8-12-2015

Dynamic Modeling of Multivariate Counts - Fitting, Diagnostics, and Applications

Volodymyr Serhiyenko
University of Connecticut - Storrs, volodymyr.serhiyenko@gmail.com

Dynamic Modeling of Multivariate Counts – Fitting, Diagnostics, and Applications

Volodymyr Serhiyenko, Ph.D.

University of Connecticut, 2015

ABSTRACT

An adequate statistical methodology is required for modeling multivariate time series of counts. The proper specification of the underlying distribution in such modeling could be very challenging, as it should account for the possibility of overdispersion, an excessive number of zero values, positive and negative association between counts, etc.

This dissertation is focused on modeling multivariate time series of counts as a function of location-specific and time-dependent covariates. The Bayesian framework for estimation and prediction is discussed. We focus on Markov chain Monte Carlo (MCMC) methods for fully Bayesian inference and the Integrated Nested Laplace Approximation (INLA) for fast implementation of approximate Bayesian modeling which is especially useful for large data sets.

The dissertation has three main contributions. First, we propose a dynamic model that combines time series compositional modeling with dynamic modeling for counts. This approach is applied to the problem of transportation engineering. We investigate the temporal behavior of injury severity levels as proportions of all pedestrian crashes in each month, taking into consideration effects of time trend, seasonal variations and

VMT (vehicle miles traveled).

Second, this dissertation discusses a hierarchical multivariate dynamic modeling framework. The use of a multivariate Poisson (MVP) sampling distribution is discussed. We show that the use of such distribution enables us to model the association between components of the multivariate response vector over time. This approach is illustrated using data from ecology on gastropod abundance in Puerto Rico.

Finally, we propose a level correlated model (LCM) to account for the association among the components of the response vector. This multivariate model accounts for overdispersion as well as for positive and negative association between counts. The flexible LCM framework allows us to combine different marginal count distributions and to build a dynamic model for the vector time series of counts. We comprehensively discuss the lower and upper limits for the association between the components of the response vector of counts. We employ the proposed modeling to ecology and marketing examples and discuss the results.

Dynamic Modeling of Multivariate Counts – Fitting, Diagnostics, and Applications

Volodymyr Serhiyenko

M.S. Mathematics, The City College of New York, 2010

B.S. Mathematics, Lviv Ivan Franko National University, 2007

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at the

University of Connecticut

2015

APPROVAL PAGE

Doctor of Philosophy Dissertation

Dynamic Modeling of Multivariate Counts – Fitting, Diagnostics, and Applications

Presented by

Volodymyr Serhiyenko, B.S. Math., M.S. Math.

Major Advisor _____
Nalini Ravishanker

Associate Advisor _____
Xiaojing Wang

Associate Advisor _____
John N. Ivan

University of Connecticut

2015

ACKNOWLEDGMENTS

Since first I came to Storrs, I could not imagine that in five years I will be leaving it with so much unforgettable memories. It was a challenging, nevertheless incredible and delightful period of my life. I would like to acknowledge all the great people that I was lucky to meet during my PhD journey at the University of Connecticut.

First of all, I would like to express my highest gratitude to my major advisor, Professor Nalini Ravishanker. Without any hesitation I can say that she has been responsible for my success during these five years. Professor Ravishanker shared her endless knowledge and gave important and crucial advise during our countless meetings, Skype talks, projects, and conferences. Her patience, enthusiasm, encouragement, and critique helped me to expand my research in numerous directions. I could not have imagined having a better advisor and mentor for my PhD study.

I would like to express my very great appreciation to my associate advisors Professor Wang and Professor Ivan for all of their vital input in my research. Their comments and guidance helped me a great deal in the writing of this work and enhanced my dissertation in various ways.

I would like to offer my special thanks to all faculty members in the Department of Statistics at University of Connecticut. Especially, I want to express my sincere appreciation to Professor Nitis Mukhopadhyay, Professor Cyr M'Lan, Professor Zhiyi Chi, and Dr. Naitee Ting for all their support and inspiration during my graduate study. I also want to thank Tracy Burke and Megan Petsa for the immense work that

they put behind my five years of study.

My sincere thanks go to Professor Mike Willig and Professor Rajkumar Venkatesan for providing data and expert opinions in the ecology and marketing data analysis examples. I'm also very fortune to have met and worked with great classmates and colleagues from Statistics, Ecology, and Civil Engineering departments.

I am extremely grateful to my mother Oleksandra Serhiyenko, my father Anatoliy Serhiyenko, my sister Nataliya Lyakhovych and all of her family. They always support me in all of my endeavors. Also, I want to thank my newly acquired Nechyporenko family for treating me like their own son.

Lastly and the most importantly, I want to thank my wife, Khrystyna Serhiyenko. I cannot express in words how grateful I am for having her by my side. Her unconditional support and love are the most important drivers in my life. Together we had many ups and downs while in the graduate school, and now we are fearlessly waiting for a new chapter in our life.

Contents

Chapter 1

Introduction and Literature Survey

1.1 Univariate and Multivariate Modeling of Counts

The scientific literature is rich with regression modeling problems where the response variables are independent counts. The standard approach uses the Poisson distribution as the response distribution and implements GLM (McCullagh and Nelder, 1989; Agresti, 2010). A serious drawback of the Poisson regression model is that it assumes a nominal dispersion, i.e., the mean is equal to the variance. In practice, count data are often overdispersed, i.e., show evidence that the variance is larger than the mean (Cameron and Trivedi, 1998).

The general solution to this problem is to assume that the Poisson mean is a mixture of fixed mean and a positive random variable. In this case, a nonnegative multiplicative random effect term is introduced into the model. The popular choice for the distribution of the random effect term is a gamma distribution. These types of models result in the negative binomial regression models (Cameron and Trivedi, 1998;

Winkelmann, 2008). In this context Bayesian inference has also been developed. Unfortunately it involves computationally intensive Markov chain Monte Carlo (MCMC) algorithms, since there is no conjugate prior for regression coefficients when the underlying distribution is Poisson or negative binomial (Chib et al., 1998; Chib and Winkelmann, 2001; Winkelmann, 2008).

A lognormal distribution can also be used as a random effect in the Poisson model, which results in the Poisson-lognormal regression model as an alternative to the negative binomial model (Breslow, 1984; Agresti, 2010). Moreover, Zhou et al. (2012) suggested that lognormal and gamma mixed negative binomial regression model for counts may fit the data better and has one extra degree of freedom to incorporate different kinds of random effects. They also developed Bayesian inference for that kind of model. The Conway-Maxwell-Poisson (COM-Poisson) was first introduced by Conway and Maxwell (1962) and just recently was used for modeling count data (Shmueli et al., 2005) in situations of under- or over-dispersion. The regression model was developed by Guikema and Coffelt (2008) for risk analysis and Lord et al. (2008) for traffic accident data. They gave a comprehensive discussion about non-Bayesian and Bayesian multivariate linear regressions, generalized linear regression models, as well as semi-parametric and non-parametric models.

The zero-inflated count models were introduced to handle data with a preponderance of zeros (Johnson and Kotz, 1969; Lambert, 1992) and applied in different fields through zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB) regression models (Shankar et al., 1997). The assumptions of zero-inflated models were discussed with respect to different applications in transportation engineering (Lord et al., 2005).

The literature on the use of the multivariate Poisson distribution (Mahamunulu,

1967) for modeling applications was sparse until recently, possibly due to the complicated form of the probability mass function. Karlis and Meligkotsidou (2005) proposed the two-way covariance structured multivariate Poisson distribution which permits a more realistic modeling of multivariate counts for several practical applications. Hu (2012) developed a Bayesian framework for regression and time series models based on the multivariate Poisson distribution. Nevertheless, this type of multivariate Poisson model assumes positive dependence between components of the vector-valued count variable, an assumption that is not realistic for several applications. Further, the marginal mean and variance of each variable coincide and, thus, this model is not appropriate for overdispersed data sets. Karlis and Meligkotsidou (2007) proposed a finite multivariate Poisson mixture as an alternative class of models for multivariate count data. These models allow for both negative and positive dependence and overdispersion. However, the computations can be very time consuming.

Another approach is the multivariate Poisson-lognormal model (Aitchison and Ho, 1989; Ma et al., 2008). The approach allows modeling dependence within the response vector for data that are possibly overdispersed. Although a Bayesian approach (Chib and Winkelmann, 2001; Ma et al., 2008) was developed for the regression model, the possibility of temporal dependence was not considered when the data are collected on different locations/segments over time. Moreover, a comprehensive description of properties for induced association between the components of the count response vector is missing from the literature. Although it was noted that the range of induced correlations is not as wide as that of the corresponding lognormal or normal distributions (Aitchison and Ho, 1989). It will be extremely useful to carefully quantify the values of the maxima and minima of the induced associations between the count

variables as a function of the marginal Poisson means and the variances and correlations of the components from the underlying lognormal distribution. Moreover, the possibility of using of the multivariate Poisson-lognormal regression in the case of time series and its extensions has not been discussed in the literature as well.

1.2 Count Time Series Modeling

The literature on count time series modeling includes observation driven models and parameter driven models. The generalized linear modeling (GLM) framework (Nelder and Wedderburn, 1972; McCullagh and Nelder, 1989) naturally combines the traditional time series models such as the autoregressive models (AR), moving average models (MA), autoregressive moving average models (ARMA), seasonal ARMA, generalized autoregressive conditional heteroskedasticity models (GARCH), etc.

The Poisson distribution is the first obvious choice of the underlying distribution for the time series modeling of counts under the GLM framework. In situations with overdisperssion, the negative binomial distribution can be considered. There are several count distributions that can be easily adopted for the time series setup such as the ZIP distribution (Lambert, 1992) and the truncated Poisson distribution (Fokianos, 2001). Excellent reviews of such time series models by Fokianos (2012), Tjøstheim (2012), and Davis et al. (2015) are valuable. We should note that the model estimation, diagnostics, and forecasting are implemented in various standard statistical packages and are readily available for the researches. The R package *tscount* gives a flexible framework for the estimation of count time series which follow generalized linear models. The R package *glarma* is another package that provides functions

for fitting generalized linear autoregressive moving average (GLARMA) models for discrete valued time series.

The model estimation can also be done using likelihood methods. Maximum likelihood inference for Poisson and negative binomial time series models has been developed by Davis et al. (2003), Fokianos et al. (2009), Fokianos and Tjøstheim (2012), and Christou and Fokianos (2014). Regression modeling for count time series using quasi-likelihood methods was discussed in Zeger (1998). Jørgensen et al. (1999) described analysis of longitudinal multivariate count data driven by a latent gamma Markov process using a state space approach. Song (2007) discussed marginal, conditional (random effects), and transitional approaches for analyzing correlated counts. Bayesian modeling of panel count data under the Poisson-lognormal model was discussed in Chib et al. (1998), while Chib and Winkelmann (2001) discussed models with latent effects for correlated count data.

One approach to develop models for count time series is based on the thinning operator (Steutel and Van Harn, 1979), where the thinning operator is generated by counting series of Bernoulli-distributed random variables. McKenzie (1985) and Al-Osh and Alzaid (1987) independently developed the first-order integer-valued autoregressive, INAR(1) model. McKenzie (2003) and Jung and Tremayne (2006) presented a good review of subsequent developments. The class of integer valued time series models based on thinning operations is more restrictive in its construction than models based on the GLM framework (Tjøstheim, 2012). One of the drawbacks is that the autocorrelation is always positive. Also, nonlinear and multivariate extensions are not easily implemented under this framework (Drost et al., 2008; McKenzie, 2003). A good review of such models is given by Weiß (2008). Multivariate INAR models for counts were also discussed in Pedeli and Karlis (2011) and references therein.

For Gaussian dynamic linear models (DLMs), often referred to as Gaussian state space models, Kalman (1960) and Kalman and Bucy (1961) popularized a recursive algorithm for optimal estimation and prediction of the state vector, which then enables the prediction of the observation vector. The use of Gaussian state space models has gained in popularity since the books of Harvey (1990) and West and Harrison (1997). These books give comprehensive overview of the class of state space models from the classical and Bayesian perspectives. Hierarchical dynamic linear models (HDLMs) combine the stratified parametric linear models (Lindley and Smith, 1972) and the DLMs into a general framework, which have been particularly useful in econometric, education, and health-care applications. The Gaussian HDLM includes a set of one or more dimension reducing structural equations along with the observation equation and state (system) equation of the DLM (Gamerman and Migon, 1993). Landim and Gamerman (2000) further extended the Gaussian HDLM to a more general class of models where the response vector has a matrix-valued normal distribution. Count data models in the state space approach were discussed in Gamerman (1998), Durbin and Koopman (2000), Frühwirth-Schnatter and Wagner (2006), and Gamerman et al. (2013).

There is a considerable amount of literature on MCMC methods for non-Gaussian and non-linear state space models; see West et al. (1985), Carlin et al. (1992), Gordon et al. (1993), Song (2007). A detailed review of the MCMC methods can be found in Fearnhead (2011) and Migon et al. (2005). The simplest method updates the components of the states in the sequential fashion (Carlin et al., 1992; Geweke and Tanizaki, 2001). However, it is well known that this type of sampler may lead to slow mixing because of possible strong correlation between states. In such cases it is better to update the states in multiple instances as blocks of states or update the

entire state process at once (Shephard and Pitt, 1997; Carter and Kohn, 1994). We note that designing MCMC algorithms and tuning those can be cumbersome and time consuming, especially for multivariate state space models and regression and time series models for counts.

1.3 Contribution of this dissertation

In several application areas, we increasingly see the need for developing accurate statistical modeling approaches for time series of multivariate count responses. The response consists of an J-dimensional vector of counts that is observed at each of n locations (or for each of n subjects) over T regularly spaced times. The objective of the statistical analysis is to understand stochastic temporal patterns in the response as a function of observed location (or subject)-specific and/or time-varying covariates. For instance, in ecology, understanding the causes and consequences of variation in the abundance of organisms as a function of topographical and environmental covariates has been a long-standing goal (Krebs, 1972; Scheiner and Willig, 2011). In business, a pharmaceutical firm may be interested in estimating and predicting the number of new prescriptions written by physicians of drugs from the firm and its competitors, as a function of the firm's promotional activities (Venkatesan et al., 2012). In a problem in transportation engineering, it would be interesting to understand stochastic patterns in the temporal behavior of crash counts categorized by injury severity across a set of highway segments, as a function of roadway geometry, traffic volume, etc. (Hu et al., 2012). For such applications, the modeling described in this dissertation enables us to adequately incorporate dependence in the response over time as well as the

dependence between the components of the response vector.

This dissertation is focused on modeling multivariate time series of counts as a function of location-specific and time-dependent covariates. The proper specification of the underlying distribution could be very challenging. It is well known that count data can exhibit overdispersion or underdispersion relative to the Poisson distribution, or an excessive number of zero values. All of these situations cannot be adequately modeled by the Poisson distribution. The question that also arises is what framework to employ if the data require multivariate modeling. We focus on Markov chain Monte Carlo (MCMC) methods for fully Bayesian inference and the Integrated Nested Laplace Approximation (INLA) (Rue et al., 2009) for fast implementation of approximate Bayesian modeling for large data sets.

Chapter 3 describes the use of dynamic modeling of compositional time series derived from multivariate counts. We introduce a new model that combines the dynamic and static models for counts and time series models of compositions under one framework. We investigate the temporal behavior of injury severity levels as proportions of all pedestrian crashes in each month, taking into consideration effects of time trend, seasonal variations and VMT (vehicle miles traveled). We describe a time series framework with vector autoregressions (VAR) for modeling and predicting compositional time series. Combining these predictions with predictions from a univariate statistical model for total crash counts enables us to predict pedestrian crash counts with different injury severity levels.

Chapter 4 gives details of a hierarchical multivariate dynamic model (HMDM) for a vector-valued time series of counts. We describe a fully Bayesian framework for estimation and prediction by assuming a multivariate Poisson (MVP) sampling distribution for the count responses. Our modeling incorporates the temporal dependence

as well as dependence between the components of the response vector.

The use of the MVP distribution also enables us to model the association between components as a function of subject/location and time specific covariates that can vary over time. We show that the use of MVP distributions enables us to model the association between components of the multivariate response vector over time. We apply this methodology to an example from ecology. Also, we discuss the computational time associated with models that use MVP distributions. We note that the proposed models can not account for the negative association between the components of the vector-valued response counts.

In Chapter 5 we propose a level correlated model (LCM) to account for the association among the components of the response vector. This model for multivariate time series of counts accounts for overdispersion as well as for positive and/or negative association between the components. The flexible LCM framework allows us to combine different marginal count distributions and to build a dynamic model for the vector time series of counts. We comprehensively discuss the lower and upper limits for the association between the components of the response vector. The maximum and minimum limit values for the strength of the association are derived for many common situations under the LCM framework. We show in detail the model for multivariate time series of counts observed on different subjects/locations as a function of subject/location and/or time-dependent covariates. We employ the Integrated Nested Laplace Approximation (INLA) approach for fast approximate Bayesian modeling. The LCM is used with Poisson marginal distributions to model the abundance of gastropod species in Puerto Rico. We also explore the possibility of using combinations of different underlying distributions of counts in the marketing example (the monthly prescription counts by physicians of a focal, leader, and challenger drugs).

We provide a description of multivariate mixture of Poisson and ZIP models under the LCM framework and discuss the performance of these models.

Chapter 2

Review of Bayesian and Approximate Bayesian Methods

2.1 Markov Chain Monte Carlo sampling algorithms

There are many data applications which require building large dimensional models. Dynamic models are examples of models with complex structure which can incorporate hierarchical as well as random effects. Under the Bayesian paradigm the model consolidates uncertainties associated with all unknown quantities whether they are explicitly observed or act through the latent state (Gamerman and Lopes, 2006). Markov chain Monte Carlo (MCMC) is a large class of techniques that enables inference in highly dimensional problems with unknown quantities and is able to handle complicated distributions. For an excellent overview of MCMC Methods in Bayesian computation please refer to books Gamerman and Lopes (2006), Robert and Casella (1999), and Chen et al. (2000). We are typically interested in computing posterior

quantities from the known, simulated, or approximated posterior distributions. Common posterior quantities include the posteriors means, standard deviations, medians, quantiles, credible intervals, etc. One of the most popular and basic techniques is Gibbs sampling (Geman and Geman, 1984; Gelfand and Smith, 1990; Robert, 1994). Most of the time this algorithm is used when the joint distribution is fairly complex, however the conditional distributions are relatively simple. Another technique is given by the Metropolis-Hastings (MH) algorithm (Metropolis et al., 1953; Hastings, 1970; Chib and Greenberg, 1995). It is used when the form of conditional distribution is not available as a known density. We give a brief description of these two techniques below.

The Gibbs sampler is one of the best known MCMC sampling algorithms for the Bayesian computations. Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)'$ denotes a vector of all parameters associated with a model and let $\boldsymbol{y}$ denote the observed data. Also let $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ denote the posterior distribution of $\boldsymbol{\theta}$ given $\boldsymbol{y}$. Then the Gibbs sampling algorithm is given as follows:

- **Step 0**. Choose an arbitrary starting point $\boldsymbol{\theta}^0 = (\theta_1^0, \ldots, \theta_p^0)'$, and set $i = 0$;
- **Step 1**. Generate the next value of $\boldsymbol{\theta}^{i+1} = (\theta_1^{i+1}, \ldots, \theta_p^{i+1})'$ as follows:
  - Generate $\theta_1^{i+1} \sim \pi(\theta_1^{i+1}|\theta_2^i, \ldots, \theta_p^i, \boldsymbol{y})$;
  - Generate $\theta_2^{i+1} \sim \pi(\theta_2^{i+1}|\theta_1^{i+1}, \theta_3^i, \ldots, \theta_p^i, \boldsymbol{y})$;
  - .....
  - Generate $\theta_p^{i+1} \sim \pi(\theta_p^{i+1}|\theta_1^{i+1}, \ldots, \theta_{p-1}^{i+1}, \boldsymbol{y})$.
- **Step 2**. Set $i = i + 1$ and repeat **Step 1**.

Once the chain has converged, the value $\boldsymbol{\theta}^i$ is a draw from the posterior distribution $\pi(\boldsymbol{\theta}|\boldsymbol{y})$. With the increase in the number of iterations, the algorithm coverages to the equilibrium condition. This approach requires the conditional distributions $\theta_k \sim \pi(\theta_k|\theta_1, \ldots, \theta_{k-1}, \theta_{k+1}, \ldots, \theta_p, \boldsymbol{y})$ to be readily available and the analytical forms to be known. If the conditional distributions are not known, then the Metropolis-Hastings algorithms can be used. The MH algorithm is based on two parts: a proposal and an acceptance of the proposal. Let $q(\boldsymbol{\theta}, \boldsymbol{\phi})$ be a proposal density and Uniform$(0, 1)$ to be the uniform distribution on the interval $(0, 1)$; then the Metropolis-Hastings sampling algorithm can be described as follows:

- **Step 0**. Choose an arbitrary starting point $\boldsymbol{\theta}^0$, and set $i = 0$;
- **Step 1**. Generate a candidate point $\boldsymbol{\theta}^*$ from $q(\boldsymbol{\theta}^i, .)$ and u from Uniform$(0, 1)$;
- **Step 2**. Set $\boldsymbol{\theta}^{i+1} = \boldsymbol{\theta}^*$ if $u \leq a(\boldsymbol{\theta}^i, \boldsymbol{\theta}^*)$ and $\boldsymbol{\theta}^{i+1} = \boldsymbol{\theta}^i$ otherwise, where the acceptance probability is given by:

$$a(\boldsymbol{\theta}^i, \boldsymbol{\theta}^*) = \min\left(\frac{\pi(\boldsymbol{\theta}^*|D)q(\boldsymbol{\theta}^*, \boldsymbol{\theta}^i)}{\pi(\boldsymbol{\theta}^i|D)q(\boldsymbol{\theta}^i, \boldsymbol{\theta}^*)}, 1\right) \tag{2.1.1}$$

- **Step 3**. Set $i = i + 1$ and go to **Step 1**.

This algorithms is stated in a very general form. The Metropolis-Hastings algorithm can be reduced to the independent chain Metropolis algorithm if $q(\boldsymbol{\theta}, \boldsymbol{\phi}) = q(\boldsymbol{\phi})$ (Tierney, 1994). Another interesting situation arises when $q(\boldsymbol{\theta}, \boldsymbol{\phi}) = q_1(\boldsymbol{\phi} - \boldsymbol{\theta})$, where $q_1(\cdot)$ is the multivariate density and the candidate $\boldsymbol{\theta}^*$ in **Step 2** is drawn according to the process $\boldsymbol{\theta}^* = \boldsymbol{\theta} + \boldsymbol{\omega}$. Here $\boldsymbol{\omega}$ represents the increment random variable and follows the distribution $q_1(\cdot)$. This case is often referred to as a random walk chain (Chib and Greenberg, 1995).

2.2 Approximate Bayesian Inference through INLA

In the case when no analytical form of the posterior distributions is available, the MCMC framework is useful. Generalized dynamic models are known to be a complex class of models in terms of dependence between different effects and states. The main sources for dependence arise from the time evolution of the state equation and the possible association within components of the response vector. It is well known that MCMC methods tend to have slow rate of convergence of the sampling scheme for such complicated problems. Extensive research was done to improve the performance of MCMC (Gamerman, 1997; Knorr-Held and Rue, 2002; Holmes and Held, 2006; Frühwirth-Schnatter and Wagner, 2006; Frühwirth-Schnatter and Frühwirth, 2007; Rue and Held, 2005). Nevertheless the construction of fast and accurate methods for the MCMC algorithms remains cumbersome and time consuming. The Integrated Nested Laplace approximations were proposed by Rue et al. (2009) to provide fast approximate Bayesian inference. INLAs provide accurate approximations to the posterior distributions of the parameters. Moreover, because INLA does not rely on the multiple sampling scheme, these approximations greatly reduce computational time. We give a brief overview of the INLA approach. For more details refer to Rue et al. (2009).

The INLA approach is usually discussed with respect to the structured additive regression models, or latent Gaussian models; see Fahrmeir and Lang (2001). Under this setup, the response variable y_t is assumed to belong to an exponential family and is observed over time. The mean μ_t is attached to a structural additive predictor η_t through a link function, i.e., $l(\mu_t) = \eta_t$. The simplest form of η_t in the case of

dynamic models can be written as follows:

$$\eta_t = \alpha + \gamma_t + \boldsymbol{z}_t'\boldsymbol{\beta}, \tag{2.2.1}$$

where γ_t introduces a temporal dependence in the model, α denotes an intercept, and $\boldsymbol{\beta}$ corresponds to the linear effect of covariates $\boldsymbol{z}$. Let $\boldsymbol{x}$ denote the vector of all the latent Gaussian variables, and $\boldsymbol{\theta}$ denote all the hyperparameters associated with a model (not necessary Gaussian). Most of the latent Gaussian models discussed in the literature are assumed to satisfy two properties. The latent field $\boldsymbol{x}$ is assumed to have conditional independence properties and the number of hyperparameters is relatively small (≤ 6) (Rue and Martino, 2007).

The marginal posterior distribution can be written in the following form:

$$\pi(x_i|\boldsymbol{y}) = \int_{\boldsymbol{\theta}} \pi(x_i|\boldsymbol{\theta}, \boldsymbol{y})\pi(\boldsymbol{\theta}|\boldsymbol{y})\boldsymbol{\theta}, \tag{2.2.2}$$

where x_i denotes each component of the latent Gaussian field $\boldsymbol{x}$, $\boldsymbol{\theta}$ denotes the vector of hyperparameters, and $\boldsymbol{y}$ is an observed data vector. Using the hierarchical structure of the joint distribution, we can rewrite $\pi(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y}) = \pi(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})\pi(\boldsymbol{\theta}|\boldsymbol{y})\pi(\boldsymbol{y})$. Then, $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ can be approximated by the Laplace approximation of a marginal posterior distribution.

$$\widetilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y}) \propto \left.\frac{\pi(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y})}{\widetilde{\pi}_G(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})}\right|_{\boldsymbol{x}=\boldsymbol{x}^*(\boldsymbol{\theta})}, \tag{2.2.3}$$

where $\boldsymbol{x}^*$ denotes the mode of the full conditional $\pi(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$. In (2.2.3) $\widetilde{\pi}_G(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$ denotes the Gaussian approximation to $\pi(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$ (Rue and Held, 2005). To integrate out $\boldsymbol{\theta}$, we need to find a good set of evaluation points θ_k for numerical integration in

(2.2.2). In order to do this we need to explore the properties of (2.2.3). In general, this is done by an iterative algorithm with appropriate choice of weights Δ_k, which are assigned to each θ_k (Rue et al., 2009).

Another part that needs to be approximated is $\pi(x_i|\boldsymbol{\theta}, \boldsymbol{y})$. According to Rue et al. (2009) and Rue and Martino (2007), there are three alternatives: a Laplace approximation, a simplified Laplace approximation and a Gaussian approximation (the simplest one). The non-normal distribution under this alternative is approximated with a Gaussian density by matching the mode and the curvature at the mode (Rue and Held, 2005). Overall, the method gives reasonable results, but the approximation can be improved by applying the Laplace or simplified Laplace approximation to $\pi(x_i|\boldsymbol{\theta}, \boldsymbol{y})$. To summarize, an approximation of the posterior marginal density (2.2.2) can be obtained by numerical integration as:

$$\widetilde{\pi}(x_i|\boldsymbol{y}) = \sum_k \widetilde{\pi}(x_i|\theta_k, \boldsymbol{y})\widetilde{\pi}(\theta_k|\boldsymbol{y})\Delta_k. \tag{2.2.4}$$

The choice of the integration points θ_k can be done using either the grid strategy (GRID) or the central composite design strategy (CCD) (for details see Rue and Martino (2007)). Thus, the approximate posterior quantities can be obtained and used as posterior summaries for the parameters of interest.

Chapter 3

Dynamic Compositional Modeling of Pedestrian Crash Counts on Urban Roads in Connecticut

Uncovering the temporal trend in crash counts provides a good understanding of the context for pedestrian safety. Since pedestrian crashes are rare events, it is impossible to investigate monthly temporal effects with individual segment/intersection level data. We study the time dependence based on data that has been suitably aggregated across road segments (as described below). Most previous studies have used annual data to investigate the differences in pedestrian crashes between different regions or countries in a given year, and/or to look at time trends of fatal pedestrian injuries annually. Use of annual data unfortunately does not provide sufficient information on patterns in time trends or seasonal effects. This chapter describes statistical methods for uncovering patterns in monthly pedestrian crashes aggregated on urban roads in Connecticut from January 1995 to December 2009. We investigate the temporal be-

havior of injury severity levels, including fatal (K), severe injury (A), evident minor injury (B), and non-evident possible injury and property damage only (C and O), as proportions of all pedestrian crashes in each month, taking into consideration effects of time trend, seasonal variations and VMT (vehicle miles traveled). This type of dependent multivariate data is characterized by positive components which sum to one, and occurs in several applications in science and engineering. We describe a dynamic framework with vector autoregressions (VAR) for modeling and predicting compositional time series. Combining these predictions with predictions from a univariate statistical model for total crash counts will then enable us to predict pedestrian crash counts with the different injury severity levels. We compare these predictions with those obtained from fitting separate univariate models to time series of crash counts at each injury severity level. We also show that the dynamic models perform better than the corresponding static models. We implement the Integrated Nested Laplace Approximation (INLA) approach to enable fast Bayesian posterior computation.

3.1 Motivation

The economic and societal losses due to motor vehicle crashes in the USA exceed US$870 billion, nearly $900 per capita based on calendar year 2010 data (NHTSA, 2014). In 2013, 32,719 people died and 2.3 million people were injured in motor vehicle crashes in the United States (IIHS, 2014). The situation is of particular interest on limited access highways, which experience significantly higher fatality rates than local and collector roads. About 60 percent of total 33,561 fatalities during 2012 occur on limited access highways.

Various studies have been performed to identify factors which affect pedestrian crashes and severity. Many factors contribute to the frequency and severity of pedestrian crashes and conflicts (Pasanen and Salmivaara, 1993; Garber and Lineau, 1996; Jensen, 1999; Klop and Khattak, 1999; LaScala et al., 2000). For example, Garber and Lineau (1996) found that the age of the pedestrian, location of the crash, the type of facility, the use of alcohol, and the type of traffic control at the site are associated with pedestrian conflicts and the likelihood of severe injury in motor vehicle crashes. This same study also found that pedestrian involvement rates are significantly higher at locations within 150 feet of an intersection stop line. Zajac and Ivan (2003) found similar results for roadway features and pedestrian characteristics having significant correlation with pedestrian injury severity from their study on rural Connecticut state-maintained highways. In addition, they also studied influence of area features on pedestrian injury severity and found that villages, downtown fringe, and low-density residential areas tend to experience higher pedestrian injury severity than downtown, compact residential, and medium- and low-density commercial areas. As one would expect, vehicle speed is seen as a significant contributor to crash severity. According to a mathematical model, a speed of 50 km/hour increases the risk of death almost eight-fold compared to a speed of 30 km/hour. Crash environment also affects crash severity as Klop and Khattak (1999) found that rain, fog, or snow as well as dark environment increases injury severity.

Along with the study on factors affecting pedestrian crash and its severity, also understanding the crash trend provides a good insight into the magnitude of the pedestrian crash problem. In a study of pedestrian crash trends from around the world, Zegeer and Bushell (2012) collected pedestrian safety statistics at the global, regional, and national levels, and studied driver factors, roadway factors, vehicle

factors, demographic factors, and pedestrian factors which affect the risk and/or severity of a pedestrian crash. They presented lessons for improving pedestrian safety learned from several countries, especially in Europe, and from USA. A few of the pedestrian safety strategies that they mentioned were to provide pedestrian-friendly geometric guidelines, implement effective traffic control and other pedestrian safety treatments, expand funding for safety education programs, and develop pedestrian friendly vehicle features.

Spainhour et al. (2006) studied pedestrian crash trends and causative factors in Florida. The paper focused on finding primary contributing factors for pedestrian crashes, and concluded that pedestrians were at fault in 78 percent of the cases reviewed. However, they did not study pedestrian crashes by severity of injuries. Rather they categorized pedestrian crashes as follows: crossing not in a crosswalk, crossing at intersection, in road, walking along roadway (with traffic), walking along roadway (against traffic), exit vehicle, vehicle turn or merge, unique (crashes with some unusual circumstances which are not likely to happen again) and other (crashes with unknown circumstances). Hu et al. (2012) introduced dynamic time series modeling in a Bayesian framework to uncover temporal patterns in the safety of senior and non-senior drivers in Connecticut.

The objective of our study is to discover temporal changes in pedestrian crashes with a particular injury severity level as proportions of total crashes of all injury levels. In other words, our goal is to investigate whether over the given time period, there is an increase in crashes of one or more injury severity levels with attendant decreases in other levels. We collected records for all crashes on state-maintained roads in Connecticut from January 1995 to December 2009 from the Connecticut Department of Transportation (ConnDOT). Crashes were classified into the following

severity groups: K= fatal injury, A = severe injury, B = evident minor injury, C = non-evident minor injury and O = property damage only.

The main characteristic of compositional data, which occurs frequently in various areas such as chemistry, demography, geology, survey analysis, consumer demand analysis, etc., is that at each time point, all components are positive and sum to one. There is a need to model different proportions or compositions that are observed over time, i.e., to model temporal changes of such compositional time series using suitable models. A compositional time series is defined as a J-variate vector of positive components $\boldsymbol{x}_t = (X_{t1}, \ldots, X_{tJ})$, for $t = 1, \ldots, T$, where the structure is completely defined by $g = J - 1$ components, so that $\boldsymbol{x}_t$ lies in a g-dimensional simplex:

$$S^g\{(X_{t1}, \ldots, X_{tJ}) : X_{t1} > 0, \ldots, X_{tJ} > 0; X_{t1} + \cdots + X_{tJ} = 1\}$$

Statistical analysis follows via a suitable transformation of the data from the g-dimensional simplex S^g into the Euclidean space R^g. An excellent approach for compositional data analysis is given by Aitchison (1982) and Aitchison (1986), who introduced the Additive Log Ratio (ALR) transformation and the Centered Log Ratio (CLR) transformation, by Rayens and Srinivasan (1991) who discussed the more general Box-Cox (Box and Cox (1964)) transformation, and by Egozcue et al. (2003) who proposed the Isometric Log Ratio transformation. Aitchison (1986) along with Brunsdon (1987), Smith and Brunsdon (1989), Brunsdon and Smith (1998), and Ravishanker et al. (2001) discussed compositional time series analysis. In these papers, compositional time series were first transformed via the ALR (or more general Box-Cox) transformation, and were then analyzed with standard time series model techniques, such as Vector AutoRegression (VAR), Vector AutoRegressive Moving

Average (*VARMA*), or Dynamic Linear Modeling via the Kalman Filter, or in a Bayesian framework.

The study design and the compositional time series modeling described in Section 3.2 enables us to model transformed crash proportions of different injury severity levels in order to discuss changes and connections among them. Section 3.3 describes a dynamic modeling framework for the time series of total pedestrian crash counts and compares it with a static model. Section 3.4 describes predictions of proportions and total counts which enables us to obtain predictions of the crash counts by injury severity levels. Section 3.5 provides a summary and discussion.

3.2 Dynamic Compositional Time Series Modeling

3.2.1 Study Design

Crash records of State-maintained roads are recorded and preserved by Connecticut Department of Transportation (ConnDOT). Crash data from January 1995 to December 2009 from ConnDOT was used. Property damage only (PDO) crashes were not reported in the database before 2007 for local roads. PDOs were reported starting in 2007. Crash data were aggregated by each month at five different severity level e.g. fatal (K), severe injury (A), evident minor injury (B), non-evident possible injury (C) and property damage only (O). The C and O severity levels were combined into one response variable for analysis because the O level is rare for pedestrian crashes, while the other severity levels were each defined as individual response variables.

Vehicle-miles-traveled (VMT) was used as a predictor variable in compositional modeling. For this purpose we needed monthly VMTs during the analysis period.

ConnDOT has average daily VMT for each year for various definitions of facility type based on urban or rural location and functional classification. Also ConnDOT has pneumatic tube and induction loop counters from which these annual average daily VMT estimates are derived. To obtain monthly VMTs, monthly expansion factors obtained from ConnDOT were used. Descriptive statistics of the response variables and VMT used in the analysis are given in Table 3.2.1.

TABLE 3.2.1: Descriptive Statistics of Pedestrian Crash Counts and VMT

Year	Total(KABCO)		K		A		B		CO		VMT	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
1995	37.5	6.65	1.92	1.38	10.5	3.5	11.42	1.88	13.67	4.42	64904	3799
1996	34.41	6.02	1.42	0.67	10.42	2.94	9.92	4.93	12.67	2.61	65450	3865
1997	37.17	9.58	1.92	1.31	9.58	4.1	12.08	3.78	13.58	3.68	66687	3894
1998	34.92	8.18	1.67	1.3	8	2.37	11.92	4.38	13.33	3.98	68810	4118
1999	35	8.66	1.83	1.27	8.5	3.63	12.42	4.01	12.25	5.03	70987	3820
2000	33	7.92	1.58	0.79	7.08	1.98	13.17	3.83	11.17	5.06	72639	3936
2001	33.75	7.82	1.67	1.23	7.08	3.58	12.83	3.04	12.17	3.69	73219	4056
2002	35.92	9.14	2.67	2.67	7.25	3.86	13.08	4.06	12.92	4.27	74690	4033
2003	35.17	10.4	1.83	0.94	6.5	3	12.5	5.32	14.33	4.5	74947	4229
2004	29.75	10.63	1.08	0.79	5.33	2.84	12.08	5.82	11.25	4.52	75245	4173
2005	30.08	10.77	1.75	1.22	5.33	3.17	10.83	4.99	12.17	4.39	75490	4593
2006	33.92	7.48	1.42	0.9	6.42	2.02	14.17	4.82	11.92	3.48	76055	4990
2007	35.5	6.4	1.75	1.66	7.42	4.19	13.17	2.62	13.17	4.3	76966	4141
2008	33.83	10.21	1.92	1.16	6.5	3.17	14.08	4.21	11.33	5.42	75405	3747
2009	31.17	6.64	1.17	1.11	6	2.76	12.42	3.29	11.58	4.87	75308	3818

3.2.2 Model Framework

In this section, we model compositional data which is represented by proportions of $J = 4$ different severity levels (K, A, B, CO) of monthly pedestrian crashes in Connecticut between 1995 and 2009. We use the Box-Cox transformation together with VAR techniques, carry out the data analysis, and highlight the important findings with respect to the interpretation of the results and the estimation of the coefficients. Given crash counts in injury severity categories K, A, B, and CO, we form the proportions X_{ti} by dividing the counts in each category by the total number of pedestrian

crashes, i.e,

$$X_{ti} = \frac{C_{ti}}{C_{t,Total}}, \tag{3.2.1}$$

where C_{ti} represents the count of pedestrian crashes for severity category i (K, A, B, CO) during month t ($t = 1, \ldots, T$), where $T = 180$, so that $C_{t,Total}$ represents the total count of pedestrian crashes during month t. The ALR transformation has the form $Y_{ti} = \log(X_{ti}/X_{tJ})$, for $i = 1, \ldots, g$, X_{tJ} being a reference (baseline) component for $t = 1, \ldots, T$, and "log" denotes the natural logarithm. Dividing by a reference (baseline) component takes care of the dimensionality problem, while taking the logarithm is intended to transform the data to normality. In the following applications $J = 4$ and $g = 3$. We use the Box-Cox transformation with a small adjustment d to avoid issues with logarithms; for $t = 1, \ldots, 180$, $i =$"K","A","B", and $J =$"CO", and $d = 10^{-5}$ (different values of d were tried and did not give different results).

$$Y_{ti} = \begin{cases} \frac{\left(\frac{X_{ti}}{X_{tJ}}+d\right)^{\lambda}-1}{\lambda} & \text{if } \lambda \neq 0 \\ \log\left(\frac{X_{ti}}{X_{tJ}}+d\right) & \text{if } \lambda = 0 \end{cases} \tag{3.2.2}$$

The next step is to fit a suitable multivariate model to the transformed data $\boldsymbol{y}_t = (Y_{tK}, Y_{tA}, Y_{tB})'$ whose components are real-valued. We fit a vector autoregressive model of order p ($VAR(p)$) with regressors (Reinsel, 1993):

$$\boldsymbol{y}_t = \sum_{k=1}^{p} \boldsymbol{\Phi}_k \boldsymbol{y}_{t-k} + \boldsymbol{\eta} u_t + \boldsymbol{\gamma} t + \boldsymbol{S\delta} + \boldsymbol{w}_t, \tag{3.2.3}$$

where $\boldsymbol{\Phi}_k$ is a 3×3 AR (autoregressive) matrix of lag k, for $k = 1, \ldots, p$; $\boldsymbol{\eta}$ is a 3-dimensional parameter vector; u_t is an exogenous variable ($\log(VMT)$); $\boldsymbol{\gamma}$ is a vector

of coefficients corresponding to the time trend; $\boldsymbol{S}$ is a 3×12 parameter matrix for the seasonal part; $\boldsymbol{\delta}$ is an indicator vector corresponding to the seasonal part, and $\boldsymbol{w}_t$ is a 3-dimensional vector of i.i.d. Normal$(\mathbf{0}, \boldsymbol{\Sigma})$ errors.

3.2.3 Model Estimation

The first step in the compositional analysis is to choose an appropriate value of the parameter λ in the Box-Cox transformation, as well as the order p for the *VAR* model. We use the Mean Absolute Error (MAE) criterion throughout the chapter for predictive validation and model selection. We calculate the within sample MAE as well as out-of-sample MAE using 6 and 12 months of holdout data using the formula. Recall that we have 4-variate counts time series data, namely number of pedestrian crashes with K, A, B and CO injury severities, for 180 months. We hold out the last $L = 6$ months or $L = 12$ months for predictive validation, and respectively use first $180 - L$, i.e., 174 or 168 observations for model fitting/calibration. The MAE is defined as

$$\text{MAE} = \frac{1}{T} \sum_{i=1}^{T} |y_i - \widehat{y}_i|, \tag{3.2.4}$$

where y_i denotes the ith observed value, and $\widehat{y}_i$ is the ith fitted or predicted value under the particular model.

We iterate λ from -2 to 2 in steps of 0.1 and record the within sample MAE for all severity levels together for the best VAR(p) model for $1 \leq p \leq 10$; the best VAR(p) model corresponds to the fitted model with lowest Akaike Information Criterion (AIC) (Akaike, 1974). That is, we compute MAE_{ALL} based on the proportions data which

we obtain by transforming back the Box-Cox transformed compositions:

$$\mathrm{MAE_{ALL}} = \mathrm{MAE_K} + \mathrm{MAE_A} + \mathrm{MAE_B} + \mathrm{MAE_{CO}} \tag{3.2.5}$$

The smallest $\mathrm{MAE_{ALL}}$ is 0.2189 with VAR of order $p = 1$ (the lowest AIC is -6.5796). We use the function *VAR* in the R package *vars* for the VAR(p) model fitting and the function *bxcx* in the R package *FitAR* for doing the Box-Cox transformation.

The fitted model for the transformed compositions corresponding to injury severity level K follows from (3.2.3) and is written as

$$\widehat{Y}_{t,K} = \widehat{\phi}_{11,K} Y_{t-1,K} + \widehat{\phi}_{12,A} Y_{t-1,A} + \widehat{\phi}_{13,B} Y_{t-1,B} + \widehat{\eta}_K u_t + \widehat{\gamma}_K t + \widehat{\boldsymbol{S}}'_K \boldsymbol{\delta}, \tag{3.2.6}$$

where $(\widehat{\phi}_{11,K}, \widehat{\phi}_{12,A}, \widehat{\phi}_{13,B})$ is the first row of estimated matrix $\boldsymbol{\Phi}$ and $\widehat{\boldsymbol{S}}_K$ is the first row of $\boldsymbol{S}$; $\widehat{\gamma}_K$ and $\widehat{\eta}_K$ are respectively the first components of the estimated vectors $\boldsymbol{\gamma}$ and $\boldsymbol{\eta}$. Partial output for the fitted coefficients in (3.2.6) is shown in the top portion of Table 3.2.2, in rows $1-5$. Note that the seasonal coefficients are suppressed from the output for brevity. March, July, August and November coefficients have the smallest p-values, which are respectively 0.0708, 0.1171, 0.0294 and 0.0871; and the corresponding estimated values of the seasonal effects, with standard errors shown in brackets, are respectively $-0.1538(0.0845)$, $-0.1325(0.0841)$, $-0.1884(0.0856)$, and $-0.1458(0.0847)$.

The fitted model for the transformed compositions corresponding to injury severity level A follows from (3.2.3) and is written as

TABLE 3.2.2: Estimated Coefficients for level K, A and B Transformed Compositions

(* denotes significant p-values with significance level $\alpha = 0.1$)

Parameter	Estimate	Std. Error	t value	Pr($>|t|$)
$\widehat{\phi}_{11,K}$	-0.0053	0.0819	-0.065	0.9481
$\widehat{\phi}_{12,A}$	0.0990	0.0507	1.952	0.0528*
$\widehat{\phi}_{13,B}$	-0.0019	0.0418	-0.047	0.9629
$\widehat{\gamma}_K$	0.0008	0.0004	1.954	0.0526*
$\widehat{\eta}_K$	-0.2597	0.0226	-11.504	<0.0001*
$\widehat{\phi}_{21,K}$	-0.2720	0.1525	-1.784	0.0765*
$\widehat{\phi}_{22,A}$	0.0354	0.0945	0.374	0.7089
$\widehat{\phi}_{23,B}$	0.0198	0.0778	0.255	0.7990
$\widehat{\gamma}_A$	-0.0019	0.0007	-2.556	0.0116*
$\widehat{\eta}_A$	-0.1213	0.0420	-2.884	0.0045*
$\widehat{\phi}_{31,K}$	-0.4279	0.1833	-2.334	0.0209*
$\widehat{\phi}_{32,A}$	0.1728	0.1136	1.521	0.1303
$\widehat{\phi}_{33,B}$	-0.0208	0.0935	-0.223	0.8241
$\widehat{\gamma}_B$	0.0029	0.0009	3.345	0.001*
$\widehat{\eta}_B$	-0.1344	0.0505	-2.659	0.0087*

$$\widehat{Y}_{t,A} = \widehat{\phi}_{21,K} Y_{t-1,K} + \widehat{\phi}_{22,A} Y_{t-1,A} + \widehat{\phi}_{23,B} Y_{t-1,B} + \widehat{\eta}_A u_t + \widehat{\gamma}_A t + \widehat{\boldsymbol{S}}'_A \boldsymbol{\delta}, \qquad (3.2.7)$$

where $(\widehat{\phi}_{21,K}, \widehat{\phi}_{22,A}, \widehat{\phi}_{23,B})$ is the second row of estimated matrix $\boldsymbol{\Phi}$, $\widehat{\boldsymbol{S}}_A$ is the second row of $\boldsymbol{S}$; $\widehat{\gamma}_A$ and $\widehat{\eta}_A$ are the second components of the estimated vectors $\boldsymbol{\gamma}$ and $\boldsymbol{\eta}$ respectively. Partial output for the fitted coefficients in (3.2.7) is shown in the middle portion of Table 3.2.2, in rows $6 - 10$. Again, results on the seasonal effects are not shown. Only January has a relatively small p-value of 0.1118, with corresponding estimated value (standard error) of 0.2566 (0.1604).

The fitted model for the transformed compositions corresponding to injury severity

level B follows from (3.2.3) and is written as

$$\widehat{Y}_{t,B} = \widehat{\phi}_{31,K} Y_{t-1,K} + \widehat{\phi}_{32,A} Y_{t-1,A} + \widehat{\phi}_{33,B} Y_{t-1,B} + \widehat{\eta}_B u_t + \widehat{\gamma}_B t + \widehat{\boldsymbol{S}}'_B \boldsymbol{\delta}, \qquad (3.2.8)$$

where $(\widehat{\phi}_{31,K}, \widehat{\phi}_{32,A}, \widehat{\phi}_{33,B})$ is the third row of estimated matrix $\boldsymbol{\Phi}$, $\widehat{\boldsymbol{S}}_B$ is the third row of $\boldsymbol{S}$; $\widehat{\gamma}_B$ and $\widehat{\eta}_B$ are the third components of the estimated vectors $\boldsymbol{\gamma}$ and $\boldsymbol{\eta}$ respectively. Partial output for the fitted coefficients in (3.2.8) is shown in the bottom portion of Table 3.2.2, in rows $11 - 15$. The seasonal coefficients for February and November have the smallest p-values of 0.132 and 0.1057 with corresponding estimated values (standard errors) of 0.2901 (0.1916) and -0.3086 (0.1896).

The estimated covariance matrix of residuals from (3.2.3) is

$$\begin{pmatrix} 0.0441 & 0.0155 & 0.0176 \\ 0.0155 & 0.153 & 0.0908 \\ 0.0176 & 0.0908 & 0.221 \end{pmatrix}$$

The matrix of residuals is an estimate of the covariance matrix $\boldsymbol{\Sigma}$ of the error term $\boldsymbol{w}_t$ in equation (3.2.3). Small off-diagonal values indicate almost no correlation between elements of vector $\boldsymbol{y_t}$. The values on the diagonal represent the estimated variance of proportions K, A and B over the baseline CO transformed by (3.2.2). The fitted model indicates that the proportion of B level crashes over the proportion of CO level crashes has the largest residual variance, followed by the proportion of A level and K level crashes over the CO level.

We ran standard diagnostics procedures to assess out model fit. Cross-correlation plots of residuals and the adjusted multivariate Portmanteau test (Hosking, 1980) indicated that the fitted models were adequate. Normal Q-Q plots of the residuals

indicated that the assumption of normality is reasonable.

3.2.4 Discussion of Results

We have modeled the data Y_{ti} which is proportional to the ratio of proportions of crashes of K, A or B levels over the CO level. Thus, a decrease in Y_{ti} is desirable in terms of transportation safety as it indicates a smaller proportion of crashes for more severe injury levels over the CO (the least severe) level. All three estimates of the elements of the vector $\boldsymbol{\gamma}$ are significant, which suggests a strong indication of a time trend for all three components. There is a substantial increase in the proportion of B level crashes over time, with $\widehat{\gamma}_B = 0.0029$ (0.0009) with respect to the baseline CO level. Coefficients also indicate a somewhat smaller decrease of severity A level crashes with $\widehat{\gamma}_A = -0.0019$ (0.0007), and just a slight increase of severity K level crashes with $\widehat{\gamma}_K = 0.0008$ (0.0004), all relative to the baseline CO level. All three estimated components of the $\boldsymbol{\eta}$ vector are significant and negative, which indicates that an increase in VMT results in a decrease in the proportions of all other crash levels (K, A and B) with respect to the baseline CO level. All significant seasonal coefficients are negative except in January for A level crashes and in February for B level crashes, which indicates an increase in those two months for A and B severity level crashes over the baseline CO level. Estimated entries of the $\boldsymbol{\Phi}$ matrix represent the serial correlation between transformed compositions at one time lag.

To visualize results from the compositional time series analysis, we show stacked plots for different injury severity proportions of pedestrian crashes in Figure 3.2.1. Proportions of K, A, B and CO injury severity levels are obtained from the fitted model (3.2.3) using the inverse transformation from (3.2.2). The solid lines at the

bottom, middle, and top of the plot denote respectively the fitted proportions of level K, level (K+A) and level (K+A+B) crashes. The dotted lines at the bottom, middle and top denote respectively the corresponding observed proportions at levels K, (K+A) and (K+A+B). The stacked plot is done to avoid the overlapping of fitted and observed proportions and is more meaningful as a graphical representation.

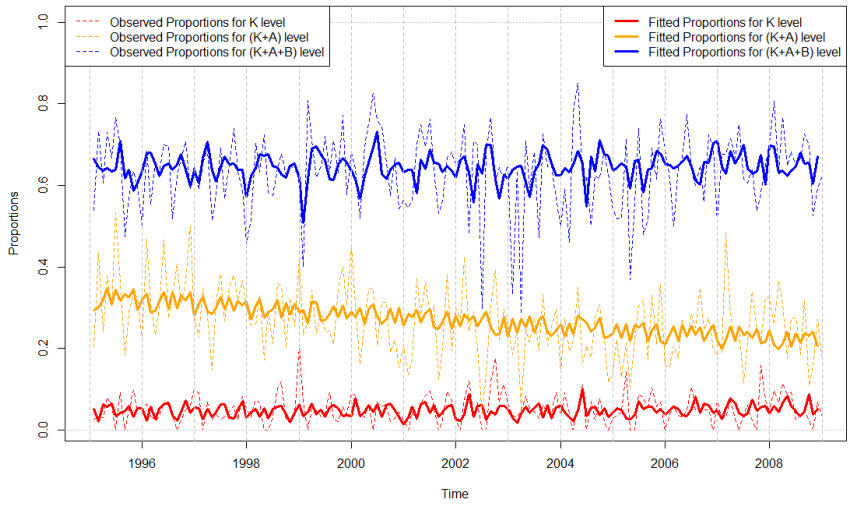

FIGURE 3.2.1: Stacked Plot of Fitted Proportions of Compositional Model

From Figure 3.2.1, we note that an overall model fit to the data is satisfactory and underline the conclusions we had made based on the estimated coefficients. We have managed to get rid of unwanted noise and extract a clear behavior of the compositional time series data. We have also verified our conclusions about time trend in the data. There is a noticeable significant decrease in the proportion of severity level A crashes over time, with an increase in pedestrian crashes with severity level B, while K and CO stay approximately on the same level over this time period. In other words, there

is an apparent migration of crashes from the higher severity level A to lower severity level B, suggesting that pedestrian injury severity is on average decreasing over time.

3.3 Dynamic and Static Modeling of Pedestrian Crash Counts

In this section, we describe the dynamic and the static modeling of the total pedestrian crash counts, aggregated over all injury severity levels. Hu et al. (2012) described the estimation of such models and the predictions for all vehicle crashes. We extend this analysis to the pedestrian crashes.

We fit a dynamic model for the total pedestrian crash counts $C_{t,Total}$ for notational simplicity, we write this simply as C_t in this section. The observation equation is written either using a conditional Poisson distribution in (3.3.1a) or a conditional negative binomial distribution in (3.3.1b):

$$C_t|\lambda_t \sim \text{Poisson}(\lambda_t) \tag{3.3.1a}$$

$$C_t|\lambda_t, k \sim \text{NegBin}(\lambda_t, k), \tag{3.3.1b}$$

where λ_t is the mean and k is the shape parameter; we model the latent variable λ_t as

$$\lambda_t = \alpha_t \times V_t^{\eta} \times \exp(S_t)$$

or

$$\log(\lambda_t) = \log(\alpha_t) + \eta \times \log(V_t) + \log(S_t) \tag{3.3.2}$$

Define $\beta_t = \log(\alpha_t)$. The trend β_t and seasonal components S_t are modeled as

$$\beta_t = \rho \times \beta_{t-1} + w_t \qquad t = 2, \ldots, 180 \tag{3.3.3a}$$

$$S_t = -(S_{t-1} + \cdots + S_{t-11}) + w_t \qquad t = 12, \ldots, 180, \tag{3.3.3b}$$

where β_t is the dynamic intercept coefficient; V_t represents VMT; η represents the static coefficient for log(VMT); ρ represents the coefficient in the autoregressive process ($AR(1)$ process); and S_t is the seasonal component to explain periodic behavior of period 12; w_t is i.i.d. Normal$(0, a^2)$; v_t is i.i.d. Normal$(0, b^2)$. We refer to this model as a Dynamic Generalized Linear Model (DGLM). The response variable is a count variable and since its conditional distribution is non-Gaussian, the standard Kalman Filter algorithm cannot be used to obtain the posterior distributions of the unknown parameters, which do not have closed forms. Gamerman (1998) suggested a fully Bayesian approach for DGLMs using Markov chain Monte Carlo techniques This approach can be computationally intensive, and we therefore implement approximate Bayesian inference using INLA approach proposed by Rue et al. (2009). The dynamic model described above enables us to fit and predict total pedestrian crash counts. We implement this method using the R package INLA available through webpage www.r-inla.org. Ruiz-Cárdenas et al. (2012) give an excellent guidance for fitting dynamic models using the R-INLA package, which includes numerous simulated and real data examples.

The static model is widely used, and is defined via (3.3.1a) or (3.3.1b), where now

$$\begin{aligned} \lambda_t &= \alpha \times V_t^{\eta} \times \exp(\gamma \times t + \boldsymbol{s}' \times \boldsymbol{\delta}) \\ &\text{or} \\ \log(\lambda_t) &= \log(\alpha) + \eta \times \log(V_t) + \gamma \times t + \boldsymbol{s}' \times \boldsymbol{\delta} \end{aligned} \tag{3.3.4}$$

Here, $\log(\alpha)$ is static intercept coefficient; η represents the static coefficient for log(VMT); γ represents time trend coefficients; $\boldsymbol{s}'$ is a 1×11 parameter vector corresponding to the seasonal part; $\boldsymbol{\delta}$ is 11×1 an indicator vector corresponding to the seasonal part and other terms are defined earlier under the dynamic model.

We fitted the negative binomial model and the Poisson model for C_t for both the static and dynamic models. The AIC criterion for the negative binomial static model was 1125.4 which was slightly larger than the AIC value of 1125.3 for the Poisson distribution static model (the overdispersion coefficient was also not significant). The Deviance Information Criterion (DIC) value was 1175.13 for the negative binomial dynamic model, which was higher than the value of 1121.3 in the Poisson case. The smaller values of AIC and DIC indicate the better models, therefore we present results from the Poisson static and dynamic models. In the static model described by (3.3.1a) and (3.3.4), the intercept α (which also represents the coefficient for January in the model with seasonal indicator variables), as well as the seasonal coefficients from February through August are significant. Estimated values (standard errors) for α, and corresponding seasonal coefficients are 7.4453 (2.4860), -0.2598 (0.0687), -0.3108 (0.0765), -0.2827 (0.0915), -0.2960 (0.1120), -0.3817 (0.1199), -0.3885 (0.1187), -0.2986 (0.1192), -0.1745 (0.1038). Estimated coefficients η and γ (effect of VMT and time) appear to be not significant. Figure 3.3.1 shows a plot of observed

total pedestrian crash counts, together with fitted values from the Poisson static and dynamic models. From the plot, we can conclude that both models fit the observed data well, and they manage to capture the seasonal part for almost all time periods. However, there are some differences in the fit of static and dynamic models between 2004 and 2006 years, which we will explain later.

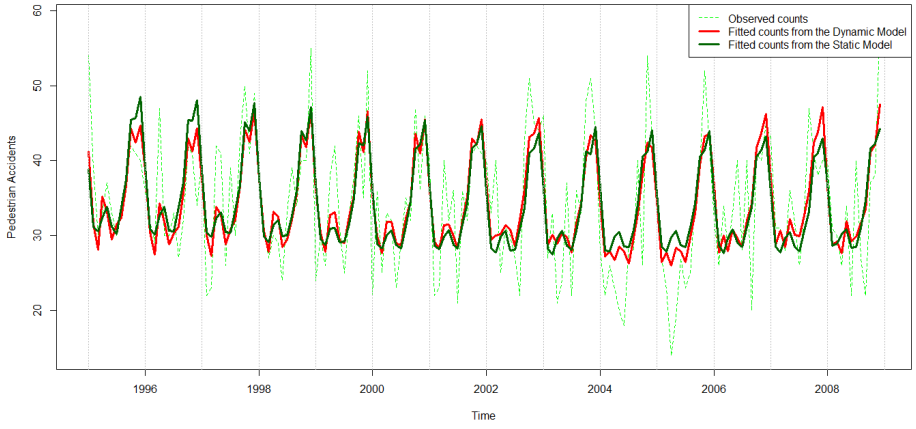

FIGURE 3.3.1: Pedestrian Crash Counts with the Model Fits

The dynamic model in (3.3.1a)-(3.3.3b) can also be used for modeling crash counts C_{ti} by injury severity levels. In order to assess differences in fitted values of static and dynamic models we need to explain the differences in the models itself. The main difference is coming from the way to model time trend coefficient. In the static model it stays the same during all time period, but for the dynamic model we allow to have dynamic intercept (represents time trend coefficient) to vary over the time

which is more natural behavior in real life situations. Figure 3.3.2 gives a plot of estimated dynamic coefficient α_t from the model (3.3.1a)-(3.3.3b) (according to the Bayesian model, it is posterior mean for intercept) for total counts as well as for level A severity injury crashes. The dynamic intercepts for K, B and CO injury severity crash levels have very slight changes in dynamic behavior, staying at the same level ($\alpha_t \approx 1$) throughout the observational period and are omitted in the Figure 3.3.2 to avoid the overlaps.

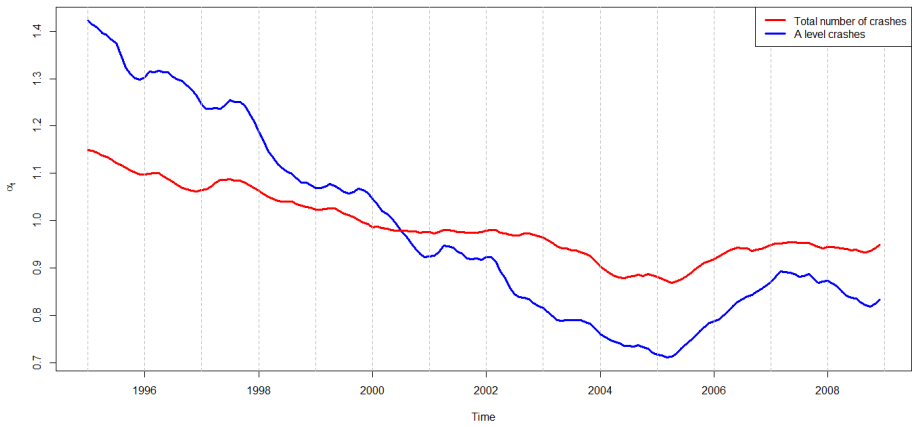

FIGURE 3.3.2: Temporal Behavior of α_t

From Figure 3.3.2, we observe an overall decrease in the total number of pedestrian crashes from 1995 to April 2005, then, an increase until May 2007, and then leveling off until the end of the time period. That is exactly the same period where the dynamic model differs from the static model. The reason is that the static model failed to

catch the dynamic behavior of the trend. We recall that the dynamic models for different injury severity crash counts account for the effect of VMT, seasonal period, and random variation in the counts due to the Poisson distribution. We observe that the main reason for the increase in total counts is an increase in the counts of level A injury severity pedestrian crashes, while counts of the other injury severity crashes do not show changes in the α_t values. In other words the results from the dynamic model show no temporal effects for K, B, or CO severity levels, thus a reduction in level A will result in fewer total crashes.

From Table 3.3.1, we conclude that an increase in VMT results in an increase in total pedestrian crashes as well as in all injury severity level crashes. The effect of the increase is the largest for the CO level crashes followed by B, A and K levels. This result coincides with the finding from the compositional model where the negative sign for VMT suggests the slower increase in K, A and B level crashes than in CO level crashes (the baseline).

TABLE 3.3.1: Estimated static coefficients for the VMT effect in the dynamic models

Parameter η (in models for)	Mean	Std. Dev.	95% Credible Interval	
Total (KABCO)	0.8259	0.0142	0.7979	0.8579
K	0.1180	0.0141	0.0900	0.1452
A	0.4739	0.0386	0.3974	0.5587
B	0.5849	0.0052	0.5746	0.5950
CO	0.5865	0.0052	0.5762	0.5966

The estimated seasonal components are shown in Figure 3.3.3. The total (KABCO) crashes have the lowest peak in March, which migrates to February after 2001. We can notice the intermediate peak around May, followed by an increase in crashes until December. Overall, the seasonal effect is negative from February until August, and stays positive for the remaining months. The seasonal effect for K level crashes is neg-

ative from February until July and in September with the lowest values during March and July and the largest value in December. The A level crashes are influenced by the negative seasonal effect during January-August months. Months January through April and June through September have negative effect on B level crashes. Similar to the total crashes, February through August have negative effects on the CO level crashes.

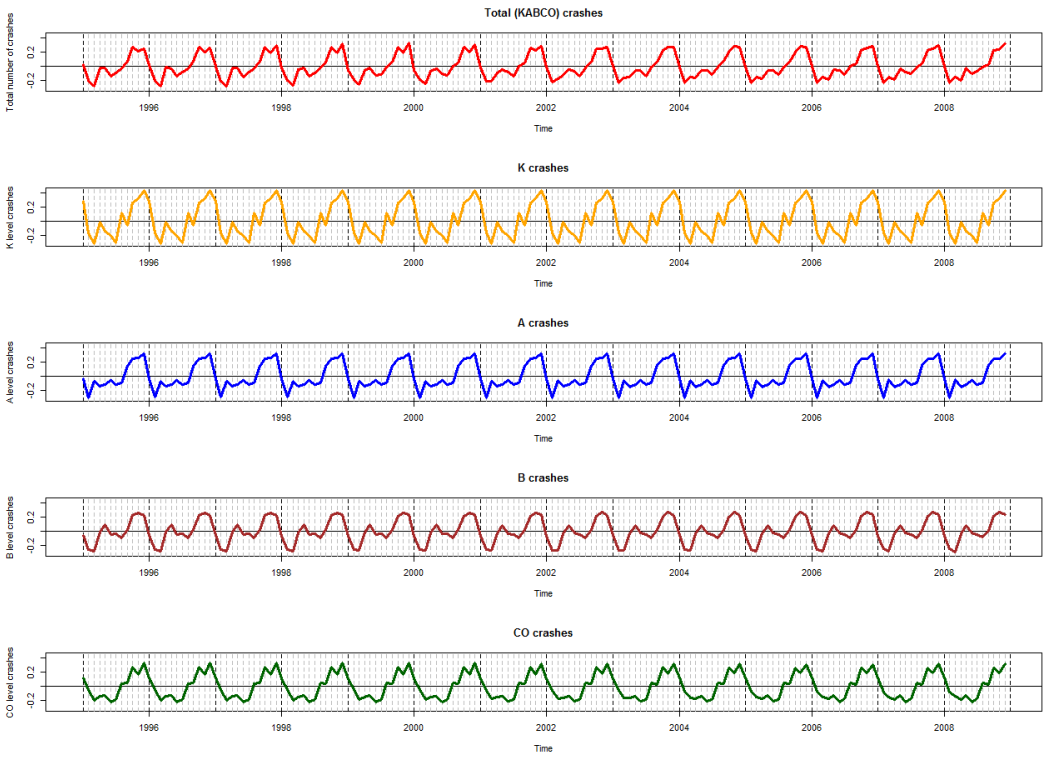

FIGURE 3.3.3: Temporal Behavior of S_t (seasonal component)

3.4 Predictions of Crash counts by Injury Severity Levels

We combine the approaches described in Sections 3.2 and 3.3 in order to construct fits/predictions for the pedestrian crash counts by injury severity levels. We also compare the predictions obtained by using the static or the dynamic models for fitting the total pedestrian crashes. For the fitting portion of the data, we use 168 observations, holding out the last 12 observations for the forecast evaluation. We evaluate 6 month and 12 month ahead predictions.

We obtain fitted or predicted compositions $\widehat{Y}_{ti}$ from the compositional model described in Section 3.2; we convert these to fitted/predicted proportions $\widehat{X}_{ti}$ via (3.2.2), and then use (3.2.1) to convert to the fitted/predicted counts at different severity levels, i.e.,

$$\widehat{C}_{ti} = \widehat{X}_{ti} \times \widehat{C}_{t,Total}$$

Here, $\widehat{C}_{ti}$ represents the fitted counts of pedestrian crashes for severity level i (K,A,B,CO) during month t ($t = 1, \ldots, 180$); $\widehat{C}_{t,Total}$ represents the fitted total counts of pedestrian crashes by static or dynamic models during month t; $\widehat{X}_{ti}$ represents the fitted proportions of pedestrian crashes by compositional model for severity level i (K,A,B,CO) during month t ($t = 1, \ldots, 180$).

Table 3.4.1 shows a comparison between different models in terms of the Mean Absolute Error criterion (MAE) by comparing C_{ti} with $\widehat{C}_{ti}$ using:

$$\text{MAE}_i = \frac{1}{T} \sum_{t=1}^{T} |C_{ti} - \widehat{C}_{ti}|$$

TABLE 3.4.1: MAE comparison

Model		Within sample (fitting portion)	6 months ahead prediction	12 months ahead prediction
Dynamic Setup	K	1.0048	0.7342	0.9790
	A	2.8424	2.4912	2.2976
	B	3.6964	2.9794	2.9172
	CO	2.9075	2.9281	2.2499
Static Setup	K	1.0094	0.7563	0.9703
	A	2.8711	2.4064	2.2467
	B	3.8180	3.0461	2.6980
	CO	3.0773	2.8506	2.3918

The Dynamic Setup and the Static Setup respectively refer to predicting the total crash counts via the dynamic model (3.3.1a)-(3.3.3b), or the static model (3.3.1a) and (3.3.4), while $\widehat{X}_{ti}$ follows from the compositional model in both cases. From Table 3.4.1, we see that the dynamic setup has lower MAE values for all within sample fits, but the static setup gives better out-of-sample predictions for 6 and 12 month holdout data.

3.5 Discussion and Conclusions

The main goal of this analysis was to explore how the proportions of different injury severity levels may be changing over time relative to one another, and to assess how pedestrian safety has changed over the observational time period. Through our statistical analysis, we were able to filter out the observational noise and estimate the temporal behavior of pedestrian crashes. The result of the compositional time series analysis revealed a substantial decrease in the proportions of pedestrian crashes in injury severity level A, and an increase in the level B proportions. The magnitude of the proportions of the K level showed relatively no change over the observational period

on urban roads in Connecticut. Moreover, we found indications of an increase in the proportions of levels A and B over the baseline (CO) during January and February with respective positive coefficients (standard errors) given by 0.2566 (0.1604) and 0.2901 (0.1916). We compared two different approaches for fitting/predicting pedestrian crash counts by injury severity level, viz., the Dynamic Setup and the Static Setup. The Dynamic Setup combines the fitted/predicted monthly proportions from the compositional model (3.2.1)-(3.2.3) and the fitted/predicted monthly total crash counts from the dynamic model (3.3.1a)-(3.3.3b). The Static Setup combines the fitted predicted monthly proportions from the compositional model (from Section 3.2) and the fitted predicted monthly total crash counts from the static model (from Section 3.3).

Additionally, the dynamic modeling for the pedestrian crash counts $\widehat{C}_{ti}$ enables us to investigate the temporal behavior of α_t for the different severity levels. We uncovered a decreasing trend in all pedestrian crash counts before April 2005, followed by a noticeable increase after that which lasted until May 2007, and then a flattening out until the end of the fitting period. This behavior appears to be largely due to changes in the severity level A pedestrian crashes.

To summarize, we conclude based on our analysis that the overall pedestrian safety on urban roads in Connecticut is increasing, because we discovered a shift into the less severe injury level (from level A to level B). Moreover, the behavior of total pedestrian crash counts appears to mirror the behavior of the level A crash counts. We suggest that more attention has to be focused on the pedestrian safety in the months of January and February, as we observed an increasing proportion of the A and B level crashes over the baseline CO level for these two months. Our analysis is essential for the segment/intersection study where the road side characteristics could

be controlled. With a rareness of pedestrian crashes it is impossible to investigate monthly temporal effects with an individual segment/intersection level data, thus the time dependence should be derived from the aggregated level data and then incorporated into segment/intersection study for example as a sampling strategy. Moreover dynamic models can uncover trends that can be used in conjunction with known application of pedestrian safety interventions on the statewide level to test their effectiveness. Our approach has also uncovered seasonal effects which are also useful for identifying at which times of year pedestrian safety is a consistently more serious issue. This can help with identifying interventions that would be appropriate for the pedestrian safety issues at those times of year. Overall the dynamic models can be used to monitor trends in pedestrian safety and can help to react in a timely manner to situations on the roads.

Chapter 4

Hierarchical Dynamic Models for Multivariate Times Series of Counts

In this chapter, we describe a hierarchical multivariate dynamic model (HMDM), with a multivariate Poisson distribution (MVP) as the sampling distribution for the response vector time series of counts, and incorporating covariates that may vary over location and/or time. The use of the MVP distribution enables us to model associations between the components of the count response vector, while the dynamic framework allows us to model the temporal behavior. The hierarchical structure enables us to capture the location (or subject) specific effects over time.

The format of this chapter follows. Section 4.1 gives a description of the ecological application, including a description of the data. Section 4.2 reviews the MVP distribution and describes fast computation of its probability mass function (pmf). Section 4.3 describes the HMDM model and gives details of the Bayesian inference.

Section 4.4 discusses model selection and prediction based on the fitted HMDM. Section 4.5 presents results for simulated data, while Section 4.6 presents an analysis of the ecological data on gastropod abundance.

4.1 Gastropod Abundance in the Luquillo Experimental Forest in Puerto Rico - Data Description

Understanding the causes and consequences of variation in the abundance of organisms has been a long-standing goal in ecology (Scheiner and Willig, 2011). Nonetheless, few long-term analyses spanning over 20 years of spatiotemporal variation in abundance exist, especially for invertebrate populations in tropical habitats that are subject to high intensity but infrequent disturbances such as hurricanes (e.g., Willig et al. (2013)). Terrestrial gastropods are of considerable ecological importance because of their abundance, diversity, and trophic position. Moreover, terrestrial gastropods, like non-marine mollusks in general, are suffering from global declines and are in need of scientifically informed conservation action and management (Lydeard et al., 2004). As such, there is considerable urgency to understand variation in gastropod abundance and the factors that affect it. In some ecosystems, gastropods respond to environmental gradients (Willig et al., 1998, 2011, 2013) and to disturbances (Bloch and Willig, 2006), including those induced by human activities, and do so at a variety of spatial scales. Because gastropods are ectothermic and not particularly vagile, they are constrained in distribution and behavior by desiccation stress (Cook and Barker, 2001). This fauna evinces a suite of attributes that suggests differential responses to spatial variation in habitat or micro-climate that might arise

as a consequence of global change. Finally, effective management or conservation of populations threatened by altered disturbance regimes requires species-specific understanding of the particular environmental aspects of change that are associated with alterations in abundance.

Long-term censuses of terrestrial gastropods were accomplished on the Luquillo Forest Dynamics Plot (LFDP; 18°20 N, 65°49 W), a16-ha grid in the northwest of the Luquillo Experimental Forest (LEF) in the Luquillo Mountains of northeastern Puerto Rico (McDowell et al., 2012). Although a modestly drier period typically extends from January to April (hereafter, the dry season), rainfall generally remains at least 20 cm in all months (Brown et al., 1983). The basic census design includes sampling during the wet and dry season of each year on each of forty circular sites (3-m radius) that are spaced evenly within a rectilinear grid such that 60-m separated adjacent points along a row or column (Willig et al., 1998; Bloch and Willig, 2010). The abundance of gastropods is sampled at each of those 40 sites during the dry and wet seasons for each year from 1991 to 2012. The number of replicate samples per season differed among years: 1 replicate in the dry season of 1991; 2 replicates in the dry season of 2003 and in the dry and wet seasons of 1992 and 1993; 3 replicates in the dry season of 1995 and in the wet and dry season of 1994; and 4 replicates in all other seasons and years between 1995 and 2012.

Although 17 species of gastropods are known to live in the Luquillo Forest Dynamics Plot, we focus on *Caracolus caracolla* and *Gaeotis nigrolineata*, the most abundant and widely distributed terrestrial gastropods in the tabonuco forest. We estimate abundance based on the minimum number known to be alive (MKNA) at each site during each season (i.e., the maximum number of individuals captured within a season at a site). All individuals were identified to species in the field and returned as

closely as possible to the point of capture and always within the site of capture. There is no ecological reason to support an assumption of negative association between the counts over time for these two species (Bloch and Willig, 2010).

In addition to species counts, environmental characteristics that are invariant over the course of the study were determined for each site. Elevation and slope (unitless) are continuous variables, ranging from 333 m asl to 428 m asl, and 0.7 to 65.1, respectively. Aspect and soil type are categorical variables that are estimated based on four 20 m by 20 m quadrants whose vertex was coincident with the center of each survey plot. The aspect of each quadrant was characterized into categories that represent the angular equivalents of the four cardinal (N, E, S, and W) or four intermediate (NE, SE, SW, and NW) compass directions. Because these categories arise from an underlying circular distribution, we quantify the central tendency after converting the angles to radians and applying a cosine transformation, doing so only when the observed aspect is no more than 135 degrees (in order to ensure environmentally informative characterizations). Aspect categories were combined into four distinct levels to increase sample sizes, viz., 1, 5, 7, and 8, of which level 1 is used as the baseline in the model shown in section 7. To ensure environmentally informative characterizations of soil type, we characterized each survey plot by the dominant soil type based on consideration of its four associated quadrants, when the most pervasive soil type was at least twice as frequent as the second most common soil type, and was dominant in at least two of the four quadrants. The soil type levels are Zarzal (1), Cristal (2), or Prieto (3).

Based on percent canopy cover (CC) evident in aerial photographs taken in 1936, US Forest Service records, and other sources, the LFDP can be subdivided into four canopy cover classes (Thompson et al., 2002). CC classes 1 and 2 were combined to

increase sample size, resulting in three areas of historic land use in 1936. The redefined CC level 1 (0–49% cover) experienced the most intensive logging and agriculture prior to 1934; CC level 2 (50–80% cover) was used for shade-coffee cultivation and other small scale mixed agriculture before 1934; and CC level 3 (80–100% cover) was lightly and selectively logged up to the 1950s.

As a consequence of disturbance and secondary succession, some habitat characteristics (e.g., canopy openness and plant apparency) vary over time at each site. Canopy openness (CO) was measured using a spherical densitometer, higher numbers representing greater canopy openness. Empirically, it equals the average number of grid cells that are not occluded by vegetation from measurements by a densitometer at the mid-point of the 4 cardinal radii of each sampling plot. Plant apparency is the volume of space in the understory that was occupied by plants, and is estimated using a plant apparency device at each of the mid-points along the cardinal radii. We estimated plant apparency (PA) via a method (Cook and Stubbendieck, 1986) that quantifies the aerial density of all living vegetation at heights up to 3 m above the forest floor (Secrest et al., 1996). Using a plant apparency device, we determine the cumulative number of foliar intercepts, defined as the sum of species-specific counts of living vegetation touching a wooden dowel at each of seven heights (0, 0.5, 1.0, 1.5, 2.0, 2.5, and 3.0 m). The device comprises a set of four 0.5 m long dowels positioned at 90^{o} angles at each height. The device is positioned 1.5 m from the center of each plot in each of the four cardinal directions. Apparency is estimated separately for *Prestoea acuminata*, the sierra palm (PAsp) and for all other plant species (PAothers). Canopy openness and plant apparency are measured only during wet seasons. In order to investigate the gastropod abundance, we propose hierarchical dynamic modeling of the bivariate count time series, which enables us to study the effect of site-specific

and time-dependent covariates, as described in the following sections.

4.2 Multivariate Poisson Distribution

The definition of an J-variate Poisson distribution is based on a mapping $g : \mathbb{N}^q \to \mathbb{N}^J$, $q \geq J$, such that $\boldsymbol{Y} = g(\boldsymbol{X}) = \mathbf{A}\boldsymbol{X}$ (Mahamunulu, 1967). Here, $\boldsymbol{X} = (X_1, \ldots, X_q)'$ is a vector of unobserved independent Poisson random variables, i.e., $X_r \sim \text{Poisson}(\lambda_r)$ for $r = 1, \ldots, q$; and $\mathbf{A}$ is an arbitrary $J \times q$ matrix which determines the properties of the multivariate Poisson distribution. The J-dimensional vector $\boldsymbol{Y} = (Y_1, \ldots, Y_J)' = \mathbf{A}\boldsymbol{X}$ follows a multivariate Poisson distribution with parameters $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_q)'$ and pmf

$$\begin{aligned} MP_J(\boldsymbol{y}|\boldsymbol{\lambda}) &= P(\boldsymbol{Y} = \boldsymbol{y}|\boldsymbol{\lambda}) = \sum_{\boldsymbol{x} \in g^{-1}(\boldsymbol{y})} P(\boldsymbol{X} = \boldsymbol{x}|\boldsymbol{\lambda}) \\ &= \sum_{\boldsymbol{x} \in g^{-1}(\boldsymbol{y})} \prod_{r=1}^{q} P(X_r = x_r|\lambda_r) \end{aligned} \tag{4.2.1}$$

where $g^{-1}(\boldsymbol{Y})$ denotes the inverse image of $\boldsymbol{Y} \in \mathbb{N}^J$ and for $r = 1, \ldots, q$, the pmf of the univariate Poisson distribution is $P(X_r = x_r|\lambda_r) = \exp(-\lambda_r)\lambda_r^{x_r}/x_r!$.

The literature on the use of the multivariate Poisson distribution for modeling applications was sparse until recently, possibly due to the complicated form of the pmf (4.2.1). Karlis and Meligkotsidou (2005) proposed the two-way covariance structured multivariate Poisson distribution which permits a more realistic modeling of multivariate counts for several practical applications. This distribution is constructed by setting $\boldsymbol{A} = [\boldsymbol{A}_1 \ \boldsymbol{A}_2]$, where $\boldsymbol{A}_1 = \boldsymbol{I}_J$ captures the main effects; $\boldsymbol{A}_2$ captures the two-way covariance effects; $\boldsymbol{A}_2$ is an $J \times \left(\frac{J(J-1)}{2}\right)$ binary matrix; each column of $\boldsymbol{A}_2$ has ex-

actly 2 ones and $(J-2)$ zeros and no duplicate columns exist; and $q = J+[J(J-1)]/2$. We correspondingly split the parameter $\boldsymbol{\lambda}$ into two parts, viz., $\boldsymbol{\lambda}^{(1)} = (\lambda_1, \cdots, \lambda_J)'$, which corresponds to the J main effects, and $\boldsymbol{\lambda}^{(2)} = (\lambda_{J+1}, \cdots, \lambda_q)'$ which corresponds to the $J(J-1)/2$ pairwise covariance effects. For example, when $J = 2$, the bivariate Poisson distribution with two-way covariance structure for $\boldsymbol{Y} = (Y_1, Y_2)'$ is expressed via $q = 3$ independent Poisson random variables as:

$$\begin{aligned} Y_1 &= X_1 + X_3 \\ Y_2 &= X_2 + X_3 \end{aligned} \tag{4.2.2}$$

where $X_i \sim \text{Poisson}(\lambda_i)$, $i = 1, 2, 3$. The joint pmf of Y_1 and Y_2 is:

$$\begin{aligned} P(Y_1 = y_1, Y_2 = y_2|\boldsymbol{\lambda}) &= \exp\{-(\lambda_1 + \lambda_2 + \lambda_3)\} \\ &\times \frac{\lambda_1^{y_1}}{y_1!}\frac{\lambda_2^{y_2}}{y_2!}\sum_{i=0}^{s}\binom{y_1}{i}\binom{y_2}{i} i! \left(\frac{\lambda_3}{\lambda_1\lambda_2}\right)^i \end{aligned} \tag{4.2.3}$$

where $s = \min(y_1, y_2)$.

Similarly, when $J = 3$, we write $\boldsymbol{Y} = (Y_1, Y_2, Y_3)'$ as:

$$\begin{aligned} Y_1 &= X_1 + X_4 + X_5 \\ Y_2 &= X_2 + X_4 + X_6 \\ Y_3 &= X_3 + X_5 + X_6 \end{aligned} \tag{4.2.4}$$

where $X_i \sim \text{Poisson}(\lambda_i)$ for $i = 1, \cdots, 6$. The joint probability mass function of Y_1,Y_2

and Y_3 is:

$$P(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3|\boldsymbol{\lambda}) = \exp\{-\sum_{i=1}^{6}\lambda_i\} \times \sum_{(X_4,X_5,X_6)\in C} \frac{\lambda_1^{y_1-X_4-X_5}\lambda_2^{y_2-X_4-X_6}}{(y_1-X_4-X_5)!(y_2-X_4-X_6)!} \times \frac{\lambda_3^{y_3-X_5-X_6}\lambda_4^{X_4}\lambda_5^{X_5}\lambda_6^{X_6}}{(y_3-X_5-X_6)!X_4!X_5!X_6!} \tag{4.2.5}$$

where the summation is over the set C such that $C = [(X_4, X_5, X_6) \in \mathbb{N}^3 : (X_4+X_5 \leq y_1) \cap (X_4 + X_6 \leq y_2) \cap (X_5 + X_6 \leq y_3)] \neq \emptyset]$

It is easy to see that the matrix $\boldsymbol{A}$ has the respective forms shown below for $J = 2$ and $J = 3$:

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \text{ and } \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}$$

The mean vector and variance-covariance matrix of $\boldsymbol{Y}$ are given by $E(\boldsymbol{Y}|\boldsymbol{\lambda}) = \boldsymbol{A\lambda}$ and $Cov(\boldsymbol{Y}|\boldsymbol{\lambda}) = \boldsymbol{A\Sigma A'}$, where $\boldsymbol{\Sigma} = \text{diag}(\lambda_1, \ldots, \lambda_q)$. As pointed out by Karlis and Meligkotsidou (2005), this covariance cannot accommodate negative associations among the components of $\boldsymbol{Y}$. When $J = 1$, the $MP_J(\boldsymbol{y}|\boldsymbol{\lambda})$ in (4.2.1) reduces to the univariate Poisson pmf $P(Y = y|\lambda) = \frac{e^{-\lambda}\lambda^y}{y!}$.

A recursive scheme was proposed by Tsiamyrtzis and Karlis (2004) for computing the multivariate Poisson pmf when $J = 2$ or $J = 3$. We use the following faster approach for the calculation of the multivariate Poisson pmf proposed in Hu (2012). When $J = 2$, let y_1 and y_2 denote the observed counts, and without loss of generality, assume that $y_1 \leq y_2$, so that $\min(y_1, y_2) = y_1$. Since X_3 is the common term in both equations in (4.2.2), it is straightforward to obtain the set of possible values that

X_3 can assume, viz., $x_3 = 0, \ldots, \min(y_1, y_2)$, and obtain the corresponding values assumed by X_1 and X_2 to be respectively $X_1 = y_1 - x_3$ and $X_2 = y_2 - x_3$. We have solved for all possible sets of values for the inverse image of $\boldsymbol{y}$, i.e., $\boldsymbol{x} \in g^{-1}(\boldsymbol{y})$. The pmf for the bivariate Poisson distribution can be calculated using (4.2.3).

When $J = 3$, without loss of generality, we assume that $y_1 \leq y_2 \leq y_3$. The possible values for x_4 and x_5 are in the set $C_1 = (0, \ldots, y_1)$, and the possible values for x_6 are in the set $C_2 = (0, \ldots, y_2)$. We have in total K different combinations for (x_4, x_5, x_6), where $K = (\text{length of set } C_1)^2 \times (\text{length of set } C_2) = (y_1+1)^2(y_2+1)$. The corresponding values for X_1, X_2, X_3 can be calculated from (4.2.4). Let C^* denote the set of K different combinations of possible values for all $q = 6$ independent Poisson variables. Since it is possible that in the set C^*, X_1, X_2, or X_3 may assume negative values, a subset of C^*, which only contains non-negative values of X_1, X_2, and X_3, is the inverse image of $\boldsymbol{y}$. The pmf of the trivariate Poisson distribution is then obtained using (4.2.5).

The computation of the multivariate Poisson likelihood directly depends on the magnitude of the counts, and a sizable portion of the computational effort in the Bayesian modeling is for the evaluation of the likelihood. In Table 4.2.1, we present the CPU times (seconds) for the likelihood calculation of $10,000$ simulated counts under three scenarios corresponding to different λ's. In these computations, we have assumed all pairwise covariance mean effects to be equal to 1, and have chosen different main effect means for each scenario. Specifically, in the univariate case when $J = 1$, we simulate Poisson random variables Y's under the three scenarios with means $\lambda = 3$, $\lambda = 5$ and $\lambda = 10$. In the bivariate case (4.2.2), we simulate X_1, X_2, X_3 under scenario 1 as univariate Poisson with respective means $\lambda_1 = 2, \lambda_2 = 2, \lambda_3 = 1$. Under scenario 2, $\lambda_1 = 4, \lambda_2 = 4, \lambda_3 = 1$, while under scenario 3, $\lambda_1 = 9, \lambda_2 = 9, \lambda_3 = 1$. In the

trivariate case (4.2.4), we simulate $X_1, \cdots, X_6$ under scenario 1 as univariate Poisson with respective means $\lambda_1 = \cdots = \lambda_6 = 1$. Under scenario 2, $\lambda_1 = \lambda_2 = \lambda_3 = 2$, and $\lambda_4 = \lambda_5 = \lambda_6 = 1$. Under scenario 3, $\lambda_1 = \lambda_2 = \lambda_3 = 8$, and $\lambda_4 = \lambda_5 = \lambda_6 = 1$. Simulations are run on Dell Optiplex 990 (Intel Core i7-2600 CPU a quad core processor with 3.4Ghz) with 16Gb of RAM using 32-bit version of Debian GNU/Linux version 6 operational system. The computational times for the simulations are comparable across the three scenarios for a given J, and increases rapidly as J increases.

TABLE 4.2.1: CPU Times (secs) for the Likelihood Computation

Likelihood	Scenario 1	Scenario 2	Scenario 3
Univariate Poisson	8×10^{-3}	8×10^{-3}	8.001×10^{-3}
Bivariate Poisson	8×10^{-3}	1.2×10^{-2}	2.4×10^{-2}
Trivariate Poisson	1.412	18.209	248.788

4.3 Hierarchical Multivariate Dynamic Model (HMDM)

Let $\boldsymbol{Y}_{it} = (Y_{1,it}, \cdots, Y_{J,it})$, for $t = 1, \cdots, T$ denote the J-dimensional time series of counts from location i, where $i = 1, \cdots, n$, and assume that $\boldsymbol{Y}_{it}$ follows a multivariate Poisson distribution (4.2.1). The observation equation of the HMDM is

$$\begin{aligned} \boldsymbol{Y}_{it}|\boldsymbol{\lambda}_{it} &\sim MP_J(\boldsymbol{y}_{it}|\boldsymbol{\lambda}_{it}) \\ \log \lambda_{j,it} &= \boldsymbol{D}'_{j,it}\boldsymbol{\delta}_{j,it} + \boldsymbol{S}'_{j,it}\boldsymbol{\eta}_j, \quad j = 1, \cdots, q \end{aligned} \tag{4.3.1}$$

where log denotes the natural logarithm, $\boldsymbol{D}_{j,it} = (D_{j,it,1}, \cdots, D_{j,it,a_j})'$ is an a_j-dimensional vector of exogenous predictors with location-time varying (dynamic) coefficients $\boldsymbol{\delta}_{j,it} = (\delta_{j,it,1}, \cdots, \delta_{j,it,a_j})'$ and $\boldsymbol{S}_{j,it} = (S_{j,it,1}, \cdots, S_{j,it,b_j})'$ is a b_j-dimensional vector of exoge-

nous predictors with static coefficients $\boldsymbol{\eta}_j = (\eta_{j,1}, \cdots, \eta_{j,it,b_j})'$. We assume that the model either includes $\delta_{j,it,1}$ which represents the location-time varying intercept, or includes $\eta_{j,1}$ which represents the static intercept, i.e., either $D_{j,it,1} = 1$ or $S_{j,it,1} = 1$. A simple formulation of (4.3.1) could set $a_j = 1$ for $j = 1, \cdots, q$, set $b_j = b > 1$ for $j = 1, \cdots, J$ and $b_j = 0$ for $j = J+1, \cdots, q$, which implies using only the location-specific and time-dependent intercept to model the Poisson means corresponding to the association portion, and the location-time intercept together with an equal number of static coefficients (corresponding to exogenous predictors) for the main effects portion of the multivariate Poisson specification.

For the reminder of this chapter, let $p_d = \sum_{j=1}^{q} a_j$ and $p_s = \sum_{j=1}^{q} b_j$. Let $\boldsymbol{\beta}_{it}$ be a p_d-dimensional vector constructed by stacking the a_j coefficients $\boldsymbol{\delta}_{j,it}$ for $j = 1, \cdots, q$. The structural equation of the HMDM relates the location-time varying parameter $\boldsymbol{\beta}_{it}$ to an aggregate (pooled) state parameter $\boldsymbol{\gamma}_t$:

$$\boldsymbol{\beta}_{it} = \boldsymbol{\gamma}_t + \boldsymbol{v}_{it} \tag{4.3.2}$$

where the errors $\boldsymbol{v}_{it}$ are assumed to be i.i.d. $N_{p_d}(\mathbf{0}, \boldsymbol{V}_i)$. The state (or system) equation of the HMDM is:

$$\boldsymbol{\gamma}_t = \boldsymbol{G}\boldsymbol{\gamma}_{t-1} + \boldsymbol{w}_t \tag{4.3.3}$$

where $\boldsymbol{G}$ is a $p_d \times p_d$ state transition matrix and the state errors $\boldsymbol{w}_t$ are assumed to be i.i.d. $N_{p_d}(\mathbf{0}, \boldsymbol{W})$.

The HMDM in (4.3.1)–(4.3.3) simplifies to the Hierarchical DGLM (HDGLM) when $J = 1$, where we replace $MP_J(\boldsymbol{y}_{it}|\boldsymbol{\lambda}_{it})$ by the univariate Poisson pmf.

4.3.1 Bayesian Inference

Let $\boldsymbol{Y}$, $\mathbf{D}$ and $\mathbf{S}$ denote all the responses $\boldsymbol{y}_{it}$, and the dynamic predictors and the static predictors for $t = 1, \cdots, T$ and $i = 1, \cdots, n$. Let $\boldsymbol{\eta}$ and $\boldsymbol{\beta}$ denote all the coefficients $\boldsymbol{\eta_j}$ and $\boldsymbol{\beta}_{it}$ for $j = 1, \cdots, q$, $t = 1, \cdots, T$ and $i = 1, \cdots, n$, and let $\boldsymbol{\gamma}$ denotes all the coefficients $\boldsymbol{\gamma}_t$ for $t = 1, \cdots, T$. The likelihood function under the model described by (4.3.1)–(4.3.3) is

$$\begin{aligned} L(\boldsymbol{\eta}, \boldsymbol{\beta}, \boldsymbol{\gamma}; \boldsymbol{Y}, \mathbf{D}, \mathbf{S}) &= \prod_{i=1}^{n} \prod_{t=1}^{T} MP_J(\boldsymbol{y}_{it} | \boldsymbol{\beta}_{it}, \boldsymbol{\gamma}_t) \\ \times \quad & p_{\text{normal}}(\boldsymbol{\eta}) \times p_{\text{normal}}(\boldsymbol{\beta}_{it} | \boldsymbol{\gamma}_t) \times p_{\text{normal}}(\boldsymbol{\gamma}_t | \boldsymbol{\gamma}_{t-1}) \end{aligned} \tag{4.3.4}$$

where we have suppressed the terms $\mathbf{D}$ and $\mathbf{S}$ on the right side for brevity. We assume multivariate Normal priors for the initial state vector and the static coefficients, i.e., $\boldsymbol{\gamma}_0 \sim N_{p_d}(\boldsymbol{J}_0, \boldsymbol{C}_0)$ and $\boldsymbol{\eta} \sim N_{p_s}(\boldsymbol{\mu}_\eta, \boldsymbol{\Sigma}_\eta)$. We assume inverse Wishart priors for the variance terms $\boldsymbol{V}_i$ and $\boldsymbol{W}$, i.e., $\boldsymbol{V}_i \sim IW(n_v, \boldsymbol{S}_v)$, and $\boldsymbol{W} \sim IW(n_w, \boldsymbol{S}_w)$, and we assume a product prior specification. The hyperparameters are selected to correspond to a vague prior specification. Prior elicitation is an important, ongoing problem of considerable interest in Bayesian analysis, and the vast, growing literature includes seminal work on objective priors (Berger, 2006), power priors (Ibrahim and Chen, 2000), expert elicited priors (O'Hagan et al., 2006), etc. In next chapters we discuss prior sensitivity for our analysis.

The joint posterior of the unknown parameters is proportional to the product of

the likelihood and the prior:

$$\begin{aligned}
\pi(\boldsymbol{\beta}_{it}, \boldsymbol{\gamma}_t, \boldsymbol{\eta}, \boldsymbol{V}_i, \boldsymbol{W} | \boldsymbol{Y}, \mathbf{D}, \mathbf{S}) &\propto [\prod_{t=1}^{T}\prod_{i=1}^{n} MP_J(\boldsymbol{y}_{it}|\boldsymbol{\lambda}_{it}) \\
\times \quad & |\boldsymbol{V}_i|^{-1/2} \exp\{-\frac{1}{2}(\boldsymbol{\beta}_{it} - \boldsymbol{\gamma}_t)' \boldsymbol{V}_i^{-1} (\boldsymbol{\beta}_{it} - \boldsymbol{\gamma}_t)\}] \\
\times \quad & |\boldsymbol{\Sigma}_\eta|^{-1/2} \exp\{-\frac{1}{2}(\boldsymbol{\eta} - \boldsymbol{\mu}_\eta)' \boldsymbol{\Sigma}_\eta^{-1} (\boldsymbol{\eta} - \boldsymbol{\mu}_\eta)\} \\
\times \quad & [\prod_{t=1}^{T} |\boldsymbol{W}|^{-1/2} \exp\{-\frac{1}{2}(\boldsymbol{\gamma}_t - \boldsymbol{G}\boldsymbol{\gamma}_{t-1})' \boldsymbol{W}^{-1} (\boldsymbol{\gamma}_t - \boldsymbol{G}\boldsymbol{\gamma}_{t-1})\}] \\
\times \quad & |\boldsymbol{C}_0|^{-1/2} \exp\{-\frac{1}{2}(\boldsymbol{\gamma}_0 - \boldsymbol{m}_0)' \boldsymbol{C}_0^{-1} (\boldsymbol{\gamma}_0 - \boldsymbol{m}_0)\} \\
\times \quad & [\prod_{i=1}^{N} |\boldsymbol{V}_i|^{-n_v/2} \exp\{-\frac{1}{2} tr(\boldsymbol{V}_i^{-1} \boldsymbol{S}_v)\}] \\
\times \quad & |\boldsymbol{W}|^{-n_w/2} \exp\{-\frac{1}{2} tr(\boldsymbol{W}^{-1} \boldsymbol{S}_w)\} \qquad (4.3.5)
\end{aligned}$$

The Gibbs sampler proceeds by sequentially sampling from the complete conditional distributions of the parameters, which are proportional to the joint posterior (4.3.5). The complete conditional densities of the unknown parameters are given below.

The complete conditional density of $\boldsymbol{\beta}_{it}$ for $i = 1, \cdots, n$ and $t = 1, \cdots, T$ is

$$\begin{aligned}
\pi(\beta_{it} | \boldsymbol{\gamma}_t, \boldsymbol{\eta}, \boldsymbol{V}_i, \boldsymbol{W}, \boldsymbol{Y}) &\propto MP_J(\boldsymbol{y}_{it}|\boldsymbol{\lambda}_{it}) \\
\times \quad & \exp\{-\frac{1}{2}(\boldsymbol{\beta}_{it} - \boldsymbol{\gamma}_t)' \boldsymbol{V}_i^{-1} (\boldsymbol{\beta}_{it} - \boldsymbol{\gamma}_t)\} \qquad (4.3.6)
\end{aligned}$$

The complete conditional density of $\boldsymbol{\gamma}_t$ for $t = 1, \cdots, T$ is

$$\begin{aligned} & \pi(\boldsymbol{\gamma}_t|\boldsymbol{\beta}_{it}, \boldsymbol{\eta}, \boldsymbol{V}_i, \boldsymbol{W}, \boldsymbol{Y}) \propto \\ \propto & \prod_{i=1}^{n} \exp\{-\frac{1}{2}(\boldsymbol{\beta}_{it} - \boldsymbol{\gamma}_t)' \boldsymbol{V}_i^{-1} (\boldsymbol{\beta}_{it} - \boldsymbol{\gamma}_t)\} \\ \times & \exp\{-\frac{1}{2}(\boldsymbol{\gamma}_t - \boldsymbol{G}\boldsymbol{\gamma}_{t-1})' \boldsymbol{W}^{-1} (\boldsymbol{\gamma}_t - \boldsymbol{G}\boldsymbol{\gamma}_{t-1})\} \\ \times & \exp\{-\frac{1}{2}(\boldsymbol{\gamma}_{t+1} - \boldsymbol{G}\boldsymbol{\gamma}_t)' \boldsymbol{W}^{-1} (\boldsymbol{\gamma}_{t+1} - \boldsymbol{G}\boldsymbol{\gamma}_t)\} \end{aligned} \tag{4.3.7}$$

The complete conditional density of $\boldsymbol{\eta}$ is

$$\begin{aligned} & \pi(\boldsymbol{\eta}|\boldsymbol{\beta}_{it}, \boldsymbol{\gamma}_t, \boldsymbol{V}_i, \boldsymbol{W}, \boldsymbol{Y}) \propto \prod_{t=1}^{T} \prod_{i=1}^{N} MP_J(\boldsymbol{y}_{it}|\boldsymbol{\lambda}_{it}) \\ \times & \exp\{-\frac{1}{2}(\boldsymbol{\eta} - \boldsymbol{\mu}_\eta)' \boldsymbol{\Sigma}_\eta^{-1} (\boldsymbol{\eta} - \boldsymbol{\mu}_\eta)\} \end{aligned} \tag{4.3.8}$$

The complete conditional density of $\boldsymbol{V}_i$ for $i = 1, \cdots, n$ is

$$\begin{aligned} & \pi(\boldsymbol{V}_i|\boldsymbol{\beta}_{it}, \boldsymbol{\gamma}_t, \boldsymbol{\eta}, \boldsymbol{W}, \boldsymbol{Y}) \propto \prod_{t=1}^{T} |\boldsymbol{V}_i|^{-1/2} \\ \times & \exp\{-\frac{1}{2}(\boldsymbol{\beta}_{it} - \boldsymbol{\gamma}_t)' \boldsymbol{V}_i^{-1} (\boldsymbol{\beta}_{it} - \boldsymbol{\gamma}_t)\} \\ \times & |\boldsymbol{V}_i|^{-n_v/2} \exp\{-\frac{1}{2} tr(\boldsymbol{V}_i^{-1} \boldsymbol{S}_v)\} \end{aligned} \tag{4.3.9}$$

The complete conditional density of $\boldsymbol{W}$ is

$$\begin{aligned}
\pi(\boldsymbol{W}|\boldsymbol{\beta}_{it}, \boldsymbol{\gamma}_t, \boldsymbol{\eta}, \boldsymbol{V}_i, \boldsymbol{Y}) &\propto \prod_{t=1}^{T} |\boldsymbol{W}|^{-1/2} \\
\times \quad & \exp\{-\frac{1}{2}(\boldsymbol{\gamma}_t - \boldsymbol{G}\boldsymbol{\gamma}_{t-1})'\boldsymbol{W}^{-1}(\boldsymbol{\gamma}_t - \boldsymbol{G}\boldsymbol{\gamma}_{t-1})\} \\
\times \quad & |\boldsymbol{W}|^{-1/2} \exp\{-\frac{1}{2}(\boldsymbol{\gamma}_0 - \boldsymbol{m}_0)'\boldsymbol{C}_0^{-1}(\boldsymbol{\gamma}_0 - \boldsymbol{m}_0)\} \\
\times \quad & |\boldsymbol{W}|^{-n_w/2} \exp\{-\frac{1}{2} tr(\boldsymbol{W}^{-1}\boldsymbol{S}_w)\} \qquad (4.3.10)
\end{aligned}$$

4.3.2 Details of Sampling Algorithms

Let $\boldsymbol{\Psi} = (\boldsymbol{V}_1, \cdots, \boldsymbol{V}_n, \boldsymbol{W})'$. Conditional on $\boldsymbol{\beta}_{it}$ and $\boldsymbol{\Psi}$, the structural and state equations of the HMDM have the form of a Gaussian Dynamic HDLM with *observations* $\boldsymbol{\beta}_{it}$ and state $\boldsymbol{\gamma}_t$. Note that $\boldsymbol{\gamma}_t$ is independent of $\boldsymbol{Y}_t$, given $\boldsymbol{\beta}_{it}$. Let $\boldsymbol{F}_t = (\boldsymbol{I}_{p_d}, \cdots, \boldsymbol{I}_{p_d})'$ denote an $Np_d \times p_d$ mapping matrix, and let $\boldsymbol{\beta}_t$ denote the vector obtained by stacking $\boldsymbol{\beta}_{it}$ for $i = 1, \cdots, n$. The complete conditional distribution of $\boldsymbol{\gamma}_t$ may be written as

$$P(\boldsymbol{\gamma}|\boldsymbol{\beta}, \boldsymbol{\Psi}) = P(\boldsymbol{\gamma}|\boldsymbol{\beta}, \boldsymbol{\Psi}) \prod_{t=1}^{T-1} P(\boldsymbol{\gamma}_t|\boldsymbol{\gamma}_{t+1}, \cdots, \boldsymbol{\gamma}_t, \boldsymbol{\beta}, \boldsymbol{\Psi}) \qquad (4.3.11)$$

The structure of the dynamic linear model implies that the second term on the right-hand side of the above equation reduces to $P(\boldsymbol{\gamma}_t|\boldsymbol{\gamma}_{t+1}, \boldsymbol{\beta}_t, \boldsymbol{\Psi})$. To generate a random sample from the complete conditional distribution of $\boldsymbol{\gamma}_t$, for $t = 1, \cdots, T$, using the Forward-Filtering-Backward-Sampling (FFBS) algorithm (Carter and Kohn, 1994), we implement the following steps:

Filter step: For $t = 1, \cdots, T$, we compute the mean $\boldsymbol{m}_t$ and variance matrix $\boldsymbol{C}_t$ of the posterior normal distributions $P(\boldsymbol{\gamma}_t|\boldsymbol{\gamma}_{t+1}, \boldsymbol{\beta}_t, \boldsymbol{\Psi})$, by applying the standard sequential updating results for Gaussian DLMs:

$$\begin{aligned}
\boldsymbol{m}_t &= \boldsymbol{G}\boldsymbol{m}_{t-1} + \boldsymbol{R}_t\boldsymbol{F}_t'\boldsymbol{Q}_t^{-1}(\boldsymbol{\beta}_t - \boldsymbol{F}_t\boldsymbol{G}\boldsymbol{m}_{t-1}) \\
\mathrm{C}_t &= \boldsymbol{R}_t - \boldsymbol{R}_t\boldsymbol{F}_t'\boldsymbol{Q}_t^{-1}\boldsymbol{F}_t\boldsymbol{R}_t \\
\boldsymbol{Q}_t &= \boldsymbol{V} + \boldsymbol{F}_t'\boldsymbol{R}_t\boldsymbol{F}_t \\
\boldsymbol{R}_t &= \boldsymbol{W} + \boldsymbol{G}\boldsymbol{C}_{t-1}\boldsymbol{G}',
\end{aligned}$$

where $\boldsymbol{V}$ is an $Np_d \times Np_d$ block diagonal matrix, each block represents the variance covariance matrix $\boldsymbol{V}_i$ for $i = 1, \cdots, n$.

Smooth step: At time $t = T$, we sample the final state vector $\boldsymbol{\gamma}_T$ from the marginal distribution, $P(\boldsymbol{\gamma}_T|\boldsymbol{\beta}_T, \boldsymbol{\Psi})$, which is $N(\boldsymbol{\gamma}_T|\boldsymbol{m}_T, \boldsymbol{C}_T)$. For time periods $t = T-1, \cdots, 0$, we sample from $P(\boldsymbol{\gamma}_t|\boldsymbol{\gamma}_{t+1}, \boldsymbol{\beta}_t, \boldsymbol{\Psi})$ which is $N(\boldsymbol{h}_t, \boldsymbol{H}_t)$ at each time, conditional on the latest value of $\boldsymbol{\gamma}_{t+1}$, where

$$\boldsymbol{H}_t = (\boldsymbol{C}_t^{-1} + \boldsymbol{G}'\boldsymbol{W}^{-1}\boldsymbol{G})^{-1} \tag{4.3.12}$$

$$\boldsymbol{h}_t = \boldsymbol{H}_t(\boldsymbol{C}_t^{-1}\boldsymbol{m}_t + \boldsymbol{G}'\boldsymbol{W}^{-1}\boldsymbol{\gamma}_{t+1}) \tag{4.3.13}$$

The results of these steps is a draw $(\boldsymbol{\gamma}_T, \cdots, \boldsymbol{\gamma}_1)$ from its complete conditional distribution.

The complete conditional distributions of the variance terms are inverse Wishart

distributions. Let $\widehat{\boldsymbol{\beta}}_{it} = \boldsymbol{\gamma}_t$. For each i, we sample $\boldsymbol{V}_i$ from $IW(n_i, \boldsymbol{S}_i)$, where

$$\begin{aligned} n_i &= T + n_v \\ \boldsymbol{S}_i &= \boldsymbol{S}_v + \sum_{t=1}^{T} \left[(\boldsymbol{\beta}_{it} - \widehat{\boldsymbol{\beta}}_{it})(\boldsymbol{\beta}_{it} - \widehat{\boldsymbol{\beta}}_{it})' \right] \end{aligned}$$

We sample $\boldsymbol{W}$ from $IW(n_w, \boldsymbol{S}_w)$ where

$$\begin{aligned} n_w &= T + n_w \\ \boldsymbol{S}_w &= \boldsymbol{S}_w + \sum_{t=1}^{T} \left[(\boldsymbol{\gamma}_t - \boldsymbol{G}\boldsymbol{\gamma}_{t-1})(\boldsymbol{\gamma}_t - \boldsymbol{G}\boldsymbol{\gamma}_{t-1})' \right] \end{aligned}$$

The Metropolis-Hastings algorithm is used for sampling $\boldsymbol{\beta}_{it}$ and $\boldsymbol{\eta}$.

The sampling based Bayesian framework for the univariate HDGLM model is similar to that of HMDM, but much simpler, since the conditional sampling distribution of Y_{it} is the univariate Poisson(λ_{it}) distribution. Since $J = q = 1$, this results in lower dimensional vectors and matrices, so that the computations become much faster.

4.4 Model Selection and Prediction

We use data on the first T time points from all n locations for model fitting, and then make predictions for the next L times. Predictions for $\boldsymbol{\lambda}_{it}$ are obtained by using the output of the Gibbs sampler to approximate the predictive density $p(\boldsymbol{\lambda}|\boldsymbol{Y}, \boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\Psi})$ using Monte Carlo integration. The Mean Absolute Deviance (MAD) and the Prediction Mean Absolute Deviance (PMAD) criteria are used for evaluating fits/predictions

in the calibration and hold-out data respectively, and are defined as

$$\begin{aligned} \text{MAD} &= \frac{1}{n}\frac{1}{T}\sum_{i=1}^{n}\sum_{t=1}^{T}\text{sign}(y_{it}-\widehat{\lambda}_{it})\sqrt{d_{it}} \\ \text{PMAD} &= \frac{1}{n}\frac{1}{L}\sum_{i=1}^{n}\sum_{t=T+1}^{T+L}\text{sign}(y_{it}-\widehat{\lambda}_{it})\sqrt{d_{it}} \end{aligned} \tag{4.4.1}$$

where $\widehat{\lambda}_{it}$ denotes the posterior mean of λ_{it}, and d_{it} denotes an individual deviance contribution and is defined as

$$d_{it} = 2\big(\log(P_{\text{Poisson}}(y_{it}|y_{it})) - \log(P_{\text{Poisson}}(y_{it}|\hat{\lambda}_{it}))\big) \tag{4.4.2}$$

where $P_{\text{Poisson}}(y|\lambda)$ denote the univariate Poisson pmf. The MAD and PMAD values reported in next sections are averaged over the Gibbs iterations.

4.5 Simulated Data Results

Data from the univariate HDGLM with counts Y_{it} ($i = 1, \cdots, 16$ and $t = 1, \cdots, 44$) is simulated according to the following model:

$$\begin{aligned} Y_{it}|\lambda_{it} &\sim \text{Poisson}(\lambda_{it}) \\ \log\lambda_{it} &= \delta_{it,1} + \delta_{it,2} + \eta S_{it} \end{aligned} \tag{4.5.1}$$

In (4.5.1), S_{it} represents a static predictor, which is simulated from Normal(0,1) and $\eta = 0.5$. Let $\boldsymbol{\beta_{it}} = (\delta_{it,1}, \delta_{it,2})'$ represent a location-time varying coefficient vector. The structural (hierarchical) and state equations are given by (4.3.2) and

(4.3.3), with $\mathbf{G} = \text{diag}(1, -1)$. The errors $\boldsymbol{v}_{it}$ in (4.3.2) are simulated from $N_2(\mathbf{0}, \boldsymbol{V}_i)$ with $\boldsymbol{V}_i = \text{diag}(0.05, 0.01)$ for $i = 1, \cdots, 16$, and $\boldsymbol{w}_t$ from (4.3.3) is simulated from $N_2(\mathbf{0}, \boldsymbol{W})$ with $\boldsymbol{W} = \text{diag}(10^{-3}, 10^{-4})$. Here, $p_d = 2$ and $p_s = 1$.

Estimation via the Bayesian framework discussed in section 4.3.1 is carried out on the first $T = 40$ observations, by assuming that the prior distribution of $\boldsymbol{W} = \text{diag}(W_1, W_2)$ is an Inverse-Wishart distribution with $2p_d + 1$ degrees of freedom and scale matrix $S_w = \text{diag}(10^3, 10^4)$. Similarly the prior distribution of $\boldsymbol{V}_i = \text{diag}(V_{1i}, V_{2i})$ is an Inverse-Wishart distribution with $2p_d + 1$ degrees of freedom and scale matrix $S_v = \text{diag}(25, 100)$. To investigate prior sensitivity, the prior for η_1 is assumed to be $N(0, 10^3)$, $N(0, 10^2)$ or $N(0, 10)$ for simulation scenarios 1, 2 and 3 respectively. In scenario 5, we assume the prior scales $S_w = \text{diag}(10^4, 10^4)$ and $S_v = \text{diag}(10^3, 10^3)$, along with a $N(0, 10^3)$ prior for η. Further, we consider different starting values for the η sampling. Thus scenarios 1 and 4 denote starting from MLE's under a static regression model and 0 respectively, both under a $N(0, 10^3)$ prior. Based on each fitted model, predictions are obtained for the next $L = 4$ times.

Data from a bivariate HMDM with the vector of counts $\boldsymbol{Y}_{it} = (Y_{1,it}, Y_{2,it})'$ ($i = 1, \cdots, 16$ and $t = 1, \cdots, 44$) is simulated according to the following model:

$$\begin{aligned} \boldsymbol{Y}_{it} | \lambda_{1,it}, \lambda_{2,it}, \lambda_{3,it} &\sim MP_2(\lambda_{1,it}, \lambda_{2,it}, \lambda_{3,it}) \\ \log \lambda_{1,it} &= \delta_{1,it,1} + \delta_{1,it,2} + \eta_1 S_{it} \\ \log \lambda_{2,it} &= \delta_{2,it,1} + \delta_{2,it,2} + \eta_2 S_{it} \\ \log \lambda_{3,it} &= \delta_{3,it,1} \end{aligned} \tag{4.5.2}$$

In (4.5.2), S_{it} represents a static predictor, which is simulated from N(0,1) and $\eta_1 =$

0.5 and $\eta_2 = -0.5$. Let $\boldsymbol{\beta}_{it} = (\delta_{1,it,1}, \delta_{2,it,1}, \delta_{3,it,1}, \delta_{1,it,2}, \delta_{2,it,2})'$ represent location-time varying coefficients. The structural and state equations are again given by (4.3.2) and (4.3.3), with $\mathbf{G} = \text{diag}(1, 1, 1, -1, -1)$. The errors $\boldsymbol{v}_{it}$ are simulated from $N_5(\mathbf{0}, \boldsymbol{V}_i)$ with $\boldsymbol{V}_i = \text{diag}(\boldsymbol{V}_{1i}, \boldsymbol{V}_{2i})$, where $\boldsymbol{V}_{1i}$ is a symmetric matrix with diagonal elements 0.05 and off-diagonal elements 0.001, for $i = 1, \cdots, 16$. Likewise, $\boldsymbol{V}_{2i}$ is a symmetric matrix with diagonal elements 0.01 and off-diagonal elements 0.001 for $i = 1, \cdots, 16$. The errors $\boldsymbol{w}_t$ are simulated from $N_5(\mathbf{0}, \boldsymbol{W})$ with $\boldsymbol{W} = \text{diag}(10^{-3}, 10^{-3}, 10^{-3}, 10^{-4}, 10^{-4})$. Here, $p_d = 5$ and $p_s = 2$.

The prior distribution of $\boldsymbol{W} = \text{diag}(W_1, \cdots, W_5)$ is an Inverse-Wishart distribution with $2p_d+1$ degrees of freedom and scale matrix $S_w = \text{diag}(10^3, 10^3, 10^3, 10^4, 10^4)$. The prior distribution of $\boldsymbol{V}_i = \text{diag}(\boldsymbol{V}_{1i}, \boldsymbol{V}_{2i})$ is an Inverse-Wishart distribution with $2p_d + 1$ degrees of freedom and scale matrix $S_v = \text{diag}(25, 25, 25, 100, 100)$. To investigate prior sensitivity, the prior for $\boldsymbol{\eta} = (\eta_1, \eta_2)'$ is assumed to be $N_2(\mathbf{0}, \text{diag}(10^3))$, $N_2(\mathbf{0}, \text{diag}(10^2))$ or $N_2(\mathbf{0}, \text{diag}(10))$ for simulation scenarios 1, 2 and 3 respectively. The prior scales $S_w = \text{diag}(10^4)$ and $S_v = \text{diag}(10^3)$ are considered for scenario 5, along with a $N_2(\mathbf{0}, \text{diag}(10^3))$ prior for $\boldsymbol{\eta}$. We again consider different starting values for the $\boldsymbol{\eta}$ sampling. Thus, scenarios 1, 2, 3 and 5 denote starting from ML estimates from a static regression model, and scenario 4 from 0 corresponding to a $N_2(\mathbf{0}, \text{diag}(10^3))$ prior.

We ran a total of 40,500 Gibbs iterations; with a burn-in of 500 iterations, and thinning of 40 iterations to obtain a posterior sample of size 1,000. Summaries from the posterior distributions of the static coefficients are shown in Table 4.5.1.

The estimated coefficients for $\boldsymbol{\eta}$ were close to the true values of the parameters across all simulation scenarios. The estimates for the other model parameters were also found to be reasonably close to true values, but have been omitted from the table

TABLE 4.5.1: Posterior Summaries for HMDM–Simulated Data

Scenario	Parameter	True value	Posterior Mean	Posterior Std. Dev.
Univariate		HDGLM		
1	η	0.5	0.5206	0.0217
2	η	0.5	0.5208	0.0220
3	η	0.5	0.5209	0.0216
4	η	0.5	0.5204	0.0215
5	η	0.5	0.5215	0.0216
Bivariate		HMDM		
1	η_1	0.5	0.4371	0.0209
	η_2	-0.5	-0.3610	0.0301
2	η_1	0.5	0.4372	0.0210
	η_2	-0.5	-0.3611	0.0303
3	η_1	0.5	0.4371	0.0211
	η_2	-0.5	-0.3610	0.0303
4	η_1	0.5	0.4370	0.0211
	η_2	-0.5	-0.3608	0.0303
5	η_1	0.5	0.4525	0.0257
	η_2	-0.5	-0.3746	0.0298

for the brevity. For simulation scenario 1 for the univariate HDGLM, MAD is 0.9080 and PMAD is 0.9829. For the bivariate HMDM, overall MAD and PMAD across both components of the count vector are respectively 0.8879 and 1.0389. The $\boldsymbol{\eta}$ estimates are also reasonably sensitive to changes in priors for the variance parameters $\mathbf{W}$ and $\mathbf{V}_i$.

4.6 Analysis of Gastropod Counts

For univariate gastropod counts Y_{it}, an HDGLM is

$$\begin{aligned}
& Y_{it} \sim \text{Poisson}(\lambda_{it}) \\
& \log \lambda_{it} = \alpha_{it} + Season_{it} + \eta_1 \text{LogRep}_t + \eta_2 \text{Elevation}_i \\
+ \;& \eta_3 \text{Slope}_i + \eta_4 \text{CO}_{it} + \eta_5 \text{PAsp}_{it} + \eta_6 \text{PAothers}_{it} \\
+ \;& \eta_7 \text{I(Aspect=5)}_i + \eta_8 \text{I(Aspect=7)}_i + \eta_9 \text{I(Aspect=8)}_i \\
+ \;& \eta_{10} \text{I(Soil Type=2)}_i + \eta_{11} \text{I(Soil Type=3)}_i \\
+ \;& \eta_{12} \text{I(CC=2)}_i + \eta_{13} \text{I(CC=3)}_i \\
= \;& \boldsymbol{D}'\boldsymbol{\beta}_{it} + \boldsymbol{S}'\boldsymbol{\eta} = \boldsymbol{I}\boldsymbol{\beta}_{it} + \boldsymbol{S}'\boldsymbol{\eta} \qquad (4.6.1)
\end{aligned}$$

In (4.6.1), let $\boldsymbol{\beta}_{it} = (\alpha_{it}, Season_{it})'$ denote the location-time varying coefficients, let $\boldsymbol{\eta}$ denote a 13-dimensional vector of static coefficients and $\boldsymbol{S}$ denote the corresponding vector of exogenous predictors. The structural and state equations are given by (4.3.2) and (4.3.3), with $\mathbf{G} = \text{diag}(1, -1)$. The specifications for Bayesian inference are what are shown for the univariate HDGLM in Section 4.5.

We next fit a bivariate HMDM to gastropod counts $\mathbf{Y}_{it} = (Carcar_{it}, Gaenig_{it})'$, where $Carcar_{it}$ and $Gaenig_{it}$ represent the abundance of *C. caracolla* and *G. nigro-lineata* respectively at location i and time t. Let $\boldsymbol{\eta}_j$ correspond to a vector of static coefficients for $j = 1, 2$ and $\boldsymbol{S}$ denote the vector of static predictors as defined in (4.6.1). As described in Section 4.3, $p_d = 5$, and $p_s = 26$. The observation equation

of the HMDM is

$$\begin{aligned} \boldsymbol{Y}_{it}|\lambda_{1,it},\lambda_{2,it},\lambda_{3,it} &\sim MP_2(\lambda_{1,it},\lambda_{2,it},\lambda_{3,it}) \\ \log\lambda_{1,it} &= \alpha_{1,it} + Season_{1,it} + \boldsymbol{S}'\boldsymbol{\eta}_1 \\ \log\lambda_{2,it} &= \alpha_{2,it} + Season_{2,it} + \boldsymbol{S}'\boldsymbol{\eta}_2 \\ \log\lambda_{3,it} &= \alpha_{3,it} \end{aligned} \tag{4.6.2}$$

The structural equation for $\boldsymbol{\beta}_{it} = (\alpha_{1,it}, \alpha_{2,it}, \alpha_{3,it}, Season_{1,it}, Season_{2,it})'$ is given in (4.3.2) and the state equation is given in (4.3.3) with $\mathbf{G} = \text{diag}(1, 1, 1, -1, -1)$.

We ran a total of 40,500 Gibbs iterations; with a burn-in of 500 iterations, and thinning of 40 iterations, to obtain a posterior sample of size 1,000. Table 4.6.1 shows prediction evaluation criteria for the two gastropod species, which are quite similar for the univariate and bivariate models. As before, MAD corresponds to the calibration data and PMAD corresponds to the holdout data.

TABLE 4.6.1: Model Selection Criteria–Gastropod Data

Model	Species	MAD	PMAD
Univariate	*C. caracolla*	1.2298	1.9130
	G. nigrolineata	1.3363	2.4915
Bivariate	Overall	1.2947	2.2087
	C. caracolla	1.3424	1.9082
	G. nigrolineata	1.2470	2.5093

The posterior means and standard deviations of the static coefficients for the univariate and bivariate models are given in Table 4.6.2. The posterior means for the first three diagonal elements of the matrix $\mathbf{W}$ are estimated to be 0.0011 with posterior standard deviation of about 0.0003 each, while the other two elements are estimated as 0.0001 with posterior standard deviation of around 0.00002 each. The estimated

posterior means for the first three diagonal elements of the matrix $\mathbf{V}_i$ are 0.044 with posterior standard deviations around 0.01, while the other diagonal elements are estimated around 0.01 with posterior standard deviation 0.002 each. All the estimated off-diagonal elements of $\mathbf{W}$ and $\mathbf{V}_i$ reveal no significant values for $i = 1, \cdots, 40$. Recall that the first three diagonal elements of $\mathbf{W}$ and $\mathbf{V}_i$ for $i = 1, \cdots, 40$ are associated with the location-specific and time-dependent intercepts, and the last two elements are associated with the location-specific seasonal components (with period equal to two, corresponding to the dry and wet periods).

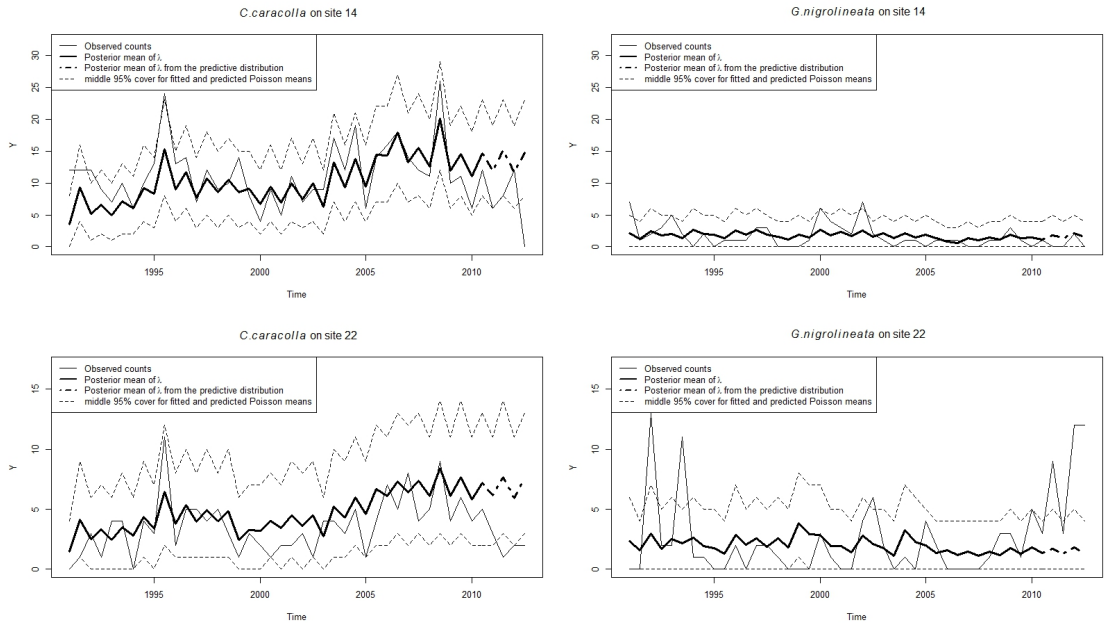

FIGURE 4.6.1: Model fits for counts of *C. caracolla* and *G. nigrolineata* on sites 14 and 22

Figure 4.6.1 shows the fitted means for the counts of *C. caracolla* and *G. nigrolineata* from Site 14 (top panel) and from Site 22 (bottom panel) based on the bivariate HMDM model. Clearly, the fitted model adequately tracks the stochastic pattern over time for almost all time points, especially for *C. caracolla*. The fit for

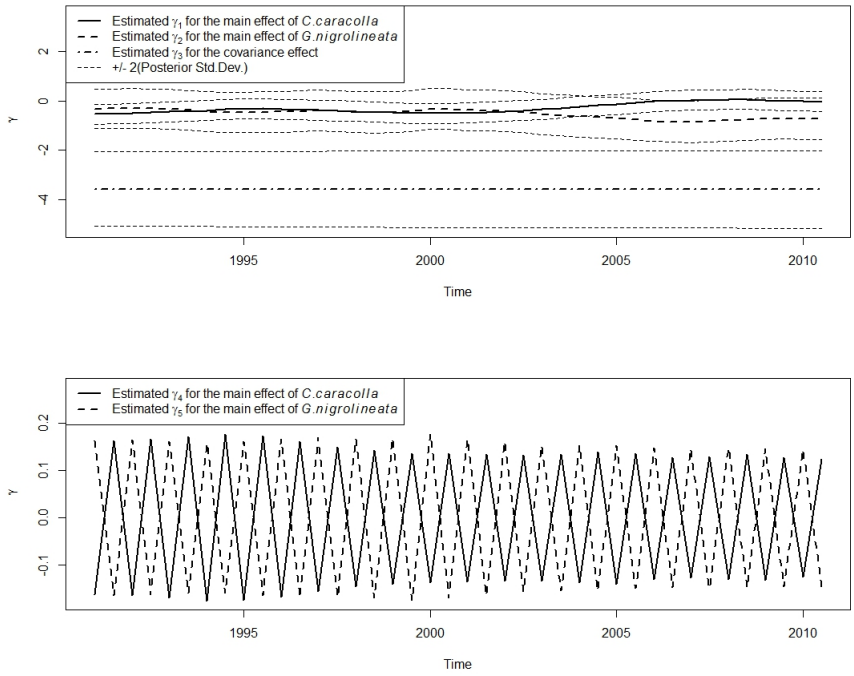

FIGURE 4.6.2: Estimated dynamic components of $\boldsymbol{\gamma}$ for *C. caracolla* and *G. nigrolineata*

G. nigrolineata is also good, although at a few time points, the model is unable to track the large empirical counts. This suggests that the abundance of *G. nigrolineata* is affected by environmental characteristics beyond those measured in this study, at least during some time periods. The significance of coefficients from the univariate and bivariate models (Table 4.6.2) are generally in accord for *C. caracolla*. Aspect and soil type have the strongest influence; the logarithm of the number of replicates has a modest influence, and other environmental characteristics (i.e., elevation, slope, canopy openness, canopy cover class, apparency of sierra palm, and plant apparency of species other than sierra palm) have little to no influence on variation in abundance. More specifically, Cristal soils as well as Zarzal soils support more individuals

than do Zarsal soils; these effects are likely mediated by differences in the nutrient characteristics of the soils and the plant assemblages that they harbor. All of the contrasts associated with aspect (over the baseline aspect level 1) are significant. Such effects could arise via two mechanisms: (i) abiotic effects of aspect on gastropods via influences on microclimate, especially diurnal temperature as well as (ii) biotic effects mediated by the well-documented responses of plant assemblages to aspect (i.e., those associated with insolation and variation in temporal patterns of temperature), especially in topographically steep environments, such as the Luquillo Mountains. Finally, the effect of logarithm of the number of replicates is to be expected, as the MNKA, the methodology used to estimate gastropod abundance, can only increase with increasing effort.

For *G. nigrolineata*, significance based on the univariate framework (see Table 4.6.2) are similar, but not identical, to those based on the bivariate model. In general, differences among sites in soil type and canopy cover class have a modest influence on abundance, whereas all other environmental characteristics have little or no significant effect. Abundance of *G. nigrolineata* on Cristal soils was lower than that on Zarzal soils, whereas abundance of *G. nigrolineata* on Prieto soils were higher than that on Zarzal soils. Moreovever, abundance on sites within cover class 2 (50–80% canopy openness in 1936) was greater than that on cover class 1 (0–49% canopy openness in 1936). The non-significant effect and negative value of the parameter estimate for the influence of the logarithm of the number of replicates on abundance is unexpected. This may be because the suite of environmental characteristics in this study do not include all variables to which *G. nigrolineata* responds most strongly (i.e., random effects capture variation in abundances associated with unmeasured environmental characteristics), or may arise because of time lags in the response of this species to

environmental variability. This is something to investigate in a future study of this species.

In summary, a comparison of static coefficients in Table 4.6.2 do not show a marked difference between the univariate and bivariate models for the two species. The top plot in Figure 4.6.2 shows that in the bivariate model, the estimated dynamic level coefficient corresponding to the main effect for *C. caracolla* exhibits a slowly increasing trend over time, while for *G. nigrolineata*, the trend is slowly decreasing over time. The estimated dynamic level coefficient corresponding to the second-order covariance effect is almost flat over time, indicating that the dependence between the species does not change over time. The bottom plot shows the decreasing amplitudes of the dynamic seasonal components for both species; the wet season shows an increased level for *C. caracolla* while the dry season shows an increased level for *G. nigrolineata*. The bivariate analysis is useful since it allows us to examine the temporal behavior of the second-order covariance effect, whose behavior would guide the modeler in terms of preferring a univariate or a multivariate model for understanding variation in species abundances. In this application for the two gastropod species, this effect is approximately constant over time, which explains why the results from the univariate and bivariate models are similar.

TABLE 4.6.2: Posterior Summaries for HMDM–Gastropod Data

Gastropod Species	Parameter η	Univariate Model Posterior Mean (Std. Dev.)	Bivariate Model Posterior Mean (Std. Dev.)
C. caracolla	LogRep	0.5492 (0.0668)	0.5650 (0.0705)
	Elevation	0.0011 (0.0007)	0.0001 (0.0005)
	Slope	0.0046 (0.0018)	0.0038 (0.0017)
	CO	0.0227 (0.0082)	0.0247 (0.0083)
	PAsp	-0.0071 (0.0015)	-0.0067 (0.0015)
	PAothers	-0.0017 (0.0009)	-0.0022 (0.0009)
	Aspect=5	1.0305 (0.0736)	1.0146 (0.0844)
	Aspect=7	0.8145 (0.0742)	0.7998 (0.0845)
	Aspect=8	0.4865 (0.0727)	0.4761 (0.0834)
	Soil Type=2	0.6975 (0.0331)	0.6827 (0.0327)
	Soil Type=3	1.8029 (0.0485)	1.7355 (0.0443)
	CC=2	0.2783 (0.0400)	0.2464 (0.0373)
	CC=3	0.0642 (0.0339)	0.0598 (0.0347)
G. nigrolineata	LogRep	-0.0644 (0.0934)	-0.0911 (0.0876)
	Elevation	0.0049 (0.0009)	0.0039 (0.0010)
	Slope	-0.0006 (0.0026)	-0.0033 (0.0025)
	CO	-0.1419 (0.0133)	-0.1328 (0.0128)
	PAsp	0.0194 (0.0018)	0.0193 (0.0019)
	PAothers	-0.0069 (0.0012)	-0.0061 (0.0012)
	Aspect=5	0.0447 (0.0951)	-0.1296 (0.0875)
	Aspect=7	-0.0207 (0.0934)	-0.0297 (0.0809)
	Aspect=8	0.1904 (0.0910)	0.2022 (0.0798)
	Soil Type=2	-0.1716 (0.0498)	-0.1644 (0.0525)
	Soil Type=3	0.2653 (0.0852)	0.1308 (0.0872)
	CC=2	0.3778 (0.0469)	0.2856 (0.0475)
	CC=3	0.1254 (0.0491)	0.1370 (0.0511)

Chapter 5

Dynamic Modeling of Multivariate Counts using Level Correlated Models

5.1 Introduction

Regression models which use a multivariate extension of the Poisson-lognormal mixture distribution are popular in different areas of applications; see Aitchison and Ho (1989); Park and Lord (2007); Ma et al. (2008). There are many situations where researchers may wish to model dependence in the response vector for data that are possibly overdispersed. Although a Bayesian framework using Markov Chain Monte Carlo (MCMC) techniques has been developed for model estimation (Chib and Winkelmann, 2001; Ma et al., 2008), this can be computationally expensive, especially for big data sets. In this chapter, we propose an approach that combines different marginal count distributions for multivariate count time series in addition

to employing the Integrated Nested Laplace Approximation (INLA) method to carry out approximate Bayesian inference.

We describe Level Correlated Models (LCMs) for multivariate count time series and illustrate the approach on an ecological example and a marketing example. We demonstrate the use of different marginal count distributions under the proposed flexible framework. Section 5.2 describes an overall structure of LCMs. Section 5.3 provides a detailed discussion about attainable associations in LCM; to our knowledge this is the first instance of such a study for multivariate count data. Section 5.4 gives a description of the ecology data, a detailed description of multivariate level correlated Poisson model fitting, and concludes with the ecology example application on gastropod abundance. Section 5.5 provides a description of the marketing data, introduces Poisson and ZIP component models under the LCM framework, and gives results and discussions for the marketing example.

5.2 Structure of Level Correlated Models

In several practical situations, responses arise as vectors of counts that vary across different observational locations. Univariate Poisson regression models for each of the components of the response vector cannot account for the dependence among the components, where the dependence may be due to omitted variables which simultaneously effect the response vector (Aitchison and Ho, 1989; Chib and Winkelmann, 2001; Ma et al., 2008). Therefore, an adequate multivariate model for counts is needed.

Let the response vector be a J-variate vector of counts. Data is collected across n observational subjects/locations. The level correlated multivariate Poisson model

can be defined as follows for $j = 1, \ldots, J$; $i = 1, \ldots, n$:

$$Y_{ji}|\lambda_{ji} \sim \text{Poisson}(\lambda_{ji}) \tag{5.2.1}$$

$$\lambda_{ji} = \exp(\alpha_{ji} + \boldsymbol{z}'_{ji}\boldsymbol{\beta}_j). \tag{5.2.2}$$

We assume that the components of the response vector are conditionally independent given their means. Throughout this section, we use the natural logarithmic function as the link function. Using the log link we can write:

$$\log(\lambda_{ji}) = \alpha_{ji} + \boldsymbol{z}'_{ji}\boldsymbol{\beta}_j, \tag{5.2.3}$$

where $i = 1, \ldots, n$ and $j = 1, \ldots, J$. Also in (5.2.3), $\boldsymbol{z}'_{ji}$ represents a p_j-dimensional vector of covariates with a vector of one's as the first column (thus there are $p_j - 1$ predictors), $\boldsymbol{\beta}_j$ is a p_j-dimensional vector of corresponding regression coefficients, and α_{ji} is a level correlated random effect component which follows a multivariate normal distribution:

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_J \end{pmatrix}_i \sim \text{Normal}(\mathbf{0}, \boldsymbol{\Sigma}), \tag{5.2.4}$$

where $i = 1, \ldots, n$, $\mathbf{0}$ is a vector of zeros and $\boldsymbol{\Sigma}$ is a variance-covariance matrix for the level correlated random effect term. Let $\boldsymbol{\alpha}_i = (\alpha_1, \ldots, \alpha_J)'$. The matrix $\boldsymbol{\Sigma}$ (acting through the random vector $\boldsymbol{\alpha}_i$) plays a crucial role in the dependence between the components of the vector of counts each of which follows a univariate

Poisson distribution. Next, we write the expectation and covariance matrix of the multivariate counts (Aitchison and Ho, 1989; Chib and Winkelmann, 2001; Ma et al., 2008). Let $\Sigma = \{\sigma_{rs}\}_{1 \le r \le s \le J}$; then for $i = 1, \ldots, n$, $j = 1, \ldots, J$ and for $r \neq s$:

$$\begin{aligned} \mathrm{E}[Y_{ji}] &= \exp(\boldsymbol{z'_{ji}}\boldsymbol{\beta_j}) \exp(\sigma_{jj}/2) = m_{ji}, && (5.2.5) \\ \mathrm{Var}[Y_{ji}] &= m_{ji} + m_{ji}^2(\exp(\sigma_{jj}) - 1), && (5.2.6) \\ \mathrm{Cov}[Y_{ri}, Y_{si}] &= m_{ri} m_{si}(\exp(\sigma_{rs}) - 1). && (5.2.7) \end{aligned}$$

The model (5.2.1)-(5.2.3) accounts for overdispersion, because the term $\sigma_{jj} > 0$ (diagonal elements of the positive definite matrix $\boldsymbol{\Sigma}$), so that $\mathrm{Var}[Y_{ji}] > \mathrm{E}[Y_{ji}]$. Moreover, the model can explain positive or negative dependence between the components of the response vector. In (5.2.7), the sign of the covariance depends directly on the value σ_{rs}. A negative value of σ_{rs} yields a negative association between the components Y_{ri} and Y_{si}, while a positive value of σ_{rs} results in a positive association.

Let $\boldsymbol{y}_i = (y_{1i}, \ldots, y_{Ji})'$ denote the multivariate response vector of counts observed on the ith subject/location. In order to derive the marginal distribution of $\boldsymbol{y}_i$, we need to integrate out $\boldsymbol{\alpha}_i$ from the joint probability density function:

$$\begin{aligned} g(\boldsymbol{y}_i | \boldsymbol{z}_{ji}, \boldsymbol{\beta}_j, \boldsymbol{\Sigma}) &= \int \cdots \int f_{Normal,J}(\boldsymbol{\alpha}_i | \boldsymbol{0}, \boldsymbol{\Sigma}) && (5.2.8) \\ &\times \prod_{j=1}^{J} f_{Poisson}(y_{ji} | \alpha_{ji}, \boldsymbol{z}_{ji}, \boldsymbol{\beta}_j) d\boldsymbol{\alpha}_i \end{aligned}$$

There is no closed algebraic solution to the J-dimensional integral, thus the marginal distribution of $\boldsymbol{y}_i$ cannot be directly derived. We will employ Bayesian methods for estimation.

The model described in equations (5.2.1)-(5.2.3) can be generalized to incorporate time dependence among counts as well as different marginal structures. Let $\boldsymbol{Y}_{it} = (Y_{1it}, \ldots, Y_{Jit})'$ be a J-variate vector of count responses over time, for $i = 1, \ldots, n$ and $t = 1, \ldots, T$. This means that we observe J types of counts on n subjects/locations over equally spaced T time points. To simplify the model specification, traditionally we can write the LCMs for multivariate time series in dynamic or state space model form with observation and state equations.

The observation equation is:

$$Y_{jit}|\lambda_{jit} \sim \text{Poisson}(\lambda_{jit}), \tag{5.2.9}$$

where $i = 1, \ldots, n$, $t = 1, \ldots, T$, $j = 1, \ldots, J$, and λ_{jit} is the random mean of Poisson distribution. We realize that the underlying distribution of counts need not be limited to the Poisson distribution since the data may fit better for example, if the zero inflated Poisson or the negative binomial is the underlying distribution. Therefore, the observation equation can be written in the more general form:

$$Y_{jit}|\boldsymbol{\theta}_{jit} \sim \text{UDC}_j(\boldsymbol{\theta}_{jit}), \tag{5.2.10}$$

where $i = 1, \ldots, n$, $t = 1, \ldots, T$ and $j = 1, \ldots, J$. UDC stands for "univariate distribution of counts" and includes, but is not limited to Poisson and ZIP distributions. The model can incorporate different marginal distributions for J types of counts in $\boldsymbol{Y}_{it}$ and any valid count data distribution could be a UDC. In (5.2.10) the vector $\boldsymbol{\theta}_{jit}$ denotes a set of parameters associated with some specific UDC.

In this section we restrict our attention to the model (5.2.9) with Poisson marginal

distribution. The case of other UDC's is a straightforward generalization. We model λ_{jit} as a function of predictors as follows:

$$\log(\lambda_{jit}) \quad = \quad \gamma_{jt} + \beta_{ji0} + \boldsymbol{z}'_{jit}\boldsymbol{\beta}_{ji} + \alpha_{jit}, \tag{5.2.11}$$

where $i = 1, \ldots, n$, $t = 1, \ldots, T$ and $j = 1, \ldots, J$. Notice that in (5.2.11), the link function for the mean can be changed according to the distribution in (5.2.10). In (5.2.11), the random effect β_{ji0} is a subject/location specific intercept which is specific to the jth type, γ_{jt} represents a specific time effect which depends on the jth response type, the random effect α_{jit} is a response type, time, and subject/location specific level correlated component, the vector $\boldsymbol{z}'_{jit}$ denotes a p_j-dimensional vector of covariates, and $\boldsymbol{\beta_{ji}}$ is a corresponding p_j-dimensional vector of coefficients. In general, $\boldsymbol{\beta_{ji}}$ can be extended to be a time varying coefficient although we do not incorporate the time effect here. The vector of covariates $\boldsymbol{z}_{jit}$ represents response type specific predictors that are observed over time on n subjects/locations.

Let $\boldsymbol{\alpha}_{it} = (\alpha_{1,it}, \ldots, \alpha_{Jit})'$. In (5.2.11), the dependence between different components of the count vector in (5.2.9) can be introduced at the observational level through $\boldsymbol{\alpha}_{it} \sim \text{Normal}(\mathbf{0}, \boldsymbol{\Sigma})$ or at the subject/location level through $\boldsymbol{\alpha}_{it} \sim \text{Normal}(\mathbf{0}, \boldsymbol{\Sigma_i})$. Here $\boldsymbol{\Sigma}$ is a constant and $\boldsymbol{\Sigma_i}$ is a subject/location specific variance-covariance matrix for the level correlated random effect term.

We assume that the components of the response type specific vector $\boldsymbol{\gamma}_t = (\gamma_{1,t}, \ldots, \gamma_{Jt})'$ evolve according to a random walk process. The state equation of the dynamic model is given by:

$$\gamma_{jt} = \gamma_{j(t-1)} + w_{jt}, \tag{5.2.12}$$

where $t = 1, \ldots, T$, $j = 1, \ldots, J$, and the error term $w_{jt} \sim \text{Normal}(0, 1/W)$, i.e., state errors for the different response types are assumed to follow the same normal distribution. We note that in general, the state error can be assumed to follow a response type specific normal distribution, i.e., $w_{jt} \sim \text{Normal}(0, 1/W_j)$.

The model defined by (5.2.9), (5.2.11), and (5.2.12) represents an LCM with underlying Poisson component distributions, where the time effect is aggregated over subjects/locations. There can be straightforward extensions of (5.2.11) and (5.2.12), where the time effect can be aggregated over response type and subjects/locations, aggregated over response type, or assumed to have separate time evolution on each subject/location for every response type.

The model where the time effect is aggregated over response types and over subjects/locations can be described by modifying equations (5.2.11) and (5.2.12) as follows:

$$\log(\lambda_{jit}) = \gamma_t + \beta_{ji0} + \boldsymbol{z}'_{jit}\boldsymbol{\beta}_{ji} + \alpha_{jit}, \tag{5.2.13}$$

$$\gamma_t = \gamma_{t-1} + w_t, \tag{5.2.14}$$

where $i = 1, \ldots, n$, $t = 1, \ldots, T$, $j = 1, \ldots, J$, and the error term is defined as $w_t \sim \text{Normal}(0, 1/W)$. All other terms are defined in the paragraphs following equations (5.2.9), (5.2.11), and (5.2.12).

The model where time effect is aggregated over response types is obtained by modifying the equations (5.2.11) and (5.2.12) as:

$$\log(\lambda_{jit}) = \gamma_{it} + \beta_{ji0} + \boldsymbol{z}'_{jit}\boldsymbol{\beta}_{ji} + \alpha_{jit}, \tag{5.2.15}$$

$$\gamma_{it} = \gamma_{i(t-1)} + w_{it}, \tag{5.2.16}$$

where $i = 1, \ldots, n$, $t = 1, \ldots, T$, $j = 1, \ldots, J$, and the error term is defined as $w_{it} \sim \text{Normal}(0, 1/W_i)$. Again all other terms are described in the paragraphs following equations (5.2.9), (5.2.11), and (5.2.12).

Finally, the model where time effect is assumed to have separate time evolution on each subject/location and for every response type, the equations (5.2.11) and (5.2.12) become:

$$\log(\lambda_{jit}) = \gamma_{jit} + \beta_{ji0} + \boldsymbol{z}'_{jit}\boldsymbol{\beta}_{ji} + \alpha_{jit}, \tag{5.2.17}$$

$$\gamma_{jit} = \gamma_{ji(t-1)} + w_{jit}, \tag{5.2.18}$$

where $i = 1, \ldots, n$, $t = 1, \ldots, T$, $j = 1, \ldots, J$, and the error term is defined as $w_{jit} \sim \text{Normal}(0, 1/W_i)$ (we assume that state errors share the same underlying distribution for all response types). All other terms stay as in the paragraphs following equations (5.2.9), (5.2.11), and (5.2.12).

5.3 Simulation Study

In this section we investigate the strength of the association structure induced by LCMs. We simulate the simplified LCM model and estimate the parameters using an approximate Bayesian inference via INLA. In the end of the section we give a CPU time comparison between INLA and a fully Bayesian inference using MCMC.

5.3.1 Attainable Correlation in LCMs

We denote the off-diagonal elements of the covariance matrix $\mathbf{\Sigma}$ in the equation (5.2.4) as follows, for $r \neq s$:

$$\sigma_{rs} = \rho_{rs}\sqrt{\sigma_{rr}\sigma_{ss}}, \tag{5.3.1}$$

where σ_{rr} and σ_{ss} are diagonal elements of $\mathbf{\Sigma}$ and ρ_{rs} is a correlation coefficient with range $-1 < \rho_{rs} < 1$. The association between counts Y_{ri} and Y_{si} is given as follows:

$$\begin{aligned} \text{Corr}[Y_{ri}, Y_{si}] &= \frac{\text{Cov}[Y_{ri}, Y_{si}]}{\sqrt{\text{Var}[Y_{ri}]\text{Var}[Y_{si}]}} \\ &= \frac{\exp(\rho_{rs}\sqrt{\sigma_{rr}\sigma_{ss}}) - 1}{\sqrt{(\exp(\sigma_{rr}) - 1 + m_{ri}^{-1})(\exp(\sigma_{ss}) - 1 + m_{si}^{-1})}} \end{aligned} \tag{5.3.2}$$

From equation (5.3.2), we observe that the correlation between counts is a function of ρ_{rs}, σ_{rr}, σ_{ss}, m_{ri}, and m_{si}, which are defined in (5.2.5). Clearly, in the case of counts, the association is not the same as the correlations employed with the multivariate normal distribution. We believe that it is important to demonstrate the range of the correlations induced by the LCMs.

To this end, we discuss in detail the limits for the association that can be attained between the count random variables Y_{ri} and Y_{si}. We note that the lower and upper limits of this association vary for different values of ρ_{rs}, σ_{rr}, σ_{ss}, m_{ri}, and m_{si}. To illustrate and quantify the nature of this relationship, we compute different values of the association in the bivariate case with one pair of observations (Y_1, Y_2). To simplify the notation and the model specification, we assume equality of the diagonal elements in $\mathbf{\Sigma}$, that is, $\sigma_{11} = \sigma_{22} = \sigma$, and take $\exp(\boldsymbol{z_1'\beta_1}) = \exp(\boldsymbol{z_2'\beta_2}) = \text{M}$. Then

(5.3.2) simplifies as follows:

$$\text{Corr}[Y_1, Y_2] \;=\; \frac{\exp(\rho\sigma) - 1}{\exp(\sigma) - 1 + (\text{M}\exp(\sigma/2))^{-1}} \tag{5.3.3}$$

In Figure 5.3.1, we plot the values of $\text{Corr}[Y_1, Y_2]$ as in equation (5.3.3), fixing the values of ρ and σ, and varying the values of M. The values of M are taken in the range from 0 to 20 and values of σ are taken to be 2, 1, 0.5, and 0.1. In Figure 5.3.2, instead of fixing σ, we fix the values of M at levels 1, 5, 10, 30, 100. In most of the real data applications, we have observed/estimated such values of the parameters. We therefore believe that the proposed values cover the most common situations encountered under the LCM framework. We take values of $\rho = 1$ (solid lines) and $\rho = -1$ (dashed lines), which represent the situations with the strongest positive and the strongest negative possible associations under LCMs, respectively.

From Figures 5.3.1 and 5.3.2 we can conclude that the positive and negative relationships introduced by (5.2.1)-(5.2.3) can cover a wide range of values. We notice that with the increase of σ, the strength of a negative association between Y_1 and Y_2 can rapidly decrease, whereas the strength of a positive association does not show such a strong decrease.

Table 5.3.1 presents maximum and minimum values of the attainable correlation $\text{Corr}[Y_1, Y_2]$ as a function of σ, keeping the values of ρ and M fixed at the levels given in rows and columns respectively. As M increases the attainable correlation $\text{Corr}[Y_1, Y_2]$ is larger in absolute value for positive values of ρ than for negative values of ρ. Even for low values of M, we can expect $\text{Corr}[Y_1, Y_2]$ to be estimated as high as 0.9884. On the other hand, the minimum values of $\text{Corr}[Y_1, Y_2]$ under the case of negative association are slowly decreasing towards -1. For moderately large counts we expect

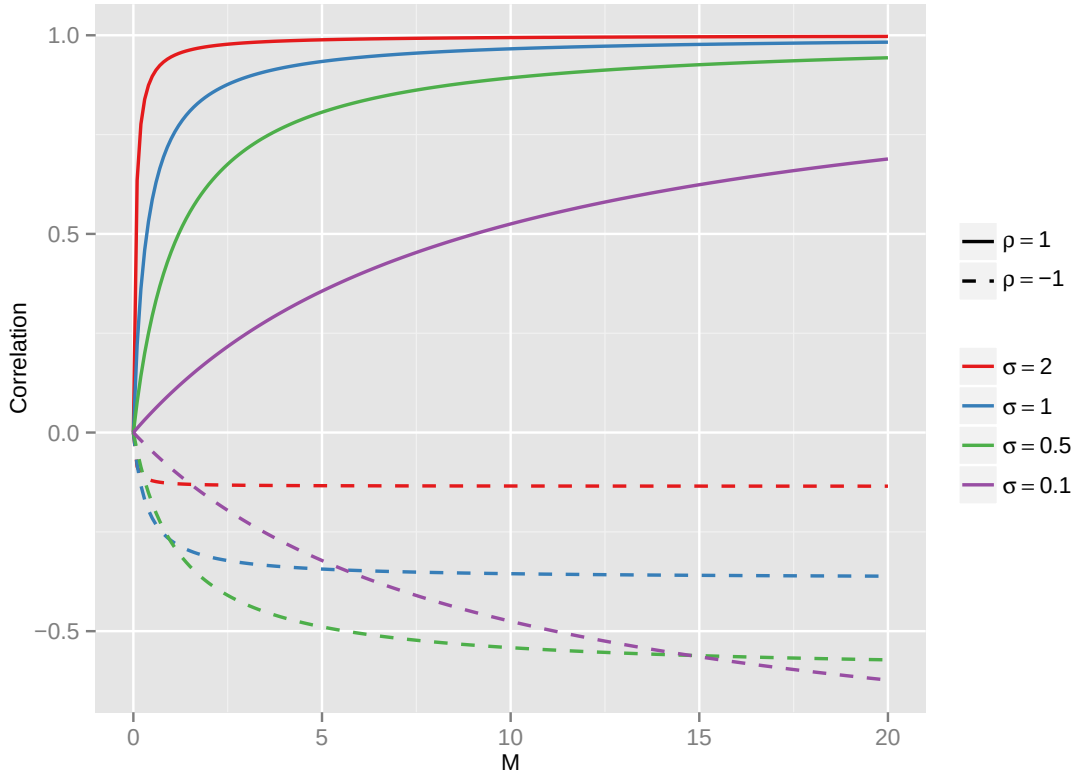

FIGURE 5.3.1: Attainable correlation $\mathrm{Corr}[Y_1, Y_2]$ of LCMs for different M

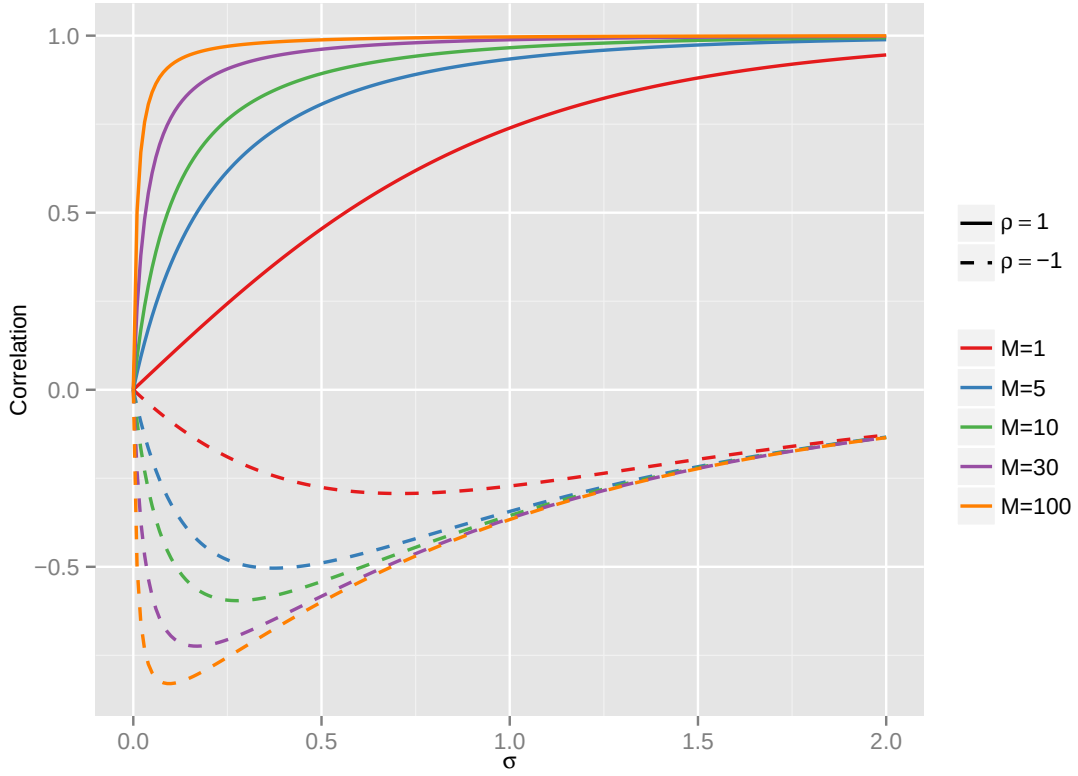

FIGURE 5.3.2: Attainable correlation $\mathrm{Corr}[Y_1, Y_2]$ of LCMs for different σ

TABLE 5.3.1: Maximum (minimum) attainable Corr$[Y_1, Y_2]$ for given positive (negative) ρ and selected values of M

ρ	M=1	M=2	M=3	M=5	M=10	M=15	M=20	M=25	M=30	M=50	M=100
1	0.9884	0.9942	0.9961	0.9977	0.9988	0.9992	0.9994	0.9995	0.9996	0.9998	0.9999
0.9	0.7477	0.7719	0.7853	0.8011	0.8206	0.8308	0.8375	0.8424	0.8461	0.8558	0.8666
0.8	0.5909	0.6251	0.6439	0.6663	0.6936	0.7078	0.7171	0.7238	0.729	0.7421	0.7567
0.7	0.4698	0.5074	0.5283	0.553	0.5834	0.5991	0.6094	0.6168	0.6226	0.637	0.6531
0.6	0.3707	0.4077	0.4285	0.4532	0.4835	0.4993	0.5096	0.5171	0.5228	0.5373	0.5533
0.5	0.287	0.3209	0.34	0.3629	0.3912	0.406	0.4156	0.4225	0.4279	0.4414	0.4564
0.4	0.2148	0.2438	0.2603	0.2801	0.3047	0.3176	0.326	0.3321	0.3368	0.3487	0.3618
0.3	0.1515	0.1743	0.1874	0.2033	0.223	0.2334	0.2402	0.2451	0.2489	0.2584	0.2691
0.2	0.0954	0.1112	0.1204	0.1314	0.1453	0.1527	0.1575	0.1609	0.1636	0.1704	0.178
0.1	0.0452	0.0534	0.0581	0.0639	0.0711	0.075	0.0775	0.0793	0.0808	0.0843	0.0883
-0.1	-0.0411	-0.0495	-0.0545	-0.0606	-0.0684	-0.0726	-0.0753	-0.0773	-0.0789	-0.0828	-0.0872
-0.2	-0.0785	-0.0956	-0.1058	-0.1184	-0.1344	-0.143	-0.1487	-0.1528	-0.156	-0.1641	-0.1732
-0.3	-0.1129	-0.1388	-0.1543	-0.1735	-0.1982	-0.2114	-0.2202	-0.2266	-0.2315	-0.2441	-0.2583
-0.4	-0.1446	-0.1793	-0.2002	-0.2263	-0.2599	-0.278	-0.29	-0.2988	-0.3056	-0.3229	-0.3424
-0.5	-0.1739	-0.2175	-0.2438	-0.2769	-0.3198	-0.343	-0.3583	-0.3696	-0.3783	-0.4006	-0.4256
-0.6	-0.2011	-0.2535	-0.2854	-0.3256	-0.378	-0.4063	-0.4252	-0.439	-0.4498	-0.4771	-0.508
-0.7	-0.2264	-0.2876	-0.3251	-0.3725	-0.4345	-0.4682	-0.4907	-0.5072	-0.52	-0.5527	-0.5896
-0.8	-0.2501	-0.32	-0.3631	-0.4178	-0.4896	-0.5288	-0.5549	-0.5741	-0.5891	-0.6272	-0.6704
-0.9	-0.2722	-0.3508	-0.3994	-0.4615	-0.5432	-0.588	-0.6179	-0.6399	-0.6571	-0.7009	-0.7505
-1	-0.293	-0.3801	-0.4343	-0.5037	-0.5956	-0.6461	-0.6798	-0.7047	-0.7241	-0.7737	-0.83

that Corr$[Y_1, Y_2]$ will only be around -0.7241. Therefore, one can use Table 5.3.1 as a reference for the expected strength of the attainable association under LCMs.

Overall we can conclude that the absolute values of the attainable association are larger when positive association is present than in the case of negative association, for fixed values of M and σ. One simple explanation of this behavior is based on the bounded support of random variables associated with LCMs. Consider the bivariate case with positive association. As one random variable increases or decreases so does the other random variable. Thus, we do not expect limitations to the correlation value as the means of counts move away from zero. In contrast, if one random variable increases to infinity, the negatively associated random variable is bounded below and cannot stretch beyond zero. Thus, we observe a slower increase in the magnitude of a negative correlation with the increase in means under the LCMs. Nevertheless,

we can conclude that overall the association structure induced by LCMs is diversely rich. The model can account for a reasonably wide range of negative and positive associations. Moreover, the flexible LCM structure can incorporate information about potentially important covariates, so that the LCMs can be useful and adequate for modeling vector count time series in various practical situations.

5.3.2 Estimation using INLA

We describe model estimation for the simulated bivariate data. Let $\boldsymbol{Y} = (Y_{1i}, Y_{2i})$ represent the data vector that is observed at n locations. We note that the simulation example is run for a simplified model described by (5.2.1)-(5.2.3) and does not assume any temporal dependence. Also, for simplicity we denote $\mathrm{M}_j = \exp(\boldsymbol{z}'_{ji}\boldsymbol{\beta}_j)$. Thus, we simulate counts accordantly to the following model:

$$Y_{ji}|\lambda_{ji} \sim \text{Poisson}(\lambda_{ji}), \tag{5.3.4}$$

$$\lambda_{ji} = \mathrm{M}_j \exp(\alpha_{ji}), \tag{5.3.5}$$

where $j = 1, 2$, $i = 1, \ldots, n$, and α_{ji} defined in the equation (5.2.4). Thus, there are five parameters associated with the simulated model, namely σ_{11}, σ_{22}, ρ, M_1, and M_2. Note that here, a different intercept M_j is assumed for each response type. We assume the true parameter values to be $\sigma_{11} = \sigma_{22} = 0.5$, $\rho = -0.99$, and $\mathrm{M}_1 = \mathrm{M}_2 = 5$. We run the simulation for $n = 50$, $n = 500$, $n = 1000$, and $n = 5000$. For the model estimation, we use the INLA framework described in Section 2.2. Table 5.3.2 gives the posterior summaries for the estimated parameters. Table 5.3.3 gives computed values of the Pearson product-moment correlation coefficient and Kendall's-tau measure of

TABLE 5.3.2: Estimated parameters in simulated LCMs

Sample Size	Parameters in LCM	Posterior mean	95% Credible Interval		True values
			0.025 quantile	0.975 quantile	
$n = 50$	σ_{11}	0.289	0.169	0.545	0.5
	σ_{22}	0.47	0.287	0.836	0.5
	ρ	-0.698	-0.87	-0.436	-0.99
	M_1	4.942	4.002	6.021	5
	M_2	4.867	3.788	6.021	5
$n = 500$	σ_{11}	0.534	0.45	0.64	0.5
	σ_{22}	0.472	0.398	0.563	0.5
	ρ	-0.902	-0.937	-0.857	-0.99
	M_1	4.669	4.31	5.047	5
	M_2	5.036	4.671	5.421	5
$n = 1000$	σ_{11}	0.494	0.438	0.558	0.5
	σ_{22}	0.545	0.484	0.616	0.5
	ρ	-0.935	-0.954	-0.909	-0.99
	M_1	5.067	4.801	5.342	5
	M_2	4.815	4.551	5.089	5
$n = 5000$	σ_{11}	0.506	0.48	0.535	0.5
	σ_{22}	0.505	0.478	0.533	0.5
	ρ	-0.958	-0.968	-0.947	-0.99
	M_1	5.018	4.898	5.14	5
	M_2	4.884	4.766	5.003	5

association (Agresti, 2013), and the estimated value of $\text{Corr}[Y_1, Y_2]$ through the LCM model fit using INLA. We note that the estimated values are close to the true values of the parameters, although when $n = 50$, the credible intervals for ρ, σ_{11}, and σ_{22} are somewhat wide. Nevertheless, with the increase of n, all the estimates get reasonably close to the true values. Under the estimated LCM models, the value of $\widehat{\rho}$, rather than the corresponding value of $\text{Corr}[Y_1, Y_2]$, provides a more reasonable indication of the association between Y_1 and Y_2.

Table 5.3.4 gives a CPU time comparison between an approximate Bayesian inference using INLA and a fully Bayesian inference using MCMC. The MCMC sampling

TABLE 5.3.3: Estimated correlation Corr$[Y_{1i}, Y_{2i}]$ in simulated LCMs

	Correlation between Y_{1i} and Y_{2i}			
Sample size	Pearson	Kendall	LCM-INLA	True
n=50	-0.446	-0.486	-0.364	-0.485
n=500	-0.478	-0.511	-0.444	-0.485
n=1000	-0.517	-0.538	-0.461	-0.485
n=5000	-0.503	-0.519	-0.472	-0.485

TABLE 5.3.4: CPU time of LCMs estimation using INLA and MCMC (in seconds)

Sample size	INLA	MCMC
n=50	1	7
n=500	4	59
n=1000	8	116
n=5000	77	582

is done using the R package $MCMCglmm$ with $105,000$ iterations. We used 5000 iterations for burn-in, and used 100 iterations for thinning, based on the autocorrelations in the generations. This resulted in a final posterior sample of size 1000, from which posterior summaries were calculated. All the posterior estimates from INLA and MCMC are reasonably close to each other. The computational time quickly increases with the increase of sample size n for the MCMC runs. INLA runs faster than MCMC, as expected in all cases, and is approximately 7 times faster for large n. We stress the fact that the $R\text{-}INLA$ package can easily incorporate time series structure in LCMs, which makes it an attractive tool from an end-user point of view. We also note that in most of our real applications, the sample size is above 1000 observations with the largest sample size reaching $100,000$ observations.

We replicate the simulation of LCM given by (5.3.4)-(5.3.5) 100 times for $n = 1000$. For each of the replicates we estimate model parameters using INLA and MCMC. Table 5.3.5 provides averages over the 100 replications of (a) the posterior means of

TABLE 5.3.5: Averages of posterior means and MSE's of 100 replicates for $n = 1000$

Parameters in LCM	INLA		MCMC		True values
	Posterior mean	MSE	Posterior mean	MSE	
σ_{11}	0.511	0.001172561	0.502	0.001274594	0.5
σ_{22}	0.511	0.001080288	0.5	0.000988587	0.5
ρ	-0.925	0.004252735	-0.991	0.0000901	-0.99
M_1	4.961	0.021279228	4.991	0.019831816	5
M_2	4.984	0.017655544	5.017	0.019169549	5

the parameters and (b) the corresponding mean squared errors (MSE's) of estimation. We conclude that INLA posterior estimates are reasonably close to MCMC estimates in most of the cases. MSE's from MCMC estimation are slightly smaller for σ_{11}, σ_{22}, and M_1 parameters. The largest difference in MSE's between INLA and MCMC is for the ρ parameter. We also note that the MSE for M_2 is in fact smaller in the case of INLA.

One drawback of *R-INLA* is that a user is restricted to the "canned" code that is provided, although someone with statistics and computing skills can easily code the algorithms in Fortran or C for more complex models that are not available on the R-INLA website. We have been able to successfully apply the R-INLA approach for the real data analyses, as shown in the following sections.

5.4 Gastropod Abundance Modeling Using LCM

5.4.1 Data Description

Understanding the causes and consequences of variation in the abundance of organisms has been a long-standing goal in ecology. The problem description for the ecological application is given in Chapter 4. We use updated data for multivariate

level correlated modeling. A brief description is given below and for further details, see Section 4.1.

Long-term censuses of terrestrial gastropods were accomplished on the 16-ha grid in the northwest of the Luquillo Experimental Forest (LEF) in the Luquillo Mountains of northeastern Puerto Rico. Figure 5.5.1 gives the map of the Caribbean showing the location of Puerto Rico, as well as the location of the Luquillo Experimental Forest (LEF). An elevation relief map of the study area illustrates topographic variability of the forest and the location and proximity (60-m spacing) of the 40 circular plots ($r = 3$ m) where sampling was done. Color of circles indicates the degree of disturbance based on historical land use. Here red denotes intensive logging with < 50 % cover, yellow stands coffee plantations with $50 - 80$ % canopy cover, blue indicates selective logging with > 80 % cover. Vertical axis represents elevation in meters. There are 17 species of gastropods that are known to live in the Luquillo Forest. We focus on *Caracolus caracolla*, *Gaeotis nigrolineata*, and *Nenia tridens*, the most abundant and widely distributed terrestrial gastropods. We estimate abundance based on the minimum number known to be alive (MKNA) at each site. All individuals were identified to species in the field and returned as close as possible to the point of capture and always within the site of capture.

The data were collected on 40 sites from 1995 to 2014 and on the additional 111 sites for the following years 1995, 1996, and 1997 during wet season. For this period of time we consistently have 4 replicates of observations. All site and time specific covariates are measured at the wet season in the data. The four variables are added, namely Litter Cover (the measure of mean leaf litter cover at a point calculated from ranks that range from 1-5), DEAD (the total number of intercepts of dead vegetation as determined with a plant apparency device), PREMON (the total

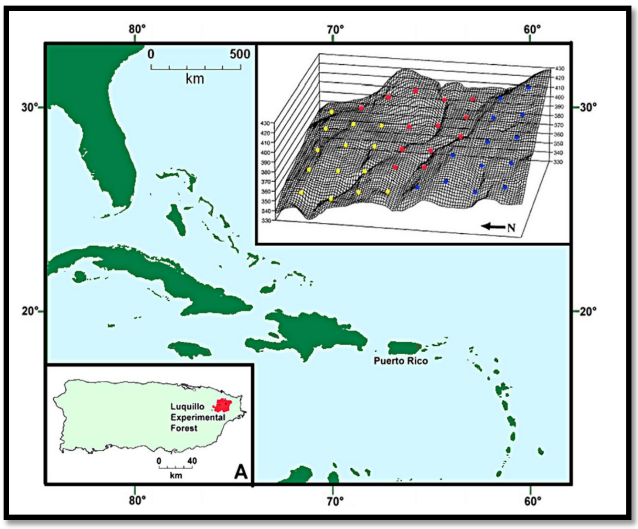

FIGURE 5.4.1: Location of the Luquillo Experimental Forest

number of intercepts of sierra palm as determined with plant apparency device), and PA (the total number of intercepts of plant species as determined with plant apparency device minus the intercepts of PREMON). The site (location) specific covariates that are used in the analysis are Elevation, Slope, Aspect, Soil Type, and Canopy Cover. We use Canopy Openness, Leaf Litter Cover, Dead Vegetation, Plant Apparency (PA) and Apparency of Sierra Palm (PREMON) as site and time specific covariates.

5.4.2 Model Framework

The problem and data description are given in Section 4.1 and updates to the data are explained in Section 5.4.1. We use the most recent data available for the gastropods counts. In the analysis presented in this section the modified set of predictors is used

in the model. The four variables are added, namely Litter Cover, DEAD, PREMON, and PA. The natural logarithm transformation is applied for all continuous variables such as Elevation, Slope, Canopy Openness (CO), Apparency of Sierra Palm (PREMON), Plant Apparency (PA) and Amount of Dead Vegetation (DEAD). The reference category for Aspect is taken to be 8 corresponding to locations that are facing South direction. Also the Cover Class=2 corresponding to locations with the most intensive logging and agriculture prior to 1934, Soil Type=1 which is the Zarzal soil, and Litter Cover=1 are taken as baseline values in the model.

In the conclusion of Section 4.6, we note that the estimated dynamic level coefficient corresponding to the second-order covariance effect does not exhibit significant variation over time, indicating that the dependence between the species stays on the same level during the observational time period. Therefore, we consider the level correlated models (LCMs) described in Section 5.2.

Let $\boldsymbol{Y}_{it} = (Y_{Cit}, Y_{Git}, Y_{Nit})'$. Throughout this section the capital letters C, G, and N stand for the three most abundant gastropod species *Caracolus caracolla*,*Gaeotis nigrolineata*, and *Nenia tridens* respectively. The observation equation corresponding to (5.2.9) and (5.2.11) for the level correlated model is:

$$Y_{jit}|\lambda_{jit} \sim \text{Poisson}(\lambda_{jit}), \tag{5.4.1}$$

where $i = 1, \ldots, n$ $(n = 111 + 40)$, $t = 1, \ldots, T$ $(T = 20)$ and $j = C, G, N$ $(J = 3)$. The natural logarithm is a link function for the random Poisson mean λ_{jit}. It is

modeled as follows:

$$\begin{aligned}
\log(\lambda_{jit}) &= \gamma_{jt} + \alpha_{jit} + \beta_{j0} + \boldsymbol{z}'_{jit}\boldsymbol{\beta}_j \\
&= \gamma_{jt} + \alpha_{jit} + \beta_{j0} + \beta_{j1}\log(\text{Elevation})_i \\
&+ \beta_{j2}\log(\text{Slope})_i + \beta_{j3}\log(\text{CO})_{it} + \beta_{j4}\log(\text{PREMON})_{it} \\
&+ \beta_{j5}\log(\text{PA})_{it} + \beta_{j6}\log(\text{DEAD})_{it} \\
&+ \beta_{j7}\text{I(Aspect=1)}_i + \beta_{j8}\text{I(Aspect=5)}_i + \beta_{j9}\text{I(Aspect=7)}_i \\
&+ \beta_{j10}\text{I(Cover Class=2)}_i + \beta_{j11}\text{I(Cover Class=3)}_i \\
&+ \beta_{j12}\text{I(Soil Type=2)}_i + \beta_{j13}\text{I(Soil Type=3)}_i \\
&+ \beta_{j14}\text{I(Litter Cover=2)}_{it} + \beta_{j15}\text{I(Litter Cover=3)}_{it} \\
&+ \beta_{j16}\text{I(Litter Cover=4)}_{it} + \beta_{j17}\text{I(Litter Cover=5)}_{it}, \qquad (5.4.2)
\end{aligned}$$

where $i = 1, \ldots, n$, $t = 1, \ldots, T$, $j = C, G, N$, and $\boldsymbol{z}'_{jit}$ is a p_j dimensional vector of covariates. Covariates that correspond to Elevation, Slope, Aspect, Soil Type, and Canopy Cover are location specific variables. Canopy Openness, Leaf Litter Cover, Dead Vegetation, Plant Apparency (PA), and Apparency of Sierra Palm (PREMON) vary over time and locations (time and location specific variables). In this setup, the level correlated term is α_{jit}. The $\boldsymbol{\alpha}_{it} = (\alpha_{Cit}, \alpha_{Git}, \alpha_{Nit})'$ is modeled according to the equation (5.2.4), i.e. $\boldsymbol{\alpha}_{it} \sim \text{Normal}(\mathbf{0}, \boldsymbol{\Sigma})$.

The term γ_{jt} is a species specific time effect and we assume a random walk evolution given by:

$$\gamma_{jt} = \gamma_{j(t-1)} + w_{jt}, \qquad (5.4.3)$$

where $t = 1, \ldots, T$ and $j = C, G, N$ and the error term $w_{jt} \sim \text{Normal}(0, 1/W_j)$. As mentioned in Section 5.2 the model (5.4.1)-(5.4.3) accounts for overdispersion. Moreover, it can explain positive or negative dependence between components of the response vector.

We use Integrated Nested Laplace Approximations (INLA) approach, which provides a mechanism for Bayesian inference based on accurate approximations to the posterior distributions of the parameters. Details are given in Section 2.2. For the approximate sampling based Bayesian framework, the usual prior specifications on the parameters are assumed here. We assume a Normal prior for β_{j0} and every component in $\boldsymbol{\beta}_j$ in equation (5.4.2), a Wishart prior for $\boldsymbol{\Sigma}$ in $\boldsymbol{\alpha}_{it}$ from (5.4.2), and a LogGamma prior for $\log(W_j)$. We use the default hyperparameter specifications in *inla* function from the R-INLA package, namely β_{j0} and every component in $\boldsymbol{\beta}_j$ from (5.4.2) have a Normal$(0, 10^3)$ prior. The matrix $\boldsymbol{\Sigma}$ has Wishart prior with $2 \times J + 1$ degrees of freedom and identity matrix as a prior precision matrix. For the state equation error precision INLA specifies $\log(W_j) \sim \text{LogGamma}(1, 5 \times 10^{-5})$. Since INLA does not rely on MCMC, the approximate approach greatly reduces computational time. The model takes about 7 minutes to run using R-INLA on Intel Core i5-2500 CPU 3.3 GHz with 8GB of RAM.

For comparison purposes, we also fit two univariate models, namely a dynamic Poisson Lognormal (DPLN) regression model and a negative binomial (NegBin) regression model. In our case the dynamic Poisson Lognormal model is an univariate version of the level correlated model (LCM) given by equations (5.4.1)-(5.4.3). We fit DPLN models separately to each component of $\boldsymbol{Y}_{it} = (Y_{Cit}, Y_{Git}, Y_{Nit})'$. For $i = 1, \ldots, n$, $t = 1, \ldots, T$, and $j = C, G, N$, the observation and the state equations stay the same as equations (5.4.1)-(5.4.3). The only difference is that in equation

(5.4.2), we assume independent structure for the level correlated term, meaning:

$$\alpha_{jit} \sim \text{Normal}(0, 1/\tau_j^2) \tag{5.4.4}$$

Thus, the models described in (5.4.1)-(5.4.3) and (5.4.4) are $J = 3$ independent dynamic Poisson Lognormal mixture models. Similar to the LCM, this regression models account for overdispersion. We perform model estimation through an approximate Bayesian framework using R-INLA and use the same prior and hyperprior specifications as given under the LCM framework.

The negative binomial regression model (Cameron and Trivedi, 1986) is defined as follows:

$$Y_{jit}|\lambda_{jit}, \delta_j \quad \sim \quad \text{NegBin}(\lambda_{jit}, \delta_j), \tag{5.4.5}$$

where $i = 1, \ldots, n$ $(n = 111 + 40)$, $t = 1, \ldots, T$ $(T = 20)$, $j = C, G, N$ $(J = 3)$, δ_j's are the overdispersion parameters and λ_{jit}'s denote means for the negative binomial distribution. The natural logarithm is a link function and the means can be written as follows:

$$\log(\lambda_{jit}) \quad = \quad \beta_{j0} + \boldsymbol{z}'_{jit}\boldsymbol{\beta}_j, \tag{5.4.6}$$

where all parameters are defined in the paragraph below the equation (5.4.2). We use the function *glm.nb* in the R package $MASS$ to perform the maximum likelihood estimation of coefficients for negative binomial regression described in equations (5.4.5)-(5.4.6). Note that this model assumes independence between the components, in other words we fit J separate regressions to each of the components of the response

TABLE 5.4.1: Posterior mean of correlation coefficients

Gastropod species	*Caracolus caracolla*	*Gaeotis nigrolineata*	*Nenia tridens*
Caracolus caracolla	1	0.141**	0.436***
Gaeotis nigrolineata		1	0.046
Nenia tridens			1

***, **, and * denote 99%, 95%, and 90% significance for NegBin model and if zero is outside 99%, 95%, and 90% Credible Intervals for multivariate LCM and univariate PLN models, respectively

vector.

5.4.3 Discussion of Results

Table 5.4.1 gives a posterior mean of the correlation coefficients defined in equation (5.3.1) of the matrix $\mathbf{\Sigma}$. We discover that *Caracolus caracolla* has a noticeable positive association with other two species. We observe that there is a moderate positive association between *Caracolus caracolla* and *Nenia tridens* with posterior mean 0.436 and 99% Credible Interval (CI) $(0.3307, 0.5287)$. There is also a weak positive association between *Caracolus caracolla* and *Gaeotis nigrolineata* with posterior mean 0.141 with 95% CI $(0.027, 0.2524)$. We can conclude that there is no indication of association between *Gaeotis nigrolineata* and *Nenia tridens*. The posterior mean for this case is estimated to be 0.046 with 90% CI $(-0.0478, 0.1395)$. The posterior means of diagonal elements of the matrix $\mathbf{\Sigma}$ are 0.0419, 0.4047, and 0.1478 respectively.

Table 5.4.2 and Table 5.4.3 give the posterior means of parameters of multivariate level correlated model (LCM), univariate dynamic Poisson Lognormal model (DPLN), and univariate negative binomial regression models (NegBin). Although there are slight numerical differences in the estimates, the results of LCM and DPLN models

TABLE 5.4.2: Estimated Coefficients for Continuous Covariates and Intercept

	Model					
Covariates	LCM		DPLN		NegBin	
Intercept.C	7.385		7.479		5.436	
Intercept.G	-6.986		-7.265		-9.276	
Intercept.N	-3.529		-4.406		-3.693	
logElevation.C	-1.199		-1.215		-0.838	
logElevation.G	1.216		1.26		1.664	
logElevation.N	0.671		0.817		0.865	
logSlope.C	0.126	*	0.126	*	0.186	**
logSlope.G	-0.028		-0.025		0.063	
logSlope.N	0.206	***	0.218	***	0.173	**
logCO.C	0.091		0.099		0.176	***
logCO.G	-0.103		-0.099		-0.384	***
logCO.N	-0.216	***	-0.217	***	-0.032	
logPREMON.C	-0.043	*	-0.04	*	-0.062	***
logPREMON.G	0.207	***	0.208	***	0.17	***
logPREMON.N	-0.054	**	-0.059	**	-0.1	***
logPA.C	0.066	**	0.062	**	0.035	
logPA.G	-0.064	**	-0.064	**	-0.113	***
logPA.N	-0.027		-0.033		-0.086	***
logDEAD.C	0.08	***	0.082	***	0.063	**
logDEAD.G	-0.113	***	-0.113	***	-0.163	***
logDEAD.N	0.153	***	0.151	***	0.081	**

***, **, and * denote 99%, 95%, and 90% significance for NegBin model and if zero is outside 99%, 95%, and 90% Credible Intervals for LCM and DPLN models, respectively

are very close to each other. We note that there are no discrepancies in the significance of the coefficients. On the other hand, the negative binomial model appears to have some discrepancies in the significance of the estimates. We expect this behavior as the negative binomial model accounts neither for time correlation nor the association between components of the response vector.

All three gastropods respond to previous land use in "similar" ways. The higher abundances are estimated in areas which were used for coffee plantations and with

selective logging (significant positive values for CoverClass=2 and CoverClass=3). Whereas the lower number of species are estimated to be within areas which experienced the most intensive logging and agriculture prior to 1934.

Variation in the topographic position of sites (Slope and Aspect) affects abundance of gastropods in a species-specific manner. *Caracollus* and *Nenia* respond similarly to such variation, whereas *Gaeotis* responds differently than the rest. The model estimates a significant and opposite sign for the coefficients which correspond to logSlope and Aspect. Such topographic variability affects abiotic and biotic characteristics that likely affect habitat quality and food availability.

Variation in apparency of vegetation also affects variation in abundances in a species-specific manner. *Caracollus* is associated with areas in which the understory is dense (positive values for logPA), but with little sierra palm (negative values for logPREMON). *Gaeotis* is associated with areas in which the understory is sparse, but with high apparency of sierra palm. *Nenia* is not associated with understory apparency of non-palms, but is associated with areas that have low apparency of sierra palm.

Variation in dead plant material (necromass) affects abundances in a species-specific manner. *Caracollus* and *Nenia* are each associated with areas with high necromass (positive coefficients for logDEAD). *Gaeotis* is associated with low necromass (negative coefficient for logDEAD).

In Figure 5.5.1 we plot temporal behavior of gastropods. The vertical red horizontal line denotes time of hurricane Georges. Hurricane Georges was a powerful Category 4 hurricane causing major damage to trees, widespread flooding, and dropped immense precipitation. Model predictions (posterior mean) accurately capture the species-specific temporal dynamics of gastropod species in tabonuco forest. The two

TABLE 5.4.3: Estimated Coefficients for Continuous Covariates

	Model					
Covariates	LCM		DPLN		NegBin	
Aspect=1.C	-0.046		-0.039		-0.034	
Aspect=1.G	0.071		0.073		0.015	
Aspect=1.N	-0.226	***	-0.21	**	-0.164	*
Aspect=5.C	0.458	***	0.46	***	0.425	***
Aspect=5.G	-0.278	***	-0.286	***	-0.264	**
Aspect=5.N	0.321	***	0.315	***	0.296	***
Aspect=7.C	0.289	***	0.297	***	0.303	***
Aspect=7.G	0.037		0.037		0.084	
Aspect=7.N	0.19	***	0.191	***	0.202	***
CoverClass=2.C	0.189	**	0.185	**	0.185	**
CoverClass=2.G	0.495	***	0.501	***	0.654	***
CoverClass=2.N	0.606	***	0.592	***	0.684	***
CoverClass=3.C	0.091		0.088		0.078	
CoverClass=3.G	0.385	***	0.393	***	0.467	***
CoverClass=3.N	0.547	***	0.546	***	0.568	***
SoilType=2.C	0.779	***	0.77	***	0.761	***
SoilType=2.G	-0.408	***	-0.416	***	-0.407	***
SoilType=2.N	0.384	***	0.371	***	0.342	***
SoilType=3.C	1.636	***	1.622	***	1.491	***
SoilType=3.G	0.214		0.202		0.414	**
SoilType=3.N	0.849	***	0.811	***	0.673	***
LitterCover=2.C	0.196	**	0.194	**	0.121	
LitterCover=2.G	0.003		0.006		-0.024	
LitterCover=2.N	-0.103		-0.101		-0.371	***
LitterCover=3.C	0.301	***	0.299	***	0.194	**
LitterCover=3.G	0.097		0.102		0.131	
LitterCover=3.N	0.102		0.102		-0.246	**
LitterCover=4.C	0.288	***	0.285	***	0.183	*
LitterCover=4.G	0.272	**	0.277	**	0.323	**
LitterCover=4.N	0.144		0.148		-0.122	
LitterCover=5.C	0.315	**	0.306	**	0.195	
LitterCover=5.G	0.196		0.2		0.041	
LitterCover=5.N	0.512	***	0.517	***	0.129	

***, **, and * denote 99%, 95%, and 90% significance for NegBin model and if zero is outside 99%, 95%, and 90% Credible Intervals for LCM and DPLN models, respectively

taxa that live in leaf litter and consume necromass (*Caracolus caracolla* and *Nenia tridens*) show parallel trajectories, gradually increasing in abundance after Hurricane Georges. In contrast, the species that live on the undersides of palm leaflets (*Gaeotis nigrolineata*) quickly increases in abundance after Hurricane Georges, ostensibly because of an increase in palm apparency associated with hurricane-induced creation of canopy openings throughout the forest, and the population fluctuates more erratically thereafter.

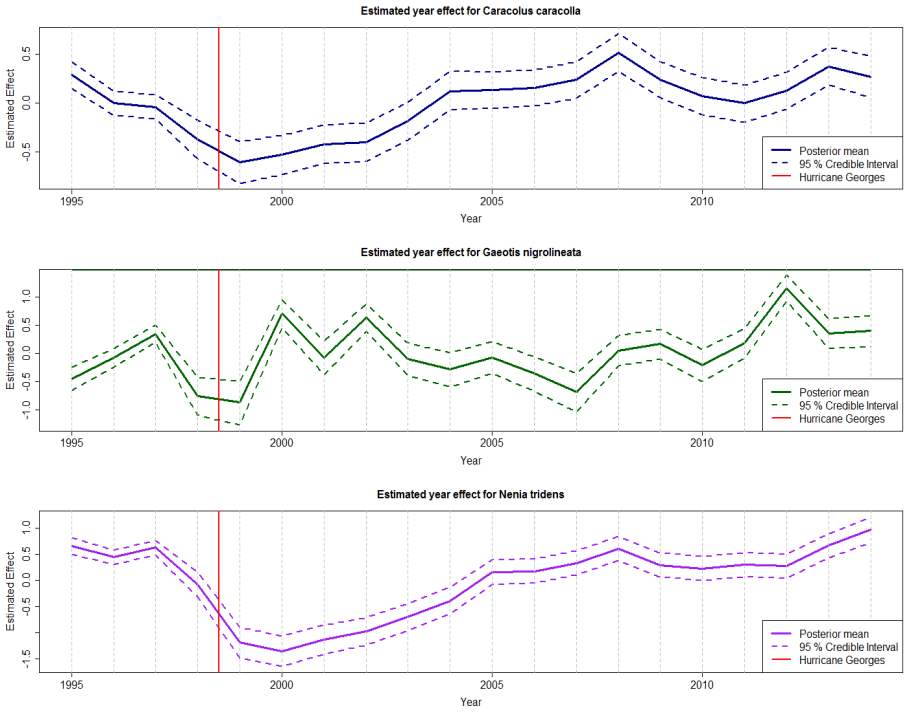

FIGURE 5.4.2: Estimated time trend for gastropod species

To evaluate the predictive ability of the models, we compute and compare the Predicted Mean Absolute Error (PMAE) which is calculated based on out-of-sample

(hold-out) observations according to two scenarios. In the first scenario, we remove counts of *Nenia tridens* on one particular location and build models on the rest of the data. In the second scenario, we remove counts for all three gastropod species on one particular location and use the rest of the data for the model estimation. To avoid the possibility of bias in the random sampling procedure, we repeat the entire procedure on all 40 locations where observations are available during the entire 1995-2014 time period. We report the average values for the predictive capabilities of models. The average PMAE can be computed as follows:

$$\text{Average PMAE}_j = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}|\widehat{Y}_{jit} - Y_{jit}|\right), \tag{5.4.7}$$

where $i = 1, \ldots, n$ ($n = 40$), $t = 1, \ldots, T$ ($T = 20$) and $j = C, G, N$ ($J = 3$). The term $\widehat{Y}_{jit}$'s represent fitted values from multivariate LCM (posterior mean), dynamic PLN (posterior mean), and negative binomial regression (point estimate of mean).

Table 5.4.4: PMAE in Scenario 1

PMAE	LCM	DPLN	NegBin
Nenia tridens	5.020	5.246	6.209

From results in Table 5.4.4 we can conclude that the multivariate LCM has better predictive abilities than both the dynamic PLN and the negative binomial models. This is in line with the results of the estimation for correlation coefficients between species reported in Table 5.4.1. All three gastropod species have some positive association among themselves, in particularly *Nenia tridens* has moderate positive association with *Caracolus caracolla* and some weak association with *Gaeotis nigrolineata*. Thus, we expect that association with *Caracolus caracolla* and *Gaeotis nigrolineata* counts increases the predictive ability of the multivariate LCM.

Table 5.4.5: PMAE in Scenario 2

PMAE	LCM	DPLN	NegBin
Caracolus caracolla	4.568	4.570	4.688
Gaeotis nigrolineata	1.789	1.789	2.015
Nenia tridens	5.243	5.246	6.209

In Table 5.4.5 we report PMAE's under the second scenario where counts for all three gastropod species are removed on one particular location. We observe that there is no considerable difference in the predictions between the multivariate LCM and the univariate dynamic PLN. Nevertheless, the multivariate LCM and the dynamic PLN over perform traditional negative binomial regression model in both scenarios.

5.5 Marketing Modeling Using LCM

5.5.1 Data Description

We describe the statistical analysis pertaining to marketing data from a large multinational pharmaceutical firm. This marketing research is concerned with the analysis of drivers for new prescriptions written by physicians. Most of the existing research focuses on physician level sales for a single drug within a therapeutic category. They do not consider the association between the sales of a drug and its competitors over time. See Venkatesan et al. (2012) for a detailed discussion. Furthermore, there is interest in knowing the effect of the firm's detailing efforts (sales visits to physicians) on the sales of its own drug and on the sales of its competitors. Mizik and Jacobson (2004) discussed existing research showing that detailing, sampling, and past behavior influence the prescription counts from physicians. Montoya et al. (2010) state that

after accounting for dynamics in physician's prescription writing behavior, detailing seems to be the most effective in acquiring new physicians, whereas sampling (i.e., providing samples of drugs to physicians) is most effective in obtaining recurring prescriptions from existing physicians. As do most other research studies in this context, we treat physicians as customers of the pharmaceutical firm.

The behavioral data collected monthly by the firm over a period of three years consists of the number of new prescriptions written by each physician (sales) and the number of sales calls directed toward such physicians (detailing). As in Venkatesan et al. (2012), our focus is on one of the newer drugs launched by the firm in a large therapeutic drug category (one of the ten largest therapeutic categories in the United States). The database consists of monthly prescription history for 43 continuous months within the last decade from a sample of physicians from the American Medical Association (AMA) database. The time window of our data starts one year after the introduction of the focal drug.

For our analysis, we have chosen three drugs in the same therapeutic category with the highest market shares. The focal drug is a drug made by the firm of interest with a market share of 13%. The leader drug has a market share of 47% and a challenger drug has a 15% market share. We denote the focal drug by the abbreviation "F", the leader drug is denoted as "L", and the challenger drug as "C". Let $\boldsymbol{Y}_{it} = (Y_{F,it}, Y_{L,it}, Y_{C,it})'$ be a 3-dimensional vector of count responses of the number of new prescriptions for each drug written by the ith physician at equally spaced times t, for $i = 1, \ldots, n$ and $t = 1, \ldots, T$.

Our analysis is based on data from $n = 5300$ physicians, and Figure 5.5.1 shows the time series of prescription counts for the focal, leader, and challenger drugs for a randomly chosen physician. Also Figure 5.5.1 shows the only available predictor,

i.e., detailing, which is the number of the sales calls to the physician from the firm's representative regarding the focal drug. Due to confidentiality concerns, we are unable to reveal any other information about the drug category or the pharmaceutical firm.

We are interested in modeling patterns in the number of prescriptions written by the physician for the focal drug, as well as for the leader and challenger drugs. Further, we want to estimate the probability of churn, also known as customer attrition, i.e., the probability that the physician stops writing prescriptions for the focal drug to patients.

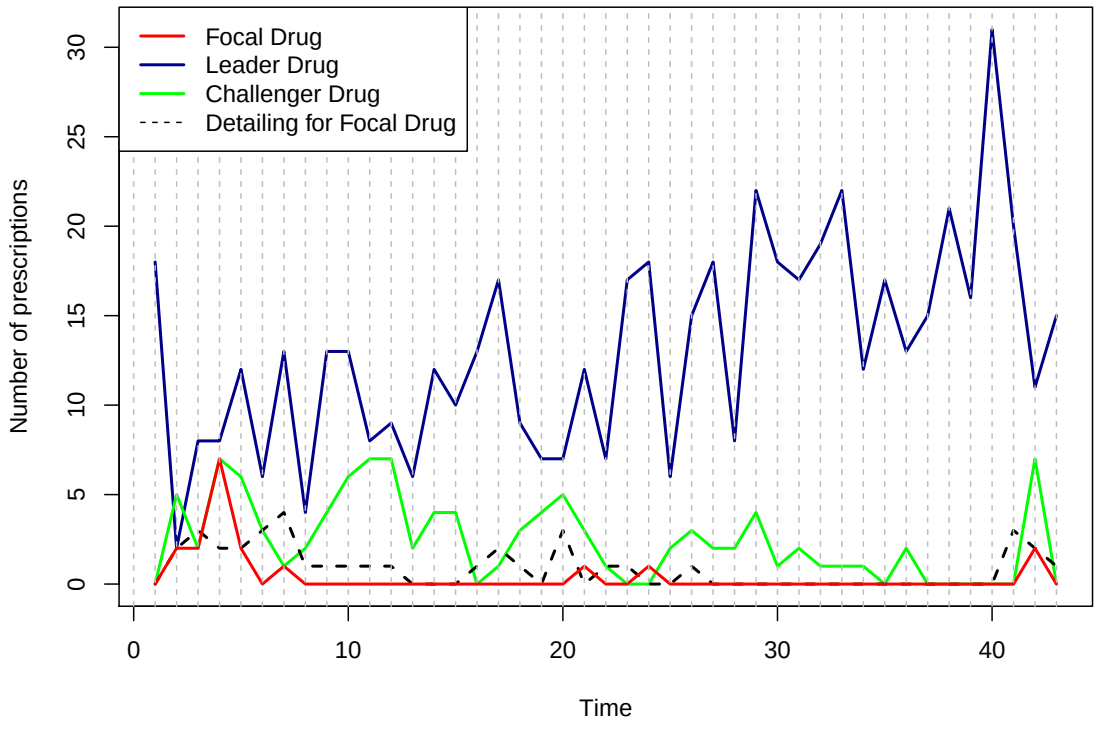

FIGURE 5.5.1: Prescription counts of three drugs and detailing for a randomly selected physician (ID# 2910)

Exploratory data analysis shows that while the firm obtains on average three new

prescriptions per month from a single physician, and on average salespeople make a call to a physician about twice a month, there is large variation in both the monthly level of sales per physician and the number of sales calls for the focal drug directed toward the physician each month. Sale calls seem to have a positive correlation with sales of the focal drug.

The focal drug represents a significantly different chemical formulation, and further targets a different function of the human body to cure the disease condition than the drugs available at the time of its introduction in the therapeutic category. It is therefore reasonable to expect that physicians will learn about the efficacy of the drug over time, resulting in a variation (either increase or decrease) in sales over time. This expectation is supported by multiple exploratory analyses of sales histories. We observe that the average level of sales (across all physicians) ranges from about 1 in the first month to 4 in the last month. An ANOVA test, reported in Venkatesan et al. (2012) rejected the null hypothesis that the mean level of sales was the same across the months.

The variation in the number of sales over time has motivated us to develop a dynamic model framework where the coefficients in the customer level sales response model could vary across customers and over time.

High and nearly infeasible computational requirements of the fully Bayesian approach based on multivariate Poisson mixtures for the large n prompted us to explore an approximate Bayesian framework for the LCM. We describe the level correlated model, which provides a useful framework for studying the evolution of sales of competing drugs (Focal, Leader, and Challenger drugs) within one therapeutic category. We use this model to decompose the association in sales among competing drugs and investigate trends induced by general industry behavior.

5.5.2 Model Framework

With such a large and diverse dataset (5300 physicians), we anticipate that some physicians would have quite different temporal behavior of drug sales than others. On the other hand, we would wish to aggregate all information about associations between drugs and estimate common model parameters for the physicians, that exhibit similar behavior. In Section 5.2 we show that with small time series, the estimated credible intervals under LCMs tend to be wide. Therefore, we anticipate that most, if not all, of the estimated coefficients will have non-significant effects. We explore an alternate approach that is more suitable for large data sets.

To proceed further, we divide our analysis into three parts: pre-clustering, clustering, and post-clustering analyses. First we fit $n = 5300$ independent level correlated models for each physician in order to estimate their individual temporal behavior. Then, we use the K-means procedure to cluster physicians according to their temporal behavior and choose K clusters. After that, we fit the LCMs within the relatively small homogeneous clusters of physicians. The fitted LCMs assume that the effects of covariates and association parameters are the same across all physicians within any given cluster. We do include a drug specific intercept in the model, meaning that this intercept varies with each component of the response vector.

For the pre-clustering part of the analysis, we fit LCMs for each physician. The general framework is given in Section 5.2. The following equations give the specific details about the fitting. The observation equation of the Level Correlated Model (LCM) for multivariate time series in the hierarchical dynamic framework can be

written as follows:

$$Y_{jit}|\lambda_{jit} \sim \text{Poisson}(\lambda_{j,it}), \tag{5.5.1}$$

where $i = 1, \ldots, n$ ($n = 5300$), $t = 1, \ldots, T$ ($T = 43$), $j = F, L, C$ ($J = 3$), and λ_{jit}'s are means of Poisson distributions (corresponding to the focal, leader, and challenger drugs). Using the natural logarithm link function for the parameter λ_{jit}, we model λ_{jit} as a function of predictors as follows:

$$\log(\lambda_{jit}) = \gamma_{jit} + \beta_{ji0} + \boldsymbol{z}'_{jit}\boldsymbol{\beta}_{ji} + \alpha_{jit}, \tag{5.5.2}$$

where $i = 1, \ldots, n$, $t = 1, \ldots, T$ and $j = F, L, C$. In (5.5.2), the random effect β_{ji0} is a drug and physician specific intercept for the number of prescriptions, γ_{jit} represents a drug and physician specific time effect, the random effect α_{jit} is a drug, time, and physician specific level correlated component, the vector $\boldsymbol{z}'_{jit}$ denotes a p_j-dimensional vector of covariates, and $\boldsymbol{\beta_{ji}}$ is a p_j-dimensional vector of coefficients corresponding to the predictors. Here, $\boldsymbol{z}_{jit} = (\log(\text{Detailing})_{F,it}, 0, 0)'$ and $\boldsymbol{\beta_{ji}} = \boldsymbol{(\beta_{F,i}, 0, 0)'}$, where $\beta_{F,i}$ is the fixed effect coefficient for the natural logarithm of detailing, i.e., the number of the sales calls to the physician from the firm's representative regarding the focal drug. A small correction term is added to avoid taking logarithms of zeros.

Let $\boldsymbol{\alpha}_{it} = (\alpha_{F,it}, \alpha_{L,it}, \alpha_{C,it})'$. In (5.5.2), the dependence between different types of counts in (5.5.1) is introduced at the physician level through $\boldsymbol{\alpha}_{it} \sim \text{Normal}(\boldsymbol{0}, \boldsymbol{\Sigma_i})$, where $\boldsymbol{\Sigma_i}$ is a physician specific variance-covariance matrix for the level correlated random effect term. Let ρ_{iFL}, ρ_{iFC}, and ρ_{iLC} denote the correlation coefficients between the focal and leader, the focal and challenger, and the leader and challenger

drugs, respectively. We assume that the physician specific random intercept follows a normal distribution, i.e., $\beta_{ji0} \sim \text{Normal}(0, 1/\tau_i)$. We also assume that the components of the drug and physician specific time effect vector $\boldsymbol{\gamma}_{it} = (\gamma_{F,it}, \gamma_{L,it}, \gamma_{C,it})'$ are independent and evolve according to a random walk process in the state equation of the hierarchical dynamic model given by:

$$\gamma_{jit} = \gamma_{ji(t-1)} + w_{jit}, \tag{5.5.3}$$

where $i = 1, \ldots, n$, $t = 1, \ldots, T$, $j = F, L, C$ and the error term $w_{jit} \sim \text{Normal}(0, 1/W_i)$.

We implement the model estimation through an approximate sampling based Bayesian framework, assuming usual prior specifications for the parameters. We assume a normal prior for $\beta_{F,i}$ in (5.5.2), a Wishart prior for $\boldsymbol{\Sigma_i}$ in the distribution of $\boldsymbol{\alpha}_{it}$ and a log gamma prior for $\log(\tau_i)$ and $\log(W_i)$ in the distribution of β_{ji0} and w_{jit} respectively in (5.5.2) and (5.5.3). We use the INLA approach (Rue et al. (2009)) which provides a mechanism for Bayesian inference based on accurate approximations to the posterior distributions of the parameters. We have slightly modified the standard setups provided in R-INLA, and therefore this approach is easily implementable from the end-user point of view.

We use the default hyperparameter specifications in the *inla* function from the R-INLA package for β_{ji0}, $\boldsymbol{\beta}_{ji}$, and $\boldsymbol{\Sigma_i}$. All the relevant values are the same as given in the paragraph after equation (5.4.3). We also explicitly specify the initial value of the precision parameters in (5.5.3) to be equal to 10. The sample data ($n = 5300$, $T = 43$ for 3 drugs) is divided into 408 cluster jobs. Each node has 2 x 4-core AMD Opteron 2350 processors (2 GHz) with 8 GB of memory. The total running time is about 4 hours. For brevity, we skip the discussion about estimated results here and

move to the clustering part of the analysis.

We use the estimated temporal behavior of physicians on drug sales from the Level Correlated Model (LCM) to cluster the physicians. The most well-known and the simplest clustering method is the K-means algorithm. This algorithm has a rich and diverse history as it was independently discovered in different scientific fields. For more in depth discussion and details see Steinhaus (1956), Ball and Hall (1965), Forgy (1965), MacQueen (1967), Hartigan and Wong (1979), Lloyd (1982), and Jain (2010). Before applying the K-means algorithm we scale the estimated time trends from the pre-clustering part. The K-Means algorithm finds the best division of n entities into k groups. In general this is done by minimizing the total distance between the group's members and the corresponding centroid. Formally, we minimize the within-cluster sum of squares. The function *kmeans* in R is used with default specifications that uses the Hartigan and Wong algorithm (Hartigan and Wong, 1979). This algorithm randomly selects the initial cluster centers and allocates to the cluster with the smallest mean. The within-cluster sum of squares under this setup is defined as:

$$\text{Sum}(k) = \sum_{i=0}^{n} \sum_{j=0}^{p} (x(i,j) - \overline{x(k,j)})^2, \tag{5.5.4}$$

where $\overline{x(k,j)}$ is the mean variable j of all elements in the group k. The algorithm iteratively searches for the optimal within-cluster sum of squares by moving points from one cluster to another. For the optimal number of clusters we use the Bayesian

information criterion (BIC) computed as follows (Schwarz, 1978):

$$\text{BIC} = \sum_{k=1}^{K} \text{Sum}(k) + mK \log(n), \tag{5.5.5}$$

where $\text{Sum}(k)$ is the within-cluster sum of squares, K is a total number of clusters, n is the number of points, and m is the number of variables (in our case $m = 43 \times 3 = 129$).

In Figure 5.5.2 we plot the values of BIC for $K = 1, \ldots, 150$. The values of BIC are quickly decreasing to somewhere around $K = 40$ (the red line). Afterwards, BIC stays on the same level until around $K = 70$ and increases for $K > 70$. Thus, we decided to choose $K = 40$ clusters for further analysis.

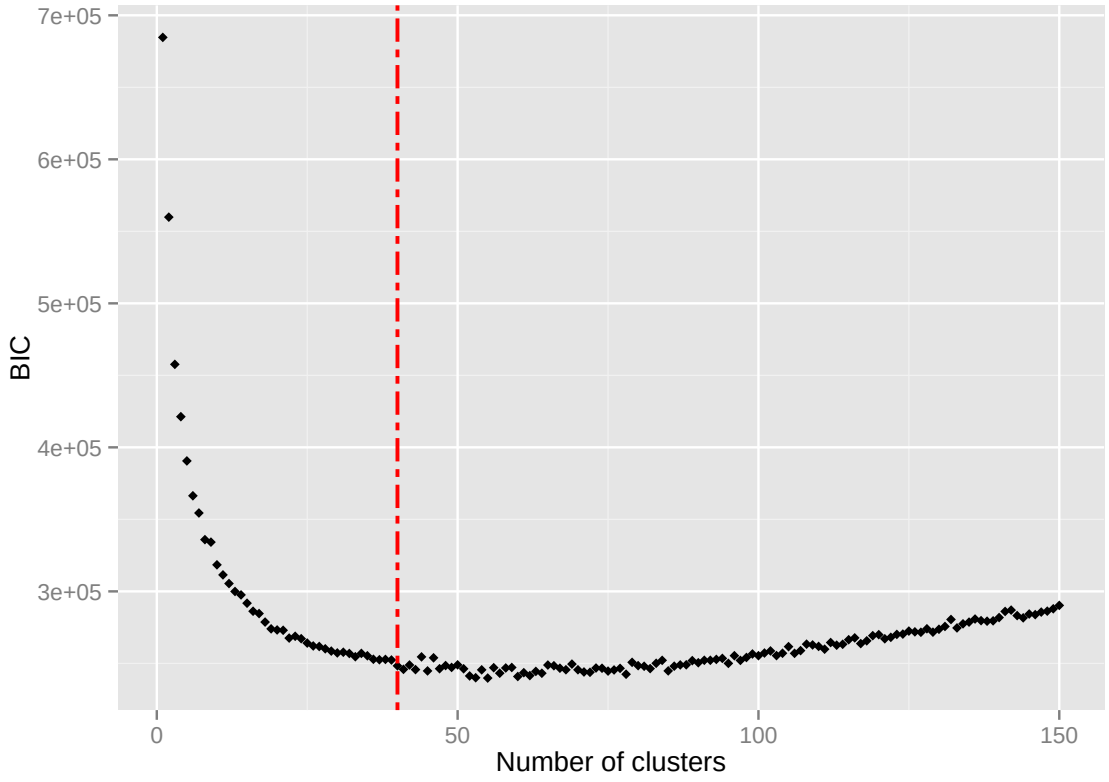

FIGURE 5.5.2: BIC values vs number of clusters

In the post-clustering part of the analysis we compare three LCMs that combine different marginal distributions. We note that the three models have different observation equations under the LCM. The Poisson model stands for the LCM which has the Poisson distribution for all marginal distributions as in equation (5.2.9). Also, let the Mixed model denote the LCM for which the focal and challenger drugs are assumed to follow the zero inflated Poisson (ZIP) distribution, whereas a Poisson distribution is assigned for the leader drug. And finally, a ZIP model describes the LCM with zero inflated Poisson marginal distributions for all three drugs. Thus the observation equations for all three models can be written as follows:

Poisson model:

$$\begin{aligned} Y_{F,it}|\lambda_{F,it} &\sim \text{Poisson}(\lambda_{F,it}), \\ Y_{L,it}|\lambda_{L,it} &\sim \text{Poisson}(\lambda_{L,it}), \\ Y_{C,it}|\lambda_{C,it} &\sim \text{Poisson}(\lambda_{C,it}) \end{aligned} \tag{5.5.6}$$

Mixed model:

$$\begin{aligned} Y_{F,it}|\pi_F, \lambda_{F,it} &\sim \text{ZIP}(\pi_F, \lambda_{F,it}), \\ Y_{L,it}|\lambda_{L,it} &\sim \text{Poisson}(\lambda_{L,it}), \\ Y_{C,it}|\pi_C, \lambda_{C,it} &\sim \text{ZIP}(\pi_C, \lambda_{C,it}) \end{aligned} \tag{5.5.7}$$

ZIP model:

$$\begin{aligned} Y_{F,it}|\pi_F, \lambda_{F,it} &\sim \text{ZIP}(\pi_F, \lambda_{F,it}), \\ Y_{L,it}|\pi_L, \lambda_{L,it} &\sim \text{ZIP}(\pi_L, \lambda_{L,it}), \\ Y_{C,it}|\pi_C, \lambda_{C,it} &\sim \text{ZIP}(\pi_C, \lambda_{C,it}), \end{aligned} \tag{5.5.8}$$

where $i = 1, \ldots, n$ ($n = 5300$), $t = 1, \ldots, T$ ($T = 43$), and λ_{jit}'s ($j = F, L, C$) are means of Poisson distributions. We model λ_{jit} as a function of predictors as follows:

$$\log(\lambda_{jit}) = \gamma_{jt} + \beta_{j0} + \boldsymbol{z}'_{jit}\boldsymbol{\beta}_j + \alpha_{jit}, \tag{5.5.9}$$

where all the terms are defined as in the paragraph below the equation (5.5.2). Here the term γ_{jt} is a drug specific time effect. We assume a random walk evolution given by:

$$\gamma_{jt} = \gamma_{j(t-1)} + w_{jt}, \tag{5.5.10}$$

where $t = 1, \ldots, T$ and $j = F, L, C$ and the error term $w_{jt} \sim \text{Normal}(0, 1/W_j)$.

Again we implement the model estimation through an approximate Bayesian framework with INLA methods, assuming the usual prior specifications for the parameters. In addition to the priors defined in the pre-clustering case, we assume normal priors for $\text{logit}(\pi_F)$, $\text{logit}(\pi_L)$, and $\text{logit}(\pi_C)$ in (5.5.7) and (5.5.8).

We use the deviance information criteria (DIC) and predicted mean absolute error (PMAE) criteria for the model comparison of each cluster. We drop the notation of

k for clusters to avoid confusion. The DIC (Spiegelhalter et al., 2002) is a popular criteria that can be used for dynamic hierarchical models. The DIC is defined as:

$$\begin{aligned} D &= -2\log(p(y|\theta)) + \text{constant}, \\ \text{DIC} &= 2\overline{D} - D(\overline{\theta}) = D(\overline{\theta}) + 2p_D, \end{aligned} \tag{5.5.11}$$

where D is the deviance, y is data, θ represents unknown parameters, $p(y|\theta)$ is a likelihood function. The constant cancels out in all calculations, thus it can be discarded. The term p_D is called the effective number of parameters for which INLA finds a good approximation (Rue et al., 2009).

To evaluate the predictive ability of the models, we compute and compare the Predicted Mean Absolute Error (PMAE) which is calculated based on out-of-sample (hold-out) observations. We take the last two time points as the hold-out observations. The PMAE can be computed as follows:

$$\text{PMAE}_j = \frac{1}{n}\sum_{i=1}^{n^*}\left(\frac{1}{T}\sum_{t=1}^{T}|\widehat{Y}_{jit} - Y_{jit}|\right), \tag{5.5.12}$$

where $i = 1, \ldots, n^*$ (n^* stands for number of physicians in the kth cluster), $t = 1, \ldots, T$ ($T = 2$) and $j = F, L, C$. The terms $\widehat{Y}_{jit}$'s represent posterior means from Poisson, Mixed, and ZIP level correlated models.

Table 5.5.1 gives the DIC and PMAE results for Poisson, Mixed, and ZIP models defined by equations (5.5.6), (5.5.7), and (5.5.8) respectively. The Poisson model has the lowest value of DIC for almost all clusters, except for the largest cluster. Here the Mixed model has the lowest DIC. The PMAE's are also lower for the Poisson model. We note that values are close to each other and for some clusters PMAE values are

almost identical. One of the benefits of using clusters is that the investigator can concentrate on one of the clusters and choose the suitable model.

From our point of view, the ZIP model is the most general among the three models. We are particularly interested in the estimated probabilities of zeros under the ZIP model given by equation (5.5.8). Thus, we report the estimated values for this model. In Figures 5.5.3 we plot the posterior means for π_F, π_L, and π_C under the ZIP level correlated model. We can conclude that in most of the clusters the probability of zero fluctuates around 0.25 and there is no difference between Focal, Leader, and Challenger drugs. Nevertheless, we notice that the cluster numbered 37 has the lowest estimated probabilities of zeros. This suggests that physicians from cluster 37 have a very low probability of not writing prescriptions in the given month. There are a couple of clusters that have different estimated probabilities of zeros for three drugs. Overall we can conclude that the Leader drug has the lowest probabilities of zeros across all clusters.

Figure 5.5.4 gives the posterior means for ρ_{FL}, ρ_{FC}, and ρ_{LC} as defined in equation (5.3.1). We can conclude that in most of the clusters the correlations between counts of different drugs are close to zero. However, we can notice that some clusters have the estimated correlations different from zero and the range approximately spans from 0.45 to -0.45. In particular, the physicians from cluster 8 have moderate negative association between Leader and Challenger drugs, and small negative association between Focal and Challenger drugs.

Table 5.5.2 gives the posterior results for the only available covariate Log(Detailing). In most clusters Detailing is positively associated with the number of prescriptions written by the physician. The posterior means fluctuate around 0.5. In few clusters Log(Detailing) does not have significant effect on the number of prescriptions

TABLE 5.5.1: DIC and PMAE comparison

Cluster	Number of members	Poisson Model		Mixed Model		ZIP Model	
		DIC	PMAE	DIC	PMAE	DIC	PMAE
1	90	52678	3.373	54566	3.775	53856	3.452
2	52	27316	3.11	27721	3.264	28534	3.272
3	137	75853	3.908	77600	3.986	80728	3.98
4	40	19917	3.917	20317	4.056	20692	4.026
5	54	29639	4.094	30164	4.161	30835	4.163
6	18	9317	2.686	9589	2.832	9750	3
7	33	17069	1.98	17300	2.028	17713	2.139
8	11	3811	3.733	3884	4.142	3937	4.126
9	24	10892	4.079	11146	4.343	11376	4.258
10	4	1779	0.149	1773	0.154	1781	0.156
11	17	7211	1.223	7351	1.237	7469	1.216
12	138	82620	4.432	85468	4.513	88533	4.51
13	18	8350	3.393	8590	3.585	8797	3.577
14	37	21124	4.156	21604	4.262	22339	5.211
15	21	10542	3.853	10544	3.858	10988	4.024
16	27	10919	3.871	11158	4.069	11284	4.08
17	40	19005	2.467	19227	2.485	20160	2.477
18	15	6288	3.621	6420	3.706	6535	3.749
19	89	54947	3.451	56583	3.766	58297	3.737
20	55	31288	4.64	32002	4.672	33433	4.693
21	14	7330	2.807	7549	2.898	7789	3.102
22	29	16090	4.503	16441	4.529	16428	4.48
23	1	130	0.151	126	1.925	128	2.596
24	16	7200	3.103	7360	3.205	7356	3.235
25	77	44588	3.096	45801	3.234	47275	3.214
26	56	30418	2.518	31045	2.567	32389	2.612
27	132	80138	4.808	82801	5.651	85774	5.65
28	51	28802	4.367	29217	4.462	30677	4.433
29	343	206754	4.339	210362	4.421	219948	7.131
30	134	80447	3.581	83228	3.602	86330	3.593
31	57	35880	4.819	37005	4.882	37997	4.874
32	116	72616	4.603	74416	4.764	77238	5.449
33	46	26774	3.17	26870	3.193	28761	3.233
34	47	27795	3.194	28174	3.214	29544	3.252
35	2855	1673979	2.152	1638671	4.499	1993518	4.482
36	32	17912	3.434	18422	3.49	18673	3.523
37	29	13207	0.961	13222	0.965	13214	0.963
38	204	127845	4.014	132833	4.08	137647	4.081
39	35	17540	3.966	17876	4.191	18587	4.096
40	106	62559	4.694	64023	4.856	66290	5.318

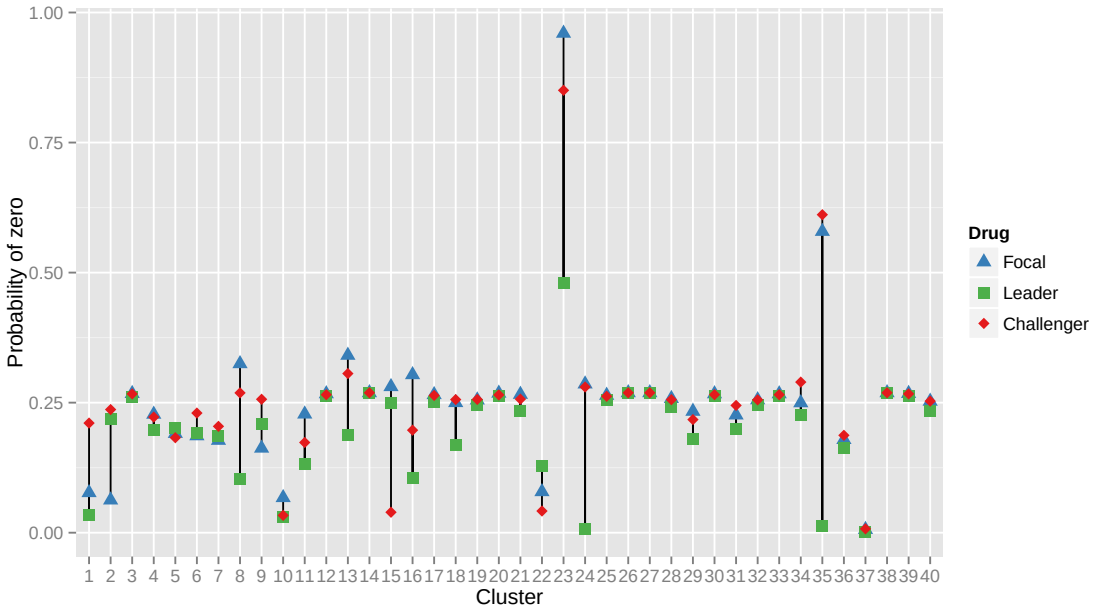

FIGURE 5.5.3: Posterior mean for probability of zero in ZIP Model

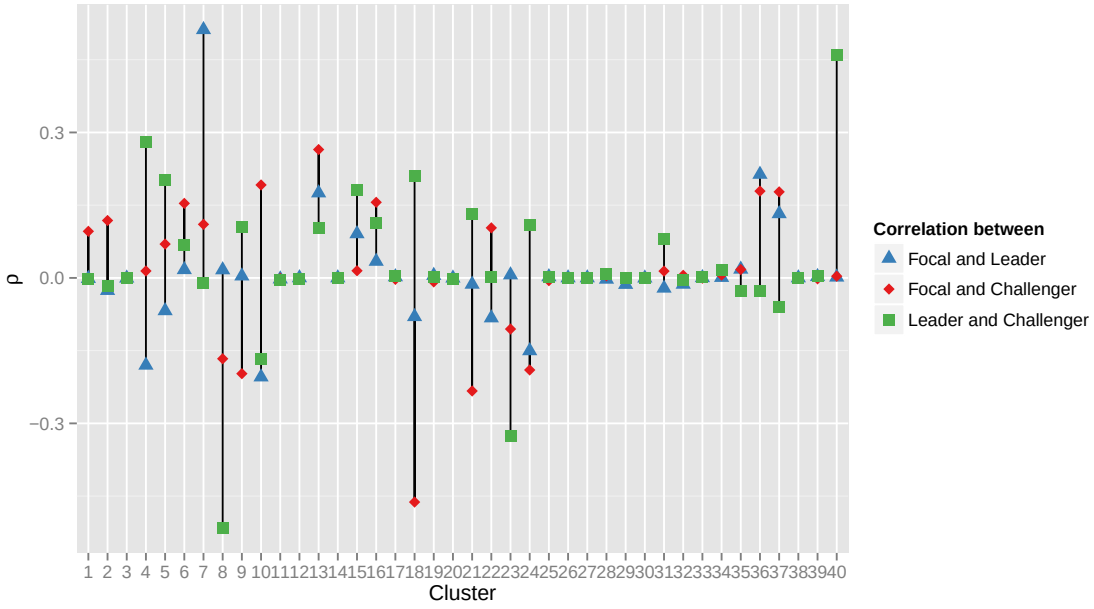

FIGURE 5.5.4: Posterior mean for ρ in ZIP Model

(clusters 2, 10, 11, and 23).

Overall, we can conclude that the posterior results follow expected behavior. Moreover, we have demonstrated that the class of LCMs can be used in these various situations. We show that the framework is very flexible and can incorporate different structures. One such structure can mix different marginal distributions. This may result in a better model fit, as well as give the researcher the flexibility to include extra parameters associated with different models (similarly to the Mixed or ZIP model setup).

TABLE 5.5.2: Posterior results for Log(Detailing) in ZIP Model

Cluster	Number of members	Posterior Estimates				
		mean	sd	0.025quant	0.5quant	0.975quant
1	90	0.469	0.029	0.412	0.469	0.526
2	52	0.01	0.05	-0.088	0.01	0.108
3	137	0.876	0.03	0.817	0.876	0.935
4	40	0.261	0.051	0.162	0.261	0.361
5	54	0.397	0.039	0.321	0.397	0.474
6	18	0.998	0.07	0.861	0.998	1.136
7	33	0.503	0.06	0.386	0.503	0.62
8	11	0.303	0.122	0.063	0.303	0.541
9	24	0.429	0.076	0.278	0.429	0.579
10	4	0.094	0.148	-0.198	0.095	0.384
11	17	0.168	0.086	-0.001	0.168	0.336
12	138	0.701	0.03	0.642	0.701	0.76
13	18	0.608	0.087	0.438	0.607	0.778
14	37	0.484	0.044	0.397	0.484	0.571
15	21	1.006	0.078	0.854	1.006	1.159
16	27	0.514	0.072	0.373	0.514	0.654
17	40	0.744	0.065	0.616	0.744	0.872
18	15	0.741	0.092	0.559	0.741	0.922
19	89	0.315	0.035	0.247	0.315	0.383
20	55	0.458	0.054	0.352	0.458	0.564
21	14	0.411	0.077	0.259	0.411	0.561
22	29	0.608	0.058	0.495	0.607	0.721
23	1	-0.154	1.002	-2.196	-0.126	1.735
24	16	0.285	0.074	0.139	0.285	0.431
25	77	0.805	0.037	0.732	0.805	0.878
26	56	0.749	0.044	0.661	0.749	0.836
27	132	0.741	0.029	0.683	0.741	0.798
28	51	0.576	0.05	0.477	0.576	0.675
29	343	0.334	0.007	0.321	0.334	0.347
30	134	0.421	0.032	0.358	0.421	0.483
31	57	0.447	0.038	0.371	0.447	0.522
32	116	0.507	0.03	0.449	0.507	0.566
33	46	0.53	0.051	0.429	0.53	0.631
34	47	0.497	0.043	0.412	0.497	0.582
35	2855	0.501	0.004	0.494	0.501	0.508
36	32	0.243	0.048	0.148	0.243	0.337
37	29	0.383	0.072	0.243	0.383	0.524
38	204	0.451	0.024	0.404	0.451	0.499
39	35	0.466	0.068	0.332	0.466	0.6
40	106	0.405	0.031	0.345	0.405	0.466

Chapter 6

Future Work

One of the natural extension of the LCM framework is to incorporate spatial dependence between locations where the counts are observed. Let $\boldsymbol{Y}_{it} = (Y_{1it}, \ldots, Y_{Jit})'$ be a J-variate vector of count responses, for $i = 1, \ldots, n$ and $t = 1, \ldots, T$. This means that we observe J types of counts on the n segments/locations over equally spaced T time points.

The observation equation of the dynamic model is:

$$Y_{jit}|\lambda_{jit} \sim \text{Poisson}(\lambda_{jit}), \tag{6.0.1}$$

$$\log(\lambda_{jit}) = \gamma_{jt} + \boldsymbol{z}'_{jit}\boldsymbol{\beta}_j + \xi_i - \alpha_{jit}, \tag{6.0.2}$$

where $i = 1, \ldots, n$, $t = 1, \ldots, T$ and $j = 1, \ldots, J$. In (6.0.2), the random effect γ_{jt} corresponds to the time effect, α_{jit} is the level correlated term, the vector $\boldsymbol{z}'_{jit}$ denotes a p_j dimensional vector of covariates with a vector of 1's as a first column, $\boldsymbol{\beta_j}$ is a p_j-dimensional vector of coefficients for covariates. In general $\boldsymbol{\beta}_j$ can be

subject/location specific as well as time-varying. The spatial random effect ξ_i can be modeled in various forms. For simplicity we consider the conditional autoregressive model proposed by Besag (Besag, 1974; Besag and Kooperberg, 1995), which can be written using a neighborhood structure as follows:

$$\xi_i|\xi_k, \tau, i \neq k \sim N\big(\frac{1}{m_i}\sum_{i\sim k}\xi_k, \frac{1}{m_i\tau}\big), \tag{6.0.3}$$

where $i = 1, \ldots, n$, m_i is the number of neighbors of segment/location i, and $i \sim k$ indicates that two segments i and k are neighbors. The random effect α_{jit} is modeled as under the LCM setup in Section 5.2.

Also as we have mentioned before, LCMs can account for overdispersion in the data. We can explore the broader class of marginal count distributions. The underlying distribution of counts should not be limited to the Poisson distribution as the data may fit better for example, if the underlying distribution is the negative binomial distribution (Zhou et al., 2012). Taking this into account we propose more general models where the underlying distribution may vary for different types of response counts. The HDLM and LCM frameworks allow us to easily modify equation (6.0.1). The observation equation can be written in the form:

$$Y_{jit}|\boldsymbol{\theta}_{jit} \;\sim\; \text{UDC}_j(y_{jit}|\boldsymbol{\theta}_{jit}), \tag{6.0.4}$$

where $i = 1, \ldots, n$, $t = 1, \ldots, T$ and $j = 1, \ldots, J$. UDC stands for the "univariate distribution of counts" and includes, but is not limited to Poisson, negative binomial, zero-inflated Poisson, zero-inflated negative binomial, Conway-Maxwell-Poisson distribution, etc. Thus, in (6.0.4) the vector $\boldsymbol{\theta}_{jit}$ denotes a set of parameters associated

with some specific distribution. Notice that in (6.0.2), the link function for the mean can be changed according to the distribution in (6.0.4).

Appendix A

Selected R code

A.1 Simulation of LCMs

```
###### libraries ####

library(MASS)

######  function to simulate LCM (with Poisson marginals)   #####

sim.LCM=function(N=N,log.lambda,level.sigma){

  log.lambda=as.matrix(log.lambda)
  level.sigma=as.matrix(level.sigma)
  J=dim(log.lambda)[1]
  log.mean=matrix(rep(log.lambda,N),ncol=J)

# simulate level alpha

  level.mu=rep(0,J)
  alpha=mvrnorm(n=N, mu=level.mu, Sigma=level.sigma)

# simulate counts
```

```
  new.mean=alpha+log.mean
  mean=exp(new.mean)
  counts=matrix(rep(NA,N*J),ncol=J)

  for (j in 1:J) {
    counts[,j]=rpois(n=N, lambda=mean[,j])
  }

  results=list(counts=counts,alpha=alpha,mean=mean)
  return(results)
}
```

A.2 Estimation of LCMs using R-INLA

```
###### libraries ####

library(INLA)
library(MCMCglmm)

###### This section gives R code used for
###### simulation and estimation of LCMs
###### in Section 5.3
###### for model in (5.3.4)-(5.3.5).
###### Results are given in
###### Tables 5.3.2, 5.3.3, and 5.3.4

######     Input for simulations      ########
######         level.sigma

rho=-0.99
sigma11=0.5
sigma22=0.5
sigma12=rho*sqrt(sigma11*sigma22)
level.sigma.true=matrix(c(sigma11,sigma12,sigma12,sigma22),ncol=2)

###### log.lambda
```

```
mean=c(5,5)
log.mean=log(mean)

####################################
######      Simulation of counts
####################################


### Sample size n
num=c(50,500,1000,5000)
out.corr=c()
out.inla=c()
time.inla=c()
time.mcmc=c()
fit.mcmc=list()
num.sim=length(num)

### Loop for n

for (k in 1:num.sim) {

  N=num[k]
  set.seed(1234567)
  sim.data=sim.LCM(N=N,
                   log.lambda=log.mean,
                   level.sigma=level.sigma.true)
  dim(sim.data$counts)

  # True correlation

  sigma=sigma11
  sig12n=sigma12
  corr.sim=(exp(sig12n)-1)/(exp(sigma)-1+
                              (mean[1]*exp(sigma/2))^(-1))

  # Pearson Correlation

  corr.p=cor.test(x=sim.data$counts[,1],y=sim.data$counts[,2],
    method = "pearson", alternative = "two.sided",conf.level=0.95)
```

```
# Kendall's tau

corr.k=cor.test(x=sim.data$counts[,1],y=sim.data$counts[,2],
  method = "kendall", alternative = "two.sided",conf.level=0.95)

#########     INLA set-up    #############

y=c(sim.data$counts[,1],sim.data$counts[,2])
N=dim(sim.data$counts)[1]
N.all=2*N
index=1:N.all
n=N.all/2

int1=c(rep(1,n),rep(0,n))
int2=c(rep(0,n),rep(1,n))

formula= y ~ int1+int2+f(index, model="iid2d", n=N.all)-1
result= inla(formula,family="poisson",
             data =data.frame(y,index,int1,int2),
             control.compute=list(dic=TRUE))
#summary(result)

# Results from INLA

t1=result$summary.hyperpar[,c(1,3,5)]
t1[1,]=1/t1[1,]
t1[2,]=1/t1[2,]
t2=exp(result$summary.fixed[,c(1,3,5)])
hyper=rbind(t1,t2)

M.inla=exp(result$summary.fixed[1,1])
M.inla.upp=exp(result$summary.fixed[1,5])
M.inla.low=exp(result$summary.fixed[1,3])

rho.inla=result$summary.hyperpar[3,1]
sig11=1/result$summary.hyperpar[1,1]
sig22=1/result$summary.hyperpar[2,1]

sig12=rho.inla*sqrt(sig11*sig22)
```

```
corr.inla=(exp(sig12)-1)/sqrt((exp(sig11)-1+
             (M.inla*exp(sig11/2))^(-1))*(exp(sig22)-1+
             (M.inla*exp(sig22/2))^(-1)))

##### MCMC ##########

y1=sim.data$counts[,1]
y2=sim.data$counts[,2]
dat1=data.frame(y1=y1,y2=y2)

# Start the clock
ptm <- proc.time()

set.seed(123456)
fit1<- MCMCglmm(cbind(y1, y2) ~ trait-1,
                rcov = ~us(trait):units,
                data = dat1, family = c("poisson", "poisson"),
                nitt = 105000,burnin=5000,thin=100,
                verbose = FALSE)

# Stop the clock
runtime=proc.time() - ptm
time.mcmc=c(time.mcmc,runtime[3])

#summary(fit1)
#plot(fit1)

fit.mcmc[[k]]=fit1

##### Set up output ########

out=data.frame(N,corr.p$estimate,
               corr.k$estimate,
               corr.inla,
               corr.sim)

out.corr=rbind(out.corr,out)
out1=data.frame(N,hyper)
out.inla=rbind(out.inla,out1)
```

```
  time.inla=c(time.inla,result$cpu.used[4])

  print(paste("Iteration number:",k))
  print(paste("Time INLA:",result$cpu.used[4]))
  print(paste("Time MCMC:",runtime[3]))

}

t1=c(sigma11,sigma22,rho,mean[1],mean[2])
true=rep(t1,4)
out.inla1=data.frame(out.inla,true)
t2=data.frame(N=w1$N,time.inla,time.mcmc)

w1=round(out.corr,3)
w2=round(out.inla1,3)
w3=round(t2,0)

###### Output Tables

w1
w2
w3
```

Bibliography

A. Agresti. *Analysis of ordinal categorical data*. John Wiley & Sons, 2010.

A. Agresti. *Categorical data analysis*. John Wiley & Sons, Inc., 2013.

J. Aitchison. The statistical analysis of compositional data. *Journal of the Royal Statistical Society*, Series B 44(2):139–177, 1982.

J. Aitchison. *The Statistical Analysis of Compositional Data*. Chapman and Hall, London, UK, 1986.

J. Aitchison and C. H. Ho. The multivariate Poisson lognormal distribution. *Biometrika*, 76(4):643–653, 1989.

M. Al-Osh and A. Alzaid. First-order integer-valued autoregressive (inar (1)) process. *Journal of Time Series Analysis*, 8(3):261–275, 1987.

G. Ball and D. Hall. ISODATA, a novel method of data analysis and pattern classification. Technical report, DTIC Document, 1965.

J. Berger. The case for objective bayesian analysis. *Bayesian analysis*, 1(3):385–402, 2006.

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

J. Besag and C. Kooperberg. On conditional and intrinsic autoregressions. *Biometrika*, 82(4):733–746, 1995.

C. Bloch and M. Willig. Context-dependence of long-term responses of terrestrial gastropod populations to large-scale disturbance. *Journal of Tropical Ecology*, 22 (02):111–122, 2006.

C. Bloch and M. Willig. Density compensation suggests interspecific competition is weak among terrestrial snails in tabonuco forest of puerto rico. *Caribb. J. Sci*, 2010.

G. Box and D. Cox. An analysis of transformations. *Journal of the Royal Statistical Society. Series B*, 26(2):211–252, 1964.

N. Breslow. Extra-poisson variation in log-linear models. *Applied statistics*, pages 38–44, 1984.

S. Brown, A. Lugo, S. Silander, and L. Liegel. Research history and opportunities in the luquillo experimental forest. 1983.

T. Brunsdon. *Time series analysis of compositional data*. PhD thesis, University of Southampton, 1987.

T. Brunsdon and T. Smith. The time series analysis of compositional data. *Journal of Official Statistics*, 14(3):237, 1998.

C. Cameron and P. Trivedi. Regression analysis of count data. 1998.

B. Carlin, N. Polson, and D. Stoffer. A Monte Carlo Approach to Nonnormal and Nonlinear State-Space Modeling. *Journal of the American Statistical Association*, 87:493–500, 1992.

C. Carter and R. Kohn. On gibbs sampling for state space models. *Biometrika*, 81 (3):541–553, 1994.

M.-H. Chen, Q.-M. Shao, and J. Ibrahim. *Monte Carlo Methods in Bayesian Computation*. Springer-Verlag: New York., 2000.

S. Chib and E. Greenberg. Understanding the metropolis-hastings algorithm. *The American statistician*, 49(4):327–335, 1995.

S. Chib and R. Winkelmann. Markov chain monte carlo analysis of correlated count data. *Journal of Business & Economic Statistics*, 2001.

S. Chib, E. Greenberg, and R. Winkelmann. Posterior simulation and bayes factors in panel count data models. *Journal of Econometrics*, 86(1):33–54, 1998.

V. Christou and K. Fokianos. Quasi-likelihood inference for negative binomial time series models. *Journal of Time Series Analysis*, 35(1):55–78, 2014.

R. Conway and W. Maxwell. A queuing model with state dependent service rates. *Journal of Industrial Engineering*, 12(2):132–136, 1962.

A. Cook and G. Barker. Behavioural ecology: on doing the right thing, in the right place at the right time. *The biology of terrestrial molluscs*, pages 447–487, 2001.

C. Cook and J. Stubbendieck. Range research: basic problems and techniques. Technical report, 1986.

R. Davis, W. Dunsmuir, and S. Streett. Observation-driven models for poisson counts. *Biometrika*, 90(4):777–790, 2003.

R. Davis, S. Holan, R. Lund, and N. Ravishanker. *Handbook of Discrete-Valued Time Series*. Chapman and Hall/CRC, 2015.

F. Drost, R. Van Den Akker, and B. Werker. Note on integer-valued bilinear time series models. *Statistics & probability letters*, 78(8):992–996, 2008.

J. Durbin and S. Koopman. Time series analysis of non-gaussian observations based on state space models from both classical and bayesian perspectives. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(1):3–56, 2000.

J. Egozcue, V. Pawlowsky-Glahn, G. Mateu-Figueras, and C. Barcelo-Vidal. Isometric logratio transformations for compositional data analysis. *Mathematical Geology*, 35(3):279–300, 2003.

L. Fahrmeir and S. Lang. Bayesian inference for generalized additive mixed models based on markov random field priors. *Applied statistics*, pages 201–220, 2001.

P. Fearnhead. Mcmc for state-space models. 2011.

K. Fokianos. Truncated poisson regression for time series of counts. *Scandinavian journal of statistics*, 28(4):645–659, 2001.

K. Fokianos. Count time series models. *Time Series–Methods and Applications*, (30): 315–347, 2012.

K. Fokianos and D. Tjøstheim. Nonlinear poisson autoregression. *Annals of the Institute of Statistical Mathematics*, 64(6):1205–1225, 2012.

K. Fokianos, A. Rahbek, and D. Tjøstheim. Poisson autoregression. *Journal of the American Statistical Association*, 104(488):1430–1439, 2009.

E. Forgy. Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *Biometrics*, 21:768–769, 1965.

S. Frühwirth-Schnatter and R. Frühwirth. Auxiliary mixture sampling with applications to logistic models. *Computational Statistics & Data Analysis*, 51(7):3509–3528, 2007.

S. Frühwirth-Schnatter and H. Wagner. Auxiliary mixture sampling for parameter-driven models of time series of counts with applications to state space modelling. *Biometrika*, 93(4):827–841, 2006.

D. Gamerman. Markov chain monte carlo (texts in statistical science), 1997.

D. Gamerman. Markov chain Monte Carlo for dynamic generalized linear models. *Biometrika*, 85(1):215–227, 1998.

D. Gamerman and H. Lopes. *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. CRC Press, 2006.

D. Gamerman and H. Migon. Dynamic hierarchical models. *Journal of the Royal Statistical Society Series (B)*, 55(3):629–642, 1993.

D. Gamerman, T. Santos, and G. Franco. A non-gaussian family of state-space models with exact marginal likelihood. *Journal of Time Series Analysis*, 34(6):625–645, 2013.

N. Garber and T. Lineau. Traffic and Highway Geometric Characteristics Associated with Pedestrian Crashes in Virginia. *Virginia Transportation Research Council. Final Report*, 96-R29, 1996.

A. Gelfand and A. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, 85(410):398–409, 1990.

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

J. Geweke and H. Tanizaki. Bayesian estimation of state-space models using the metropolis–hastings algorithm within gibbs sampling. *Computational Statistics & Data Analysis*, 37(2):151–170, 2001.

N. Gordon, D. Salmond, and A. Smith. Novel approach to nonlinear/non-gaussian bayesian state estimation. In *IEE Proceedings F (Radar and Signal Processing)*, volume 140, pages 107–113. IET, 1993.

S. Guikema and J. Coffelt. Modeling count data in risk analysis and reliability engineering. In *Handbook of Performability Engineering*, pages 579–594. Springer, 2008.

J. Hartigan and M. Wong. Algorithm AS 136: A k-means clustering algorithm. *Applied statistics*, pages 100–108, 1979.

A. Harvey. *Forecasting, structural time series models and the Kalman filter*. Cambridge university press, 1990.

W. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

C. Holmes and L. Held. Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Analysis*, 1(1):145–168, 2006.

J. Hosking. The multivariate portmanteau statistic. *Journal of the American Statistical Association*, 75(371):602–608, 1980.

S. Hu. *Dynamic modeling of discrete-valued time series with applications*. PhD thesis, University of Connecticut, 2012.

S. Hu, J. Ivan, N. Ravishanker, and J. Mooradian. Temporal modeling of highway crash counts for senior and non-senior drivers. *Accident Analysis and Prevention*, 50:1003–1013, 2012.

J. Ibrahim and M.-H. Chen. Power prior distributions for regression models. *Statistical Science*, pages 46–60, 2000.

Insurance Institute for Highway Safety IIHS, 2014. Available at: http://www.iihs.org/iihs/topics/t/roadway-and-environment/fatalityfacts.

A. Jain. Data clustering: 50 years beyond K-means. *Pattern recognition letters*, 31 (8):651–666, 2010.

S. Jensen. Pedestrian Safety in Denmark. *Transport*, 1674:61–69, 1999.

N. Johnson and S. Kotz. *Distributions in statistics*. J. Wiley, 1969.

B. Jørgensen, S. Lundbye-Christensen, PX-K. Song, and L. Sun. A state space model for multivariate longitudinal count data. *Biometrika*, 86(1):169–181, 1999.

R. Jung and A. Tremayne. Binomial thinning models for integer time series. *Statistical Modelling*, 6(2):81–96, 2006.

R. Kalman. A new approach to linearfiltering and prediction theory. *Transactions of the ASME. Series D, Journal of Basic Engineering*, 82:35–45, 1960.

R. Kalman and R. Bucy. New results in filtering and prediction theory. *Transactions of the ASME. Series D, Journal of Basic Engineering*, 83:95–108, 1961.

D. Karlis and L. Meligkotsidou. Multivariate poisson regression with covariance structure. *Statistics and Computing*, 15:255–265, 2005.

D. Karlis and L. Meligkotsidou. Fiinite mixtures of multivariate poisson regression with application. *Journal of Statistical Planning and Inference*, 137:1942–1960, 2007.

J. Klop and A. Khattak. Factors influencing bicycle crash severity on two-lane, undivided roadways in north carolina. *Transportation Research Record: Journal of the Transportation Research Board*, 1674:78–85, 1999.

L. Knorr-Held and H. Rue. On block updating in markov random field models for disease mapping. *Scandinavian Journal of Statistics*, pages 597–614, 2002.

C. Krebs. *Ecology: the experimental analysis of distribution and abundance*. Harper and Row, New York, 1972.

D. Lambert. Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics*, 34(1):1–14, 1992.

F. Landim and D. Gamerman. Dynamic hierarchical models: an extension to matrix-variate observations. *Computational statistics & data analysis*, 35(1):11–42, 2000.

E. LaScala, D. Gerber, and P. Gruenewald. Demographic and environmental correlates of pedestrian injury collisions: A spatial analysis. *Accident Analysis and Prevention*, 32:651–658, 2000.

D. Lindley and A. Smith. Bayes estimates for the linear model. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–41, 1972.

S. Lloyd. Least squares quantization in PCM. *Information Theory, IEEE Transactions on*, 28(2):129–137, 1982.

D. Lord, S. Washington, and J. Ivan. Poisson, poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. *Accident Analysis & Prevention*, 37(1):35–46, 2005.

D. Lord, S. Guikema, and S. Geedipally. Application of the conway–maxwell–poisson generalized linear model for analyzing motor vehicle crashes. *Accident Analysis & Prevention*, 40(3):1123–1134, 2008.

C. Lydeard, R. Cowie, W. Ponder, A. Bogan, P. Bouchet, S. Clark, K. Cummings, T. Frest, O. Gargominy, and D. Herbert. The global decline of nonmarine mollusks. *BioScience*, 54(4):321–330, 2004.

J. Ma, K. Kockelman, and P. Damien. A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods. *Accident Analysis & Prevention*, 40(3):964–975, 2008.

J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA., 1967.

D. Mahamunulu. A note on regression in the multivariate poisson distribution. *Journal of the American Statistical Association*, 62(317):251–258, 1967.

P. McCullagh and J. Nelder. *Generalized Linear Models, 2nd edition*. Chapman and Hall, London, 1989.

W. McDowell, F. Scatena, R. Waide, N. Brokaw, G. Camilo, A. Covich, T. Crowl, G. González, E. Greathouse, and P. Klawinski. Geographic and ecological setting of the luquillo mountains. 2012.

E. McKenzie. Some simple models for discrete variate time series, 1985.

E. McKenzie. Discrete variate time series. *Handbook of statistics*, 21:573–606, 2003.

N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21 (6):1087–1092, 1953.

H. Migon, D. Gamerman, H. Lopes, and M. Ferreira. Dynamic models. *Handbook of Statistics*, 25:553–588, 2005.

N. Mizik and R. Jacobson. Are Physicians Easy Marks? Quantifying the Effects of Detailing and Sampling on New Prescriptions. *Management Science*, 50(12): 1704–1715, 2004.

R. Montoya, O. Netzer, and K. Jedidi. Dynamic Allocation of Pharmaceutical Detailing and Sampling for Long-term Profitability. *Marketing Science*, 29(5):909–924, 2010.

J. Nelder and R. Wedderburn. Generalized linear models. *Journal of the Royal Statistical Society*, 135(3):370–384, 1972.

National Highway Traffic Safety Administration NHTSA, 2014. Retrieved from: http://www-nrd.nhtsa.dot.gov/Pubs/811625.pdf.

A. O'Hagan, C. Buck, A. Daneshkhah, R. Eiser, P. Garthwaite, D. Jenkinson, J. Oakley, and T. Rakow. *Uncertain judgements: eliciting experts' probabilities*. John Wiley & Sons, 2006.

E. Park and D. Lord. Multivariate poisson-lognormal models for jointly modeling crash frequency by severity. *Transportation Research Record: Journal of the Transportation Research Board*, 2009(1):1–6, 2007.

E. Pasanen and H. Salmivaara. Driving speeds and pedestrian safety in the city of helsinki. *Traffic Eng. Control*, 34 (6):1308–310, 1993.

X. Pedeli and D. Karlis. A bivariate inar (1) process with application. *Statistical modelling*, 11(4):325–349, 2011.

N. Ravishanker, R. Venkatesan, and S. Hu. Compositional time series analysis of mortality proportions. *Communications in Statistics-Theory and Methods*, 30(11): 2281–2291, 2001.

W. Rayens and C. Srinivasan. Box-Cox transformations in the analysis of compositional data. *Chemometrics*, 5(3):227–239, 1991.

C. Robert. *The Bayesian Choice: a decision-theoretic motivation*. Springer, New York, 1994.

C. Robert and G. Casella. *Monte Carlo statistical methods*. Springer Science & Business Media, 1999.

H. Rue and L. Held. *Gaussian Markov Random Fields. Theory and Applications*. Chapman and Hall/CRC: New York., 2005.

H. Rue and S. Martino. Approximate Bayesian inference for hierarchical Gaussian Markov random fields models. *Journal of Statistical Planning and Inference*, 137: 3177–3192, 2007.

H. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian models by using Integrated Nested Laplace Approximations. *Journal of the Royal Statistical Society Series B*, 71:319–392, 2009.

R. Ruiz-Cárdenas, E. Krainski, and H. Rue. Direct fitting of dynamic models using integrated nested Laplace approximations – INLA. *Computational Statistics & Data Analysis*, 56(6):1808–1828, 2012.

S. Scheiner and M. Willig. *The theory of ecology*. University of Chicago Press, 2011.

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

Michele F Secrest, Michael R Willig, and Lottie L Peppers. The legacy of disturbance on habitat associations of terrestrial snails in the luquillo experimental forest, puerto rico. *Biotropica*, pages 502–514, 1996.

V. Shankar, J. Milton, and F. Mannering. Modeling accident frequencies as zero-altered probability processes: an empirical inquiry. *Accident Analysis & Prevention*, 29(6):829–837, 1997.

N. Shephard and M. Pitt. Likelihood analysis of non-gaussian measurement time series. *Biometrika*, 84(3):653–667, 1997.

G. Shmueli, T. Minka, J. Kadane, S. Borle, and P. Boatwright. A useful distribution for fitting discrete data: revival of the conway–maxwell–poisson distribution. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(1):127–142, 2005.

T. Smith and T. Brunsdon. The time series analysis of compositional data. *In Proceedings of the Survey Research Methods Section, American Statistical Association*, pages 26–32, 1989.

PX-K. Song. *Correlated data analysis: modeling, analytics, and applications*. Springer Science & Business Media, 2007.

L. Spainhour, I. Wootton, J. Sobanjo, and P. Brady. Causative factors and trends in florida pedestrian crashes. *Transportation Research Record: Journal of the Transportation Research Board*, 1982:90–98, 2006.

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

H. Steinhaus. Sur la division des corp materiels en parties. *Bull. Acad. Polon. Sci*, 1:801–804, 1956.

F. Steutel and K. Van Harn. Discrete analogues of self-decomposability and stability. *The Annals of Probability*, pages 893–899, 1979.

J. Thompson, N. Brokaw, J. Zimmerman, R. Waide, E. Everham III, D. Lodge, C. Taylor, D. García-Montiel, and M. Fluet. Land use history, environment, and tree composition in a tropical forest. *Ecological applications*, 12(5):1344–1363, 2002.

L. Tierney. Markov chains for exploring posterior distributions. *the Annals of Statistics*, pages 1701–1728, 1994.

D. Tjøstheim. Some recent theory for autoregressive count time series. *Test*, 21(3): 413–438, 2012.

R. Venkatesan, W. Reinartz, and N. Ravishanker. The role of attitudinal information in CLV-based customer management. *MSI Working Paper Series*, 12:107, 2012.

C. Weiß. Thinning operations for modeling time series of countsa survey. *AStA Advances in Statistical Analysis*, 92(3):319–341, 2008.

M. West and P. Harrison. *Bayesian Forecasting & Dynamic Models*. Springer Verlag, 1997.

M. West, P. Harrison, and H. Migon. Dynamic generalized linear models and bayesian forecasting. *Journal of the American Statistical Association*, 80(389):73–83, 1985.

M. Willig, E. Sandlin, and M. Gannon. Structural and taxonomic correlates of habitat selection by a puerto rican land snail. *The Southwestern Naturalist*, pages 70–79, 1998.

M. Willig, S. Presley, C. Bloch, I. Castro-Arellano, L. Cisneros, C. Higgins, and B. Klingbeil. Tropical metacommunities along elevational gradients: effects of forest type and other environmental factors. *Oikos*, 120(10):1497–1508, 2011.

M. Willig, S. Presley, C. Bloch, and J. Alvarez. Population, community, and metacommunity dynamics of terrestrial gastropods in the luquillo mountains: a gradient perspective. *Ecological Gradient Analyses in a Tropical Landscape. Ecological Bulletins*, 54:117–140, 2013.

R. Winkelmann. *Econometric analysis of count data*. Springer Science & Business Media, 2008.

S. Zajac and J. Ivan. Factors influencing injury severity of motor vehicle - crossing pedestrian crashes in rural Connecticut. *Accident Analysis and Prevention*, 35(3): 369–379, 2003.

C. Zegeer and M. Bushell. Pedestrian crash trends and potential countermeasures from around the world. *Accident Analysis and Prevention*, 44:3–11, 2012.

S. Zeger. A regression model for time series of counts. *Biometrika*, 75:621–629, 1998.

M. Zhou, L. Li, D. Dunson, and L. Carin. Lognormal and gamma mixed negative binomial regression. In *Machine learning: proceedings of the International Conference. International Conference on Machine Learning*, volume 2012, page 1343. NIH Public Access, 2012.