the detailed derivation of mathematical expressions related to our spatial-longitudinal bent-cable model. We derive the joint distribution in detail in section A.1. We work out the first stage of hierarchy in Section A.2 while the full conditional distributions are derived in Section A.3.

A.1 Joint Distribution

Let $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{in})$, $\mathbf{y} = (\mathbf{y}_1', \mathbf{y}_2', \ldots, \mathbf{y}_m')'$ and $\boldsymbol{\Theta}$ denotes all the model parameters collectively i.e. $\boldsymbol{\Theta} = \left(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\mu}_\beta, \boldsymbol{\mu}_\alpha, \boldsymbol{\Sigma}_\beta^{-1}, \boldsymbol{\Sigma}_\alpha^{-1}, \boldsymbol{\sigma}^{-2}, \delta\right)$. Then the joint distribution of model parameters and data can be expressed as:

$$
\begin{aligned}
\pi\left(\boldsymbol{\Theta}, \mathbf{y}\right) = {} & \pi\left(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\mu}_\beta, \boldsymbol{\mu}_\alpha, \boldsymbol{\Sigma}_\beta^{-1}, \boldsymbol{\Sigma}_\alpha^{-1}, \boldsymbol{\sigma}^{-2}, \delta\right) \\
& \times \prod_{i=1}^{m} \pi\left(\boldsymbol{\beta}_i|\boldsymbol{\alpha}, \boldsymbol{\mu}_\beta, \boldsymbol{\mu}_\alpha, \boldsymbol{\Sigma}_\beta^{-1}, \boldsymbol{\Sigma}_\alpha^{-1}, \boldsymbol{\sigma}^{-2}, \delta\right) \\
& \times \prod_{i=1}^{m} \pi\left(\boldsymbol{\alpha}_i|\boldsymbol{\mu}_\beta, \boldsymbol{\mu}_\alpha, \boldsymbol{\Sigma}_\beta^{-1}, \boldsymbol{\Sigma}_\alpha^{-1}, \boldsymbol{\sigma}^{-2}, \delta\right) \\
& \times \left\{\pi\left(\boldsymbol{\mu}_\beta\right)\ \pi\left(\boldsymbol{\mu}_\alpha\right)\ \pi\left(\boldsymbol{\Sigma}_\beta^{-1}\right)\ \pi\left(\boldsymbol{\Sigma}_\alpha^{-1}\right)\ \pi\left(\boldsymbol{\sigma}^{-2}\right)\ \pi\left(\delta\right)\right\} \\
= {} & \prod_{i=1}^{m} \mathcal{N}_n\left(\mathbf{y}_i|\boldsymbol{\mu}_i, \sigma_i^2 \mathbb{I}_i\right) \times \prod_{i=1}^{m} \mathcal{N}_3\left(\boldsymbol{\beta}_i|\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta\right) \times \prod_{i=1}^{m} \mathcal{LN}_2\left(\boldsymbol{\alpha}_i|\boldsymbol{\mu}_\alpha, \boldsymbol{\Sigma}_\alpha\right) \\
& \times \mathcal{N}_3\left(\boldsymbol{\mu}_\beta|\mathbf{h}_\beta, \mathbb{H}_\beta\right) \times \mathcal{N}_2\left(\boldsymbol{\mu}_\alpha|\mathbf{h}_\alpha, \mathbb{H}_\alpha\right) \times \mathcal{W}\left(\boldsymbol{\Sigma}_\beta^{-1}|\nu_\beta, (\nu_\beta \mathbb{A})^{-1}\right) \\
& \times \mathcal{W}\left(\boldsymbol{\Sigma}_\alpha^{-1}|\nu_\alpha, (\nu_\alpha \mathbb{A})^{-1}\right) \times \prod_{i=1}^{m} \mathcal{G}\left(\sigma_i^{-2}|\frac{\mathbf{a}_0}{2}, \frac{\mathbf{a}_1}{2}\right) \times \mathcal{U}\left(\delta|\mathbf{b}_0, \mathbf{b}_1\right) \qquad \text{(A.1)}
\end{aligned}
$$

where $\mathcal{N}_p$, $\mathcal{LN}_2$, $\mathcal{W}$, $\mathcal{G}$, and $\mathcal{U}$ denote $p-$variate normal distribution, bivariate lognormal distribution, Wishart distribution, Gamma distribution, and Uniform distribution respec-

tively. We derive the full conditional distribution for each parameter based on equation A.1 to implement the Metropolis within Gibbs algorithm for our spatial-longitudinal bent-cable model.

A.2 First Stage of Hierarchy

Recalling from Section 2.2, we denote the CFC measurement observed at i^{th} station, $i = 1, 2, \ldots, m$ at j^{th} time period, $j = 1, 2, \ldots, n$ by y_{ij}. From equations (2.1), (2.4) and (2.5), we have:

$$y_{ij} = f_{ij} + e_{ij}$$

$$e_{ij} = \delta \sum_{k=1}^{m} w_{ik} e_{kj} + \epsilon_{ij}$$

$$\left[\epsilon_{ij} | \sigma_i^2 \right] \sim \mathcal{N} \left(0, \sigma_i^2 \right)$$

Now,

$$
\begin{aligned}
y_{ij} &= f_{ij} + e_{ij} \\
&= f_{ij} + \delta \sum_{k=1}^{m} w_{ik} e_{kj} + \epsilon_{ij} \\
&= f_{ij} + \delta \sum_{k=1}^{m} w_{ik} \left(y_{kj} - f_{kj} \right) + \epsilon_{ij} \\
&= f_{ij} + \delta \sum_{k=1}^{m} w_{ik} y_{kj} - \delta \sum_{k=1}^{m} w_{ik} f_{kj} + \epsilon_{ij} \\
&= \delta \sum_{k=1}^{m} w_{ik} y_{kj} + f_{ij} - \delta \sum_{k=1}^{m} w_{ik} f_{kj} + \epsilon_{ij} \\
&= \delta \sum_{k \neq i=1}^{m} w_{ik} y_{kj} + f_{ij} - \delta \sum_{k \neq i=1}^{m} w_{ik} f_{kj} + \epsilon_{ij} \qquad [since\ w_{ii} = 0] \\
&= \delta \sum_{k \neq i=1}^{m} w_{ik} y_{kj} + \beta_{0i} + \beta_{1i} t_{ij} + \beta_{2i} q_{ij} - \delta \sum_{k \neq i=1}^{m} w_{ik} f_{kj} + \epsilon_{ij}
\end{aligned}
$$

$$= \delta \sum_{k \neq i=1}^{m} w_{ik} y_{kj} + \beta_{0i} + \beta_{1i} t_{ij} + \beta_{2i} q_{ij} - \delta \left(w_{i1} f_{1j} + w_{i2} f_{2j} + \ldots\ldots + w_{im} f_{mj} \right) + \epsilon_{ij}$$

$$= \delta \sum_{k \neq i=1}^{m} w_{ik} y_{kj} + \beta_{0i} + \beta_{1i} t_{ij} + \beta_{2i} q_{ij} - \delta w_{i1} \left(\beta_{01} + \beta_{11} t_{1j} + \beta_{21} q_{1j} \right)$$

$$+ \delta w_{i2} \left(\beta_{02} + \beta_{12} t_{2j} + \beta_{22} q_{2j} \right) + \ldots\ldots + \delta w_{im} \left(\beta_{0m} + \beta_{1m} t_{mj} + \beta_{2m} q_{mj} \right) + \epsilon_{ij}$$

$$= \mu_{ij} + \epsilon_{ij}$$

Since $[\epsilon_{ij} | \sigma_i^2] \sim \mathcal{N}\left(0, \sigma_i^2\right)$ for all j, we get the first stage of hierarchy as:

$$\left[\mathbf{y}_i | \boldsymbol{\theta}_i, \delta, \sigma_i^2 \right] \sim \mathcal{N}_n \left(\boldsymbol{\mu}_i, \sigma_i^2 \mathbb{I}_i \right)$$

where $\boldsymbol{\mu}_i \left(\boldsymbol{\theta}_i, \delta \right) \equiv \boldsymbol{\mu}_i = \left(\mu_{i1}, \mu_{i2}, \ldots, \mu_{in} \right)'$, $\mu_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + \beta_{2i} q_{ij} + \delta \sum_{k \neq i=1}^{m} w_{ik} y_{kj} - \delta \sum_{k \neq i=1}^{m} w_{ik} f_{kj}$, $\mathbb{I}_i$ is an identity matrix of order n and $\mathcal{N}_n$ denotes a $n-$variate normal distribution.

A.3 Full Conditionals

In general, Metropolis sampling is based on the full conditional distribution of a particular parameter when it is being updated. These full conditional distributions also play central role in Gibbs sampling. The full conditionals are usually obtained from the joint distribution and in many cases reduce to standard density from which direct sampling is straight forward. Full conditional distributions are extracted by abstracting out the elements involving parameter of interest from the joint density and treating other components as constants (referred in Congdon (2010)). For example, we derive the full conditional for $\boldsymbol{\beta}_i$ by picking out the terms from joint density (equation A.1) which involves $\boldsymbol{\beta}_i$. It is to be noted that any terms,

not depending on $\boldsymbol{\beta}_i$, is taken as proportionality constant in the full conditional.

An appealing feature of bent-cable function is that it is partially linear (Khan, 2010). This fact leads to closed-form full conditional for some parameters while the conditional for others can be expressed up to a proportionality constant. As our model is an extension of longitudinal bent-cable model (Khan et al., 2009), conditional distributions for some hyper-parameters are same for both models. We reproduce them here in our context for completeness of this thesis. We define the following notations to extract the full conditional distributions for parameters involved in our spatial-longitudinal bent-cable model.

$$z_{ij} = y_{ij} - \delta \sum_{k=1}^{m} w_{ik}(y_{kj} - f_{kj}) \text{ and } \mathbf{z}_i = (z_{i1}, z_{i2}, \ldots\ldots\ldots, z_{in})'$$

$$\mathbb{X}_i = \begin{pmatrix} 1 & t_{i,1} & q_{i,1} \\ 1 & t_{i,2} & q_{i,2} \\ \vdots & \vdots & \vdots \\ 1 & t_{i,n} & q_{i,n} \end{pmatrix}$$

$$\mathbb{M}_i^{-1} = \sigma_i^{-2} \mathbb{X}_i' \mathbb{X}_i + \Sigma_\beta^{-1}, \ \mathbb{U}_\beta^{-1} = m\Sigma_\beta^{-1} + \mathbb{H}_\beta^{-1} \text{ and } \mathbb{U}_\alpha^{-1} = m\Sigma_\alpha^{-1} + \mathbb{H}_\alpha^{-1}$$

$$\boldsymbol{\xi}_i = \log(\boldsymbol{\alpha}_i) = (\log \gamma_i, \log \tau_i)'$$

$$\tilde{\boldsymbol{\beta}} = \sum_{i=1}^{m} \boldsymbol{\beta}_i \text{ and } \tilde{\boldsymbol{\xi}} = \sum_{i=1}^{m} \boldsymbol{\xi}_i$$

A.3.1 Full Conditional for $\boldsymbol{\beta}_i$

Abstracting out the terms involving $\boldsymbol{\beta}_i$ from the joint density (equation A.1), we have

$$\pi\left(\boldsymbol{\beta}_i|.\right) \propto \exp\left\{ -\frac{1}{2\sigma_i^2}\left(\mathbf{y}_i - \boldsymbol{\mu}_i\right)'\left(\mathbf{y}_i - \boldsymbol{\mu}_i\right) - \frac{1}{2}\left(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta\right)'\Sigma_\beta^{-1}\left(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta\right) \right\}$$

From Section A.1, we can write

$$y_{ij} - \mu_{ij} = y_{ij} - \delta \sum_{k=1}^{m} w_{ik} \left(y_{kj} - f_{kj} \right) - \beta_{0i} - \beta_{1i} t_{ij} - \beta_{2i} q_{ij}$$

$$= z_{ij} - \beta_{0i} - \beta_{1i} t_{ij} - \beta_{2i} q_{ij} \quad for \; j = 1, 2, \ldots\ldots, n$$

In vector-matrix notation, we can write $\mathbf{y}_i - \boldsymbol{\mu}_i = \mathbf{z}_i - \mathbb{X}_i \boldsymbol{\beta}_i$. Using the result, we have

$$\pi \left(\boldsymbol{\beta}_i | . \right) \propto \exp \left\{ -\frac{1}{2} \left[\sigma_i^{-2} \left(\mathbf{z}_i - \mathbb{X}_i \boldsymbol{\beta}_i \right)' \left(\mathbf{z}_i - \mathbb{X}_i \boldsymbol{\beta}_i \right) + \left(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta \right)' \Sigma_\beta^{-1} \left(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta \right) \right] \right\}$$

$$= \exp \left\{ -\frac{1}{2} \left[\sigma_i^{-2} \mathbf{z}_i' \mathbf{z}_i - \sigma_i^{-2} \mathbf{z}_i' \mathbb{X}_i \boldsymbol{\beta}_i - \sigma_i^{-2} \boldsymbol{\beta}_i' \mathbb{X}_i' \mathbf{z}_i + \sigma_i^{-2} \boldsymbol{\beta}_i' \mathbb{X}_i' \mathbb{X}_i \boldsymbol{\beta}_i \right. \right.$$

$$\left. \left. + \; \boldsymbol{\beta}_i' \Sigma_\beta^{-1} \boldsymbol{\beta}_i - \boldsymbol{\beta}_i' \Sigma_\beta^{-1} \boldsymbol{\mu}_\beta - \boldsymbol{\mu}_\beta' \Sigma_\beta^{-1} \boldsymbol{\beta}_i + \boldsymbol{\mu}_\beta' \Sigma_\beta^{-1} \boldsymbol{\mu}_\beta \right] \right\}$$

$$\propto \exp \left\{ -\frac{1}{2} \left[\sigma_i^{-2} \mathbf{z}_i' \mathbb{X}_i \boldsymbol{\beta}_i - \sigma_i^{-2} \boldsymbol{\beta}_i' \mathbb{X}_i' \mathbf{z}_i + \sigma_i^{-2} \boldsymbol{\beta}_i' \mathbb{X}_i' \mathbb{X}_i \boldsymbol{\beta}_i + \boldsymbol{\beta}_i' \Sigma_\beta^{-1} \boldsymbol{\beta}_i \right. \right.$$

$$\left. \left. - \boldsymbol{\beta}_i' \Sigma_\beta^{-1} \boldsymbol{\mu}_\beta - \boldsymbol{\mu}_\beta' \Sigma_\beta^{-1} \boldsymbol{\beta}_i \right] \right\}$$

[proportionality follows because $\sigma_i^{-2} \mathbf{z}_i' \mathbf{z}_i$ and $\boldsymbol{\mu}_\beta' \Sigma_\beta^{-1} \boldsymbol{\mu}_\beta$ are independent of $\boldsymbol{\beta}_i$]

$$= \exp \left\{ -\frac{1}{2} \left[-2\sigma_i^{-2} \boldsymbol{\beta}_i' \mathbb{X}_i' \mathbf{z}_i + \sigma_i^{-2} \boldsymbol{\beta}_i' \mathbb{X}_i' \mathbb{X}_i \boldsymbol{\beta}_i + \boldsymbol{\beta}_i' \Sigma_\beta^{-1} \boldsymbol{\beta}_i - 2\boldsymbol{\beta}_i' \Sigma_\beta^{-1} \boldsymbol{\mu}_\beta \right] \right\}$$

[$\sigma_i^{-2} \mathbf{z}_i' \mathbb{X}_i \boldsymbol{\beta}_i$ and $\boldsymbol{\mu}_\beta' \Sigma_\beta^{-1} \boldsymbol{\beta}_i$ are scalers]

$$= \exp \left\{ -\frac{1}{2} \left[-2\boldsymbol{\beta}_i' \left(\sigma_i^{-2} \mathbb{X}_i' \mathbf{z}_i + \Sigma_\beta^{-1} \boldsymbol{\mu}_\beta \right) + \boldsymbol{\beta}_i' \left(\sigma_i^{-2} \mathbb{X}_i' \mathbb{X}_i + \Sigma_\beta^{-1} \right) \boldsymbol{\beta}_i \right] \right\}$$

$$= \exp \left\{ -\frac{1}{2} \left[-2\boldsymbol{\beta}_i' \left(\sigma_i^{-2} \mathbb{X}_i' \mathbf{z}_i + \Sigma_\beta^{-1} \boldsymbol{\mu}_\beta \right) + \boldsymbol{\beta}_i' \mathbb{M}_i^{-1} \boldsymbol{\beta}_i \right] \right\}$$

$$\propto \exp \left\{ -\frac{1}{2} \left[-\boldsymbol{\beta}_i' \left(\sigma_i^{-2} \mathbb{X}_i' \mathbf{z}_i + \Sigma_\beta^{-1} \boldsymbol{\mu}_\beta \right) - \left(\sigma_i^{-2} \mathbb{X}_i' \mathbf{z}_i + \Sigma_\beta^{-1} \boldsymbol{\mu}_\beta \right)' \boldsymbol{\beta}_i + \boldsymbol{\beta}_i' \mathbb{M}_i^{-1} \boldsymbol{\beta}_i \right. \right.$$

$$\left. \left. + \left(\sigma_i^{-2} \mathbb{X}_i' \mathbf{z}_i + \Sigma_\beta^{-1} \boldsymbol{\mu}_\beta \right)' \mathbb{M}_i \left(\sigma_i^{-2} \mathbb{X}_i' \mathbf{z}_i + \Sigma_\beta^{-1} \boldsymbol{\mu}_\beta \right) \right] \right\}$$

[$\boldsymbol{\beta}_i' (\sigma_i^{-2} \mathbb{X}_i' \mathbf{z}_i + \Sigma_\beta^{-1} \boldsymbol{\mu}_\beta)$ is a scaler and so is $\boldsymbol{\beta}_i' (\sigma_i^{-2} \mathbb{X}_i' \mathbf{z}_i + \Sigma_\beta^{-1} \boldsymbol{\mu}_\beta) = (\sigma_i^{-2} \mathbb{X}_i' \mathbf{z}_i + \Sigma_\beta^{-1} \boldsymbol{\mu}_\beta)' \boldsymbol{\beta}_i$ and

proportionality follows as $(\sigma_i^{-2} \mathbb{X}_i' \mathbf{z}_i + \Sigma_\beta^{-1} \boldsymbol{\mu}_\beta)' \mathbb{M}_i (\sigma_i^{-2} \mathbb{X}_i' \mathbf{z}_i + \Sigma_\beta^{-1} \boldsymbol{\mu}_\beta)$ does not depend on $\boldsymbol{\beta}_i$]

$$= \exp \left\{ -\frac{1}{2} \left[\boldsymbol{\beta}_i - \mathbb{M}_i \left(\sigma_i^{-2} \mathbb{X}_i' \mathbf{z}_i + \Sigma_\beta^{-1} \boldsymbol{\mu}_\beta \right) \right]' \mathbb{M}_i^{-1} \left[\boldsymbol{\beta}_i - \mathbb{M}_i \left(\sigma_i^{-2} \mathbb{X}_i' \mathbf{z}_i + \Sigma_\beta^{-1} \boldsymbol{\mu}_\beta \right) \right] \right\}$$

which is proportional to a density function of trivariate normal distribution with mean vector $\mathbb{M}_i \left(\sigma_i^{-2} \mathbb{X}_i' \mathbf{z}_i + \Sigma_\beta^{-1} \boldsymbol{\mu}_\beta \right)$ and covariance matrix $\mathbb{M}_i$. Thus,

$$[\boldsymbol{\beta}_i|.] \sim \mathcal{N}_3 \left(\mathbb{M}_i \left(\sigma_i^{-2} \mathbb{X}_i' \mathbf{z}_i + \Sigma_\beta^{-1} \boldsymbol{\mu}_\beta \right), \mathbb{M}_i \right)$$

A.3.2 Full Conditional for $\boldsymbol{\alpha}_i$

The full conditional distribution of $\boldsymbol{\alpha}_i$ can't be expressed in closed-form, only upto the following proportionality constant:

$$\pi \left(\boldsymbol{\alpha}_i|. \right) \propto \exp \left\{ -\frac{1}{2\sigma_i^2} \left(\mathbf{z}_i - \mathbb{X}_i \boldsymbol{\beta}_i \right)' \left(\mathbf{z}_i - \mathbb{X}_i \boldsymbol{\beta}_i \right) \right\} \times \left[\frac{1}{\gamma_i \tau_i} \exp \left\{ -\frac{1}{2} \left(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha \right)' \Sigma_\alpha^{-1} \left(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha \right) \right\} \right]$$

We use random-walk Metropolis to generate samples in the MCMC scheme.

A.3.3   Full Conditional for $\boldsymbol{\mu}_\beta$

Taking out the terms, which involves $\boldsymbol{\mu}_\beta$, from the joint density (equation A.1), we get

$$\pi\left(\boldsymbol{\mu}_\beta|.\right) \propto \prod_{i=1}^{m} \exp\left\{-\frac{1}{2}\left(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta\right)' \Sigma_\beta^{-1}\left(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta\right)\right\} \exp\left\{-\frac{1}{2}\left(\boldsymbol{\mu}_\beta - \mathbf{h}_\beta\right)' \mathbb{H}_\beta^{-1}\left(\boldsymbol{\mu}_\beta - \mathbf{h}_\beta\right)\right\}$$

$$= \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{m}\left(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta\right)' \Sigma_\beta^{-1}\left(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta\right) + \left(\boldsymbol{\mu}_\beta - \mathbf{h}_\beta\right)' \mathbb{H}_\beta^{-1}\left(\boldsymbol{\mu}_\beta - \mathbf{h}_\beta\right)\right]\right\}$$

$$= \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{m}\left(\boldsymbol{\beta}_i'\Sigma_\beta^{-1}\boldsymbol{\beta}_i - \boldsymbol{\beta}_i'\Sigma_\beta^{-1}\boldsymbol{\mu}_\beta - \boldsymbol{\mu}_\beta'\Sigma_\beta^{-1}\boldsymbol{\beta}_i + \boldsymbol{\mu}_\beta'\Sigma_\beta^{-1}\boldsymbol{\mu}_\beta\right) + \left(\boldsymbol{\mu}_\beta'\mathbb{H}_\beta^{-1}\boldsymbol{\mu}_\beta\right.\right.\right.$$

$$\left.\left.\left.-\boldsymbol{\mu}_\beta'\mathbb{H}_\beta^{-1}\mathbf{h}_\beta - \mathbf{h}_\beta'\mathbb{H}_\beta^{-1}\boldsymbol{\mu}_\beta + \mathbf{h}_\beta'\mathbb{H}_\beta^{-1}\mathbf{h}_\beta\right)\right]\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{m}\left(-\boldsymbol{\beta}_i'\Sigma_\beta^{-1}\boldsymbol{\mu}_\beta - \boldsymbol{\mu}_\beta'\Sigma_\beta^{-1}\boldsymbol{\beta}_i + \boldsymbol{\mu}_\beta'\Sigma_\beta^{-1}\boldsymbol{\mu}_\beta\right) + \boldsymbol{\mu}_\beta'\mathbb{H}_\beta^{-1}\boldsymbol{\mu}_\beta - \boldsymbol{\mu}_\beta'\mathbb{H}_\beta^{-1}\mathbf{h}_\beta\right.\right.$$

$$\left.\left.-\mathbf{h}_\beta'\mathbb{H}_\beta^{-1}\boldsymbol{\mu}_\beta\right]\right\}$$

[proportionality follows because $\boldsymbol{\beta}_i'\Sigma_\beta^{-1}\boldsymbol{\beta}_i$ and $\mathbf{h}_\beta'\mathbb{H}_\beta^{-1}\mathbf{h}_\beta$ are independent of $\boldsymbol{\mu}_\beta$]

$$= \exp\left\{-\frac{1}{2}\left[-2\boldsymbol{\mu}_\beta'\Sigma_\beta^{-1}\sum_{i=1}^{m}\boldsymbol{\beta}_i + m\boldsymbol{\mu}_\beta'\Sigma_\beta^{-1}\boldsymbol{\mu}_\beta + \boldsymbol{\mu}_\beta'\mathbb{H}_\beta^{-1}\boldsymbol{\mu}_\beta - 2\boldsymbol{\mu}_\beta'\mathbb{H}_\beta^{-1}\mathbf{h}_\beta\right]\right\}$$

[$\boldsymbol{\beta}_i'\Sigma_\beta^{-1}\boldsymbol{\mu}_\beta$ and $\mathbf{h}_\beta'\mathbb{H}_\beta^{-1}\boldsymbol{\mu}_\beta$ are scalers and so are $\boldsymbol{\beta}_i'\Sigma_\beta^{-1}\boldsymbol{\mu}_\beta = \boldsymbol{\mu}_\beta'\Sigma_\beta^{-1}\boldsymbol{\beta}_i$ and $\mathbf{h}_\beta'\mathbb{H}_\beta^{-1}\boldsymbol{\mu}_\beta = $

$\boldsymbol{\mu}_\beta'\mathbb{H}_\beta^{-1}\mathbf{h}_\beta$]

$$= \exp\left\{-\frac{1}{2}\left[-2\boldsymbol{\mu}_\beta'\left(\Sigma_\beta^{-1}\sum_{i=1}^{m}\boldsymbol{\beta}_i + \mathbb{H}_\beta^{-1}\mathbf{h}_\beta\right) + \boldsymbol{\mu}_\beta'\left(m\Sigma_\beta^{-1} + \mathbb{H}_\beta^{-1}\right)\boldsymbol{\mu}_\beta\right]\right\}$$

$$= \exp\left\{-\frac{1}{2}\left[-2\boldsymbol{\mu}_\beta'\left(\Sigma_\beta^{-1}\tilde{\boldsymbol{\beta}} + \mathbb{H}_\beta^{-1}\mathbf{h}_\beta\right) + \boldsymbol{\mu}_\beta'\mathbb{U}_\beta^{-1}\boldsymbol{\mu}_\beta\right]\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left[\boldsymbol{\mu}_\beta'\left(\Sigma_\beta^{-1}\tilde{\boldsymbol{\beta}} + \mathbb{H}_\beta^{-1}\mathbf{h}_\beta\right) - \left(\Sigma_\beta^{-1}\tilde{\boldsymbol{\beta}} + \mathbb{H}_\beta^{-1}\mathbf{h}_\beta\right)'\boldsymbol{\mu}_\beta + \boldsymbol{\mu}_\beta'\mathbb{U}_\beta^{-1}\boldsymbol{\mu}_\beta\right.\right.$$

$$\left.\left.+ \left(\Sigma_\beta^{-1}\tilde{\boldsymbol{\beta}} + \mathbb{H}_\beta^{-1}\mathbf{h}_\beta\right)'\mathbb{U}_\beta\left(\Sigma_\beta^{-1}\tilde{\boldsymbol{\beta}} + \mathbb{H}_\beta^{-1}\mathbf{h}_\beta\right)\right]\right\}$$

[$\boldsymbol{\mu}_\beta'(\Sigma_\beta^{-1}\tilde{\boldsymbol{\beta}} + \mathbb{H}_\beta^{-1}\mathbf{h}_\beta)$ is a scaler, and so $\boldsymbol{\mu}_\beta'(\Sigma_\beta^{-1}\tilde{\boldsymbol{\beta}} + \mathbb{H}_\beta^{-1}\mathbf{h}_\beta) = (\Sigma_\beta^{-1}\tilde{\boldsymbol{\beta}} + \mathbb{H}_\beta^{-1}\mathbf{h}_\beta)'\boldsymbol{\mu}_\beta$; propor-

tionality follows as $(\Sigma_\beta^{-1}\tilde{\boldsymbol{\beta}} + \mathbb{H}_\beta^{-1}\mathbf{h}_\beta)'\mathbb{U}_\beta(\Sigma_\beta^{-1}\tilde{\boldsymbol{\beta}} + \mathbb{H}_\beta^{-1}\mathbf{h}_\beta)$ is independent of $\boldsymbol{\mu}_\beta$]

$$= \exp\left\{-\frac{1}{2}\left[\boldsymbol{\mu}_\beta - \mathbb{U}_\beta\left(\Sigma_\beta^{-1}\tilde{\boldsymbol{\beta}} + \mathbb{H}_\beta^{-1}\mathbf{h}_\beta\right)\right]'\mathbb{H}_\beta^{-1}\left[\boldsymbol{\mu}_\beta - \mathbb{U}_\beta\left(\Sigma_\beta^{-1}\tilde{\boldsymbol{\beta}} + \mathbb{H}_\beta^{-1}\mathbf{h}_\beta\right)\right]\right\}$$

which is the density function of a trivariate normal distribution with mean vector $\mathbb{U}_\beta \left(\Sigma_\beta^{-1} \tilde{\boldsymbol{\beta}} + \mathbb{H}_\beta^{-1} \mathbf{h}_\beta \right)$ and covariance matrix $\mathbb{U}_\beta$. Hence,

$$[\boldsymbol{\mu}_\beta | .] \sim \mathcal{N}_3 \left(\mathbb{U}_\beta \left(\Sigma_\beta^{-1} \tilde{\boldsymbol{\beta}} + \mathbb{H}_\beta^{-1} \mathbf{h}_\beta \right), \mathbb{U}_\beta \right)$$

A.3.4 Full Conditional for $\boldsymbol{\mu}_\alpha$

Picking out the terms involving $\boldsymbol{\mu}_\alpha$ from the joint density (equation A.1), we have

$$\pi\left(\boldsymbol{\mu}_\alpha|.\right) \propto \prod_{i=1}^{m} \exp\left\{-\frac{1}{2}\left(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha\right)' \Sigma_\alpha^{-1}\left(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha\right)\right\} \exp\left\{-\frac{1}{2}\left(\boldsymbol{\mu}_\alpha - \mathbf{h}_\alpha\right)' \mathbb{H}_\alpha^{-1}\left(\boldsymbol{\mu}_\alpha - \mathbf{h}_\alpha\right)\right\}$$

$$= \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{m}\left(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha\right)' \Sigma_\alpha^{-1}\left(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha\right) + \left(\boldsymbol{\mu}_\alpha - \mathbf{h}_\alpha\right)' \mathbb{H}_\alpha^{-1}\left(\boldsymbol{\mu}_\alpha - \mathbf{h}_\alpha\right)\right]\right\}$$

$$= \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{m}\left(\boldsymbol{\xi}_i' \Sigma_\alpha^{-1}\boldsymbol{\xi}_i - \boldsymbol{\xi}_i' \Sigma_\alpha^{-1}\boldsymbol{\mu}_\alpha - \boldsymbol{\mu}_\alpha' \Sigma_\alpha^{-1}\boldsymbol{\xi}_i + \boldsymbol{\mu}_\alpha' \Sigma_\alpha^{-1}\boldsymbol{\mu}_\alpha\right)\right.\right.$$

$$\left.\left. + \left(\boldsymbol{\mu}_\alpha' \mathbb{H}_\alpha^{-1}\boldsymbol{\mu}_\alpha - \boldsymbol{\mu}_\alpha' \mathbb{H}_\alpha^{-1}\mathbf{h}_\alpha - \mathbf{h}_\alpha' \mathbb{H}_\alpha^{-1}\boldsymbol{\mu}_\alpha + \mathbf{h}_\alpha' \mathbb{H}_\alpha^{-1}\mathbf{h}_\alpha\right)\right]\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{m}\left(-\boldsymbol{\xi}_i' \Sigma_\alpha^{-1}\boldsymbol{\mu}_\alpha - \boldsymbol{\mu}_\alpha' \Sigma_\alpha^{-1}\boldsymbol{\xi}_i + \boldsymbol{\mu}_\alpha' \Sigma_\alpha^{-1}\boldsymbol{\mu}_\alpha\right) + \boldsymbol{\mu}_\alpha' \mathbb{H}_\alpha^{-1}\boldsymbol{\mu}_\alpha\right.\right.$$

$$\left.\left. - \boldsymbol{\mu}_\alpha' \mathbb{H}_\alpha^{-1}\mathbf{h}_\alpha - \mathbf{h}_\alpha' \mathbb{H}_\alpha^{-1}\boldsymbol{\mu}_\alpha\right]\right\}$$

[proportionality follows because $\boldsymbol{\xi}_i' \Sigma_\alpha^{-1}\boldsymbol{\xi}_i$ and $\mathbf{h}_\alpha' \mathbb{H}_\alpha^{-1}\mathbf{h}_\alpha$ are independent of $\boldsymbol{\mu}_\alpha$]

$$= \exp\left\{-\frac{1}{2}\left[-2\boldsymbol{\mu}_\alpha' \Sigma_\alpha^{-1}\sum_{i=1}^{m}\boldsymbol{\xi}_i + m\boldsymbol{\mu}_\alpha' \Sigma_\alpha^{-1}\boldsymbol{\mu}_\alpha + \boldsymbol{\mu}_\alpha' \mathbb{H}_\alpha^{-1}\boldsymbol{\mu}_\alpha - 2\boldsymbol{\mu}_\alpha' \mathbb{H}_\alpha^{-1}\mathbf{h}_\alpha\right]\right\}$$

[$\boldsymbol{\xi}_i' \Sigma_\alpha^{-1}\boldsymbol{\mu}_\alpha$ and $\mathbf{h}_\alpha' \mathbb{H}_\alpha^{-1}\boldsymbol{\mu}_\alpha$ are scalers, and so are $\boldsymbol{\xi}_i' \Sigma_\alpha^{-1}\boldsymbol{\mu}_\alpha = \boldsymbol{\mu}_\alpha' \Sigma_\alpha^{-1}\boldsymbol{\xi}_i$ and $\mathbf{h}_\alpha' \mathbb{H}_\alpha^{-1}\boldsymbol{\mu}_\alpha =$

$\boldsymbol{\mu}_\alpha' \mathbb{H}_\alpha^{-1}\mathbf{h}_\alpha$]

$$= \exp\left\{-\frac{1}{2}\left[-2\boldsymbol{\mu}_\alpha'\left(\Sigma_\alpha^{-1}\sum_{i=1}^{m}\boldsymbol{\xi}_i + \mathbb{H}_\alpha^{-1}\mathbf{h}_\alpha\right) + \boldsymbol{\mu}_\alpha'\left(m\Sigma_\alpha^{-1} + \mathbb{H}_\alpha^{-1}\right)\boldsymbol{\mu}_\alpha\right]\right\}$$

$$= \exp\left\{-\frac{1}{2}\left[-2\boldsymbol{\mu}_\alpha'\left(\Sigma_\alpha^{-1}\tilde{\boldsymbol{\xi}} + \mathbb{H}_\alpha^{-1}\mathbf{h}_\alpha\right) + \boldsymbol{\mu}_\alpha'\mathbb{U}_\alpha^{-1}\boldsymbol{\mu}_\alpha\right]\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left[-\boldsymbol{\mu}_\alpha'\left(\Sigma_\alpha^{-1}\tilde{\boldsymbol{\xi}} + \mathbb{H}_\alpha^{-1}\mathbf{h}_\alpha\right) - \left(\Sigma_\alpha^{-1}\tilde{\boldsymbol{\xi}} + \mathbb{H}_\alpha^{-1}\mathbf{h}_\alpha\right)'\boldsymbol{\mu}_\alpha + \boldsymbol{\mu}_\alpha'\mathbb{U}_\alpha^{-1}\boldsymbol{\mu}_\alpha\right.\right.$$

$$\left.\left. + \left(\Sigma_\alpha^{-1}\tilde{\boldsymbol{\xi}} + \mathbb{H}_\alpha^{-1}\mathbf{h}_\alpha\right)'\mathbb{U}_\alpha\left(\Sigma_\alpha^{-1}\tilde{\boldsymbol{\xi}} + \mathbb{H}_\alpha^{-1}\mathbf{h}_\alpha\right)\right]\right\}$$

[$\boldsymbol{\mu}_\alpha'(\Sigma_\alpha^{-1}\tilde{\boldsymbol{\xi}} + \mathbb{H}_\alpha^{-1}\mathbf{h}_\alpha)$ is a scaler, and so is $\boldsymbol{\mu}_\alpha'(\Sigma_\alpha^{-1}\tilde{\boldsymbol{\xi}} + \mathbb{H}_\alpha^{-1}\mathbf{h}_\alpha) = (\Sigma_\alpha^{-1}\tilde{\boldsymbol{\xi}} + \mathbb{H}_\alpha^{-1}\mathbf{h}_\alpha)'\boldsymbol{\mu}_\alpha$; propor-

tionality follows because $(\Sigma_\alpha^{-1}\tilde{\boldsymbol{\xi}} + \mathbb{H}_\alpha^{-1}\mathbf{h}_\alpha)'\mathbb{U}_\alpha(\Sigma_\alpha^{-1}\tilde{\boldsymbol{\xi}} + \mathbb{H}_\alpha^{-1}\mathbf{h}_\alpha)$ is independent of $\boldsymbol{\mu}_\alpha$]

$$= \exp\left\{-\frac{1}{2}\left[\boldsymbol{\mu}_\alpha - \mathbb{U}_\alpha\left(\Sigma_\alpha^{-1}\tilde{\boldsymbol{\xi}} + \mathbb{H}_\alpha^{-1}\mathbf{h}_\alpha\right)\right]'\mathbb{U}_\alpha^{-1}\left[\boldsymbol{\mu}_\alpha - \mathbb{U}_\alpha\left(\Sigma_\alpha^{-1}\tilde{\boldsymbol{\xi}} + \mathbb{H}_\alpha^{-1}\mathbf{h}_\alpha\right)\right]\right\}$$

which is proportional to the density function of a bivariate normal distribution with mean vector $\mathbb{U}_\alpha \left(\Sigma_\alpha^{-1} \tilde{\boldsymbol{\xi}} + \mathbb{H}_\alpha^{-1} \mathbf{h}_\alpha \right)$ and covariance matrix $\mathbb{U}_\alpha$. Therefore,

$$[\boldsymbol{\mu}_\alpha |.] \sim \mathcal{N}_2 \left(\mathbb{U}_\alpha \left(\Sigma_\alpha^{-1} \tilde{\boldsymbol{\xi}} + \mathbb{H}_\alpha^{-1} \mathbf{h}_\alpha \right), \mathbb{U}_\alpha \right)$$

A.3.5   Full Conditional for Σ_β^{-1}

Taking out the terms involving Σ_β^{-1} from the joint density (equation A.1), we have

$$\pi\left(\Sigma_\beta^{-1}|.\right) \propto \prod_{i=1}^{m} \frac{1}{|\Sigma_\beta|^{1/2}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta\right)' \Sigma_\beta^{-1}\left(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta\right)\right\} |\Sigma_\beta^{-1}|^{\frac{\nu_\beta-3-1}{2}} \exp\left\{-\frac{\nu_\beta}{2} tr\left(\mathbb{A}_\beta \Sigma_\beta^{-1}\right)\right\}$$

$$= \frac{1}{|\Sigma_\beta|^{m/2}} \exp\left\{-\frac{1}{2} \sum_{i=1}^{m}\left(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta\right)' \Sigma_\beta^{-1}\left(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta\right)\right\} |\Sigma_\beta^{-1}|^{\frac{\nu_\beta-3-1}{2}} \exp\left\{-\frac{\nu_\beta}{2} tr\left(\mathbb{A}_\beta \Sigma_\beta^{-1}\right)\right\}$$

$$= |\Sigma_\beta^{-1}|^{\frac{m+\nu_\beta-3-1}{2}} \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{m}(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta)'\Sigma_\beta^{-1}(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta) + tr(\nu_\beta \mathbb{A}_\beta \Sigma_\beta^{-1})\right]\right\}$$

$$= |\Sigma_\beta^{-1}|^{\frac{(m+\nu_\beta)-3-1}{2}} \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{m} tr\left(\left(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta\right)\left(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta\right)' \Sigma_\beta^{-1}\right) + tr\left(\nu_\beta \mathbb{A}_\beta \Sigma_\beta^{-1}\right)\right]\right\}$$

[Assuming $\mathbf{d}_{1i} = (\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta)'\Sigma_\beta^{-1}$ and $\mathbf{d}_{2i} = (\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta)$, we can write $\mathbf{d}_{1i}\mathbf{d}_{2i} = tr(\mathbf{d}_{i1}\mathbf{d}_{i2}) =$

$tr(\mathbf{d}_{2i}\mathbf{d}_{1i})$ by the property of trace of a matrix]

$$= |\Sigma_\beta^{-1}|^{\frac{(m+\nu_\beta)-3-1}{2}} \exp\left\{-\frac{1}{2}\left[tr\left(\sum_{i=1}^{m}\left(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta\right)\left(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta\right)' \Sigma_\beta^{-1}\right) + tr\left(\nu_\beta \mathbb{A}_\beta \Sigma_\beta^{-1}\right)\right]\right\}$$

$$= |\Sigma_\beta^{-1}|^{\frac{(m+\nu_\beta)-3-1}{2}} \exp\left\{-\frac{1}{2}\left[tr\left(\sum_{i=1}^{m}\left(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta\right)\left(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta\right)' + \nu_\beta \mathbb{A}_\beta\right) \Sigma_\beta^{-1}\right]\right\}$$

which is a proportional to the density function of Wishart distribution with degrees of freedom $m + \nu_\beta$ and scale matrix $\left[\sum_{i=1}^{m}\left(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta\right)\left(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta\right)' + \nu_\beta \mathbb{A}_\beta\right]^{-1}$. Thus,

$$\left[\Sigma_\beta^{-1}|.\right] \sim \mathcal{W}\left(m + \nu_\beta, \left[\sum_{i=1}^{m}(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta)(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta)' + \nu_\beta \mathbb{A}_\beta\right]^{-1}\right)$$

A.3.6 Full Conditional for Σ_α^{-1}

Picking out the terms involving Σ_α^{-1} from the joint density (equation A.1), we get

$$\pi\left(\Sigma_\alpha^{-1}|.\right) \propto \prod_{i=1}^{m} \frac{1}{|\Sigma_\alpha|^{1/2}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha\right)' \Sigma_\alpha^{-1} \left(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha\right)\right\} |\Sigma_\alpha^{-1}|^{\frac{\nu_\alpha-2-1}{2}} \exp\left\{-\frac{\nu_\alpha}{2} tr\left(\mathbb{A}_\alpha \Sigma_\alpha^{-1}\right)\right\}$$

$$= \frac{1}{|\Sigma_\alpha|^{m/2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{m}\left(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha\right)' \Sigma_\alpha^{-1} \left(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha\right)\right\} |\Sigma_\alpha^{-1}|^{\frac{\nu_\alpha-2-1}{2}} \exp\left\{-\frac{\nu_\alpha}{2} tr\left(\mathbb{A}_\alpha \Sigma_\alpha^{-1}\right)\right\}$$

$$= |\Sigma_\alpha^{-1}|^{\frac{m+\nu_\alpha-2-1}{2}} \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{m}\left(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha\right)' \Sigma_\alpha^{-1} \left(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha\right) + tr\left(\nu_\alpha \mathbb{A}_\alpha \Sigma_\alpha^{-1}\right)\right]\right\}$$

$$= |\Sigma_\alpha^{-1}|^{\frac{(m+\nu_\alpha)-2-1}{2}} \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{m} tr\left(\left(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha\right)\left(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha\right)' \Sigma_\alpha^{-1}\right) + tr\left(\nu_\alpha \mathbb{A}_\alpha \Sigma_\alpha^{-1}\right)\right]\right\}$$

[Again, assuming $\mathbf{d}_{1i} = \left(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha\right)' \Sigma_\alpha^{-1}$ and $\mathbf{d}_{2i} = \mathbb{I}_i\left(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha\right)$, we can write $\mathbf{d}_{1i}\mathbf{d}_{2i} =$

$tr\left(\mathbf{d}_{i1}\mathbf{d}_{i2}\right) = tr\left(\mathbf{d}_{2i}\mathbf{d}_{1i}\right)$ by the property of trace of a matrix]

$$= |\Sigma_\alpha^{-1}|^{\frac{(m+\nu_\alpha)-2-1}{2}} \exp\left\{-\frac{1}{2}\left[tr\left(\sum_{i=1}^{m}\left(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha\right)\left(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha\right)' \Sigma_\alpha^{-1}\right) + tr\left(\nu_\alpha \mathbb{A}_\alpha \Sigma_\alpha^{-1}\right)\right]\right\}$$

$$= |\Sigma_\alpha^{-1}|^{\frac{(m+\nu_\alpha)-2-1}{2}} \exp\left\{-\frac{1}{2}\left[tr\left(\sum_{i=1}^{m}\left(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha\right)\left(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha\right)' + \nu_\alpha \mathbb{A}_\alpha\right) \Sigma_\alpha^{-1}\right]\right\}$$

which is proportional to the density function of Wishart distribution with degrees of freedom

$m + \nu_\alpha$ and scale matrix $\left[\sum_{i=1}^{m}\left(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha\right)\left(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha\right)' + \nu_\alpha \mathbb{A}_\alpha\right]^{-1}$. Thus,

$$\left[\Sigma_\alpha^{-1}|.\right] \sim \mathcal{W}\left(m + \nu_\alpha, \left[\sum_{i=1}^{m}\left(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha\right)\left(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha\right)' + \nu_\alpha \mathbb{A}_\alpha\right]^{-1}\right)$$

A.3.7 Full Conditional for σ_i^{-2}

Taking out the relevant terms for σ_i from the joint density (equation A.1), we have

$$\pi\left(\sigma_i^{-2}|.\right) \propto \frac{1}{\mid \sigma_i^2 \mathrm{I}_i \mid^{1/2}} \exp\left\{-\frac{1}{2\sigma_i^2}\left(\mathbf{y}_i - \boldsymbol{\mu}_i\right)'\left(\mathbf{y}_i - \boldsymbol{\mu}_i\right)\right\}\left(\sigma_i^{-2}\right)^{\frac{\mathbf{d}_0}{2}-1}\exp\left\{-\frac{\mathbf{d}_1}{2}\sigma_i^{-2}\right\}$$

$$= \frac{1}{\mid \sigma_i^2 \mathrm{I}_i \mid^{1/2}} \exp\left\{-\frac{1}{2\sigma_i^2}\left(\mathbf{z}_i - \mathbb{X}_i\boldsymbol{\beta}_i\right)'\left(\mathbf{z}_i - \mathbb{X}_i\boldsymbol{\beta}_i\right)\right\}\left(\sigma_i^{-2}\right)^{\frac{\mathbf{d}_0}{2}-1}\exp\left\{-\frac{\mathbf{d}_1}{2}\sigma_i^{-2}\right\}$$

$$= \left(\sigma_i^{-2}\right)^{\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma_i^2}\left(\mathbf{z}_i - \mathbb{X}_i\boldsymbol{\beta}_i\right)'\left(\mathbf{z}_i - \mathbb{X}_i\boldsymbol{\beta}_i\right)\right\}\left(\sigma_i^{-2}\right)^{\frac{\mathbf{d}_0}{2}-1}\exp\left\{-\frac{\mathbf{d}_1}{2}\sigma_i^{-2}\right\}$$

$$= \left(\sigma_i^{-2}\right)^{\frac{n+\mathbf{d}_0}{2}-1} \exp\left\{-\frac{\left(\mathbf{z}_i - \mathbb{X}_i\boldsymbol{\beta}_i\right)'\left(\mathbf{z}_i - \mathbb{X}_i\boldsymbol{\beta}_i\right) + \mathbf{d}_1}{2}\sigma_i^{-2}\right\}$$

which is proportional to the density function of a Gamma distribution with shape parameter

as $\frac{n+\mathbf{d}_0}{2}$ and rate parameter $\frac{(\mathbf{z}_i-\mathbb{X}_i\boldsymbol{\beta}_i)'(\mathbf{z}_i-\mathbb{X}_i\boldsymbol{\beta}_i)+\mathbf{d}_1}{2}$. Thus,

$$\left[\sigma_i^{-2}|.\right] \sim \mathcal{G}\left(\frac{n+\mathbf{d}_0}{2}, \frac{\left(\mathbf{z}_i - \mathbb{X}_i\boldsymbol{\beta}_i\right)'\left(\mathbf{z}_i - \mathbb{X}_i\boldsymbol{\beta}_i\right) + \mathbf{d}_1}{2}\right)$$

A.3.8 Full Conditional for δ

The full conditional distribution of δ cannot be expressed as a standard density function. Abstracting out the relevant terms from the joint density (equation A.1), we can write it as the following proportionality constant:

$$\pi\left(\delta|.\right) \propto \mid \mathrm{I}_m - \delta W \mid^n \exp\left\{-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n}\sigma_i^{-2}\left[y_{ij} - \delta\sum_{k=1}^{m}w_{ik}\left(y_{kj} - f_{kj}\right) - f_{ij}\right]^2\right\}$$

where $\mid \mathrm{I}_m - \delta W \mid^n$ is the Jacobian term taking into account the effect of $\sum_{k=1}^{m}w_{ik}e_{kj}$

$$= \mid \mathrm{I}_m - \delta W \mid^n \exp\left\{-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n}\sigma_i^{-2}\left(\mathbb{Z}_i - \mathbb{X}_i\boldsymbol{\beta}_i\right)'\left(\mathbb{Z}_i - \mathbb{X}_i\boldsymbol{\beta}_i\right)\right\}$$