UNIVERSITY OF CALIFORNIA
SANTA CRUZ

SOME COUNT TIME SERIES RESULTS

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

STATISTICAL SCIENCE

by

Jiajie Kong

December 2023

The Dissertation of Jiajie Kong
is approved:

Professor Robert Lund, Chair

Professor Raquel Prado

Professor Ju Hee Lee

Peter Biehl
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

List of Tables

Abstract

SOME COUNT TIME SERIES RESULTS

by

Jiajie Kong

Count time series are now widely encountered in practice. This dissertation contains three projects on count time series.

Our first project uses a recent advance in stationary count time series to develop a general seasonal count time series modeling paradigm. The model constructed here permits any marginal distribution for the series and the most flexible autocorrelations possible, including those with negative dependence. Likelihood methods of inference are explored. The project first develops the modeling methods, which entail a discrete transformation of a Gaussian process having seasonal dynamics. Properties of this model class are then established and particle filtering likelihood methods of parameter estimation are developed. A simulation study demonstrating the efficacy of the methods is presented and an application to the number of rainy days in successive weeks in Seattle, Washington is given.

Our second project reviews and compares popular methods that produce count time series having Poisson marginal distributions. The project begins by reviewing common ways that count series with Poisson marginal distributions can be produced. Statistical estimation methods are next discussed for some of the more worthy methods. Modeling nonstationary series with covariates motivates consideration of methods where

the Poisson parameter depends on time. The methods are illustrated in the analysis of two series: 1) a count sequence of major hurricanes occurring in the North Atlantic Basin since 1970, and 2) the number of no-hitter games pitched in major league baseball since 1893.

Our third project develops a mathematical model and statistical methods to quantify trends in presence/absence observations of snow cover (not depths) and applies these in an analysis of Northern Hemispheric observations extracted from satellite flyovers during 1967-2021. A two-state Markov chain model with periodic dynamics is introduced to analyze changes in the data in a grid by grid fashion. Trends, converted to the number of weeks of snow cover lost/gained per century, are estimated for each study grid. Uncertainty margins for these trends are developed from the model and used to assess the significance of the trend estimates. Grids with questionable data quality are explicitly identified. Among trustworthy grids, snow presence is seen to be declining in almost twice as many cells as it is advancing. While Arctic and southern latitude snow presence is found to be rapidly receding, other locations, such as Eastern Canada, are experiencing advancing snow cover.

This Dissertation is dedicated to my parents,

Yongqi Kong,

and,

Xiue Kong,

who have provided me with invaluable educational opportunities

and have also demonstrated examples of how to be a good person.

Acknowledgments

During my six years of study in Santa Cruz, I have been profoundly grateful for the unwavering support and guidance given to me by many remarkable individuals on my academic journey.

Foremost, my deepest gratitude goes to my doctoral supervisor, Robert Lund. He has not only been an exceptional professional mentor, but also a true friend. With over 30 years of expertise as a statistician and reviewer, he generously imparted his knowledge, teaching me the art of writing scholarly articles, and cultivating my own academic prowess. Our countless discussions on mathematical concepts and writing dilemmas at his home are among the most cherished memories of my life. Beyond academics, Robert stood by me through many highs and lows, offering unwavering encouragement when I doubted myself. It was he who instilled in me the belief that I am not alone, and that I possess the strength to overcome any challenge.

Next, I extend my heartfelt appreciation to my committee members, Ju Hee and Raquel. Their invaluable time spent reading and providing constructive feedback on my thesis is deeply appreciated. It is an honor to have had you as part of my committee. I am also grateful for the guidance and education you bestowed upon me during my student years. I must also acknowledge my former mentor, Abel Rodriguez, whose fresh perspectives on problem-solving and coding significantly improved my skills. Your research kindled my fascination with stochastic calculus and trajectory prediction.

I am indebted to Herbie Lee and Tony Pourmohamed for leading me into

the captivating realm of Bayesian optimization and collaborating on my paper "An Acquisition Function Family for Bayesian Optimization". Their expertise enriched my understanding and propelled my contributions in the field. Furthermore, I express my gratitude to Professors Kottas, Sanso, Guhaniyogi, Li, and the many staff members whose assistance played an indispensable role in my accomplishments. Your unwavering support has been pivotal to my success.

In addition to my UCSC comrades, I extend my appreciation to my collaborators in various projects: Yisu Jia, Jon Woody, Jamie Dyer, and Steve Marron, who worked with me on the third project in this dissertation. I am immensely grateful for the opportunities that expanded my scientific research horizons and instilled in me a rigorous research attitude. I also extend my thanks to collaborators of "An R package for copula count time series": Stefanos Kechagias, James Livsey, and Vladas Pipiras. Collaborating with you has broadened my scientific horizons and enriched my research experience.

Finally, my heartfelt thanks go to my family. Without your unwavering support, I would not be where I am today. Your love and encouragement have been the foundation of my journey, and for that, I am eternally grateful.

Chapter 1

Introduction

This dissertation seeks to further develop flexible methods to model correlated count time series. Much of the work furthers a recent breakthrough in Jia et al. (2023) in modeling count series having a particular preset marginal distribution. Having a particular marginal distribution for the series allows one to make more accurate inferences. For example, probabilities of the variables at a specific time may be of interest; the series' marginal distribution is needed to compute these. As we will see, hundreds of recent papers have focused on developing strictly stationary time series with a specific marginal distribution.

The work here is partitioned into three projects. In the first project, we develop a model for seasonal (periodic) count data and establish some asymptotic theory for the setup. Our methods are evaluated through simulation studies and application to rainy day counts recorded in Seattle, Washington. This project has been published in *The Journal of Time Series Analysis.*

Our second project considers the classical Poisson count distribution, studying ways that construct count series having a marginal Poisson distribution. Time-varying dynamics, essentially allowing one to conduct a Poisson regression in correlated settings, are developed. This paper will be submitted to *The Journal of Time Series Analysis* shortly.

Our third project is a case study of snow absence/presence trends in the Northern Hemisphere over the last 50 years. This data are recorded weekly, and hence are highly seasonal with snow cover being absent except at the most mountainous or Arctic stations. The data are zero-one valued, with one indicating snow cover. As such, the methods need to account for seasonality and trends in binary data. This study has been published in *The Journal of Hydrometeorology*.

1.1 Background

Correlated random phenomena have been widely studied in recent decades. In time series settings, Gaussian models have dominated the landscape due to their flexibility and ease of interpretation (Brockwell and Davis (2009); Shumway and Stoffer (2017); Cressie (2015); Cressie and Wikle (2015); Banerjee et al. (2014) among many others). Gaussian methods are completely specified by their first and second moments (we call the second moment the covariance or autocovariance function). Unfortunately, count series may not be completely specified by their first two moments. Non-Gaussian issues become particularly important to account for in the case of small counts, the

extreme being a binary series supported on $\{0, 1\}$. As such, non-Gaussian models are needed to adequately describe count series.

As the majority of count data are defined on subsets of $\{0, 1, 2, \cdots\}$, we isolate to this support set here. For example, the number of rainy days in a week, which is examined in our first project, must lies in $\{0, 1, 2, 3, 4, 5, 6, 7\}$, for example.

While the popular generalized linear modeling methods capably describe independent count data, they cannot describe scenarios where the counts are correlated. Researchers began to devise count models having correlation in the 1980s (McKenzie (1985, 1986, 1988)); the subject has become a very active research area today. We now briefly review some count time series approaches of the past.

Inspired by the highly successful autogressive moving-average (ARMA) models, early researchers mimicked the ARMA modeling paradigm to construct count series. For one example, a first order integer autoregression (INAR(1)) series is built with a thinning operator $\circ$ and success parameter $p \in (0, 1)$ via the AR(1)-type equation

$$X_t = p \circ X_{t-1} + \epsilon_t.$$

Here, $\{\epsilon_t\}$ is an independent and identically distributed (IID) count-valued random sequence and $p \circ Y$ is defined as

$$p \circ Y = \sum_{i=1}^{Y} B_i$$

for any count valued random variable Y, where $\{B_i\}_{i=1}^{\infty}$ is a collection of IID Bernoulli trials, independent of Y, with success probability p.

The INAR(1) covariance function is known to have the form $\text{Corr}(X_{t+h}, X_t) = p^h$. Since $p \in [0, 1]$ must be non-negative, negatively correlated series cannot be produced from INAR(1) models. Other authors Franke and Rao Subba (1993); Latour (1997); Pedeli and Karlis (2013a,b); Scotto et al. (2014) have added higher order autoregressive terms and even moving-average components, but negatively correlated series can still not be produced with this model class (INARMA is the general acronym).

Joe (1996, 2016) quantifies the marginal distributions that can be made from INARMA models; this is the so-called convolution closed class. Unfortunately, many common marginal distributions cannot be produced with INARMA models. Perhaps more problematic, INARMA model likelihoods are largely intractable and parameter estimation can be challenging.

Discrete autoregressive moving-average (DARMA) methods are another early count series construction having ARMA features. Unlike INARMA models, the marginal distribution $F_X(\cdot)$ of a strictly stationary DAR(1) series $\{X_t\}$ can be anything desired. After generating X_1 from $F_X(\cdot)$, X_t is obtained via the recursion

$$X_t = B_t X_{t-1} + (1 - B_t)Y_t, \quad t \in \{2, \cdots, N\},$$

where $\{B_t\}$ is a collection of IID Bernoulli trials with success parameter $p \in (0, 1)$, and the Y_t are independently generated from $F_X(\cdot)$. Equivalently, a new observation

has probability p of equalling the previous observation and probability $1-p$ as being a new independent value drawn from $F_X(\cdot)$. The marginal distribution of the DAR(1) series above is easily shown to be $F_X(\cdot)$ via induction. However, DAR(1) series are likely to remain constant whenever the correlation is high (p is close to unity). As such, practitioners abandoned DAR and DARMA methods soon after their discovery.

Blight (1989) and Cui and Lund (2009) took a very different approach to construct count time series. Their methods combine correlated binary (Bernoulli) processes in some manner to achieve the desired marginal distribution.

Consider a Bernoulli series $\{B_t\}$ taking values in $\{0,1\}$ only. One example of this is a renewal point process in which the distance between consecutive renewals are IID and supported in $\{1,2,\ldots\}$. Specifically, a random walk series $\{S_n\}_{n=0}^{\infty}$ is defined from the IID "lifetimes" $\{L_i\}_{i=1}^{\infty}$ via

$$S_n = L_0 + L_1 + L_2 + \cdots + L_n, \quad n \geq 0$$

and the binary sequence is given by $B_t = 1$ if $S_n = t$ for some $n \geq 0$. That is, $B_t = 1$ at times t that are renewal times. In this setup, each $L_1, L_2, \ldots$ have the same distribution as L. The distribution of L_0 may be different from that of the other L_is. The process is called non-delayed if $L_0 = 0$; if L_0 is chosen to have the first derived distribution from the tails of L_1, then $\{B_t\}$ will be a stationary sequence. Otherwise, B_t converges to a stationary setting.

Another method for constructing a correlated Bernoulli series $\{B_t\}$ clips a sta-

tionary Gaussian process $\{Z_t\}$ with zero mean and unit variance into zero-one categories. For example,

$$B_t = 1_A(Z_t)$$

gives a point if Z_t lies in the set A.

Both methods allow $\{B_t\}$ to have negative correlations. In fact, if $\{B_t\}$ is stationary, its covariance function is

$$\gamma_B(h) = \mathbb{E}[L]^{-1}(u_h - \mathbb{E}[L]^{-1}),$$

where u_h denotes the probability of obtaining a renewal (point) at time h in the non-delayed process ($u_0 = 1$).

The count series $\{X_t\}$ is constructed by combining multiple IID copies of $\{B_t\}$ in a way to create the desired marginal distribution. For example, to produce a count series $\{X_t\}$ with a marginal binomial distribution, simply set

$$X_t = \sum_{j=1}^{M_t} B_{t,j}.$$

Here, $M_t = M$ is the number of trials desired at time t and $\{B_{t,1}\}, \{B_{t,2}\}, \cdots, \{B_{t,M}\}$ are independent. Methods to construct $\{X_t\}$ from the $\{B_{t,i}\}$s that produce Poisson, geometric, and other classic count distributions are discussed in Blight (1989), Cui and Lund (2009), Fralix et al. (2012), Lund and Livsey (2016), Lund et al. (2016), and Livsey et al. (2018). This method produces flexible correlation structures — negative

correlations are possible. In fact,

$$\mathrm{Cov}(X_t, X_{t+h}) = \frac{M}{\mathbb{E}(L)} \left(u_h - \mathbb{E}(L)^{-1} \right)$$

in the renewal case, which is negative at lag h whenever $u_h < 1/E[L]$.

This said, some important marginal distributions, such as generalized Poisson, are hard to construct via these methods; also, the most negative correlation possible may not be achieved — see Jia et al. (2023) and our third project for a discussion.

Another mainstream framework for count series modeling involves generalized state space modeling and dynamic linear modeling. A state space model contains two parts: a model for the observation X_t that depends on the state of the system at time t (denoted by α_t), and a model that evolves the state process $\{\alpha_t\}$. A common example is the hierarchical setup

$$X_t | \alpha_t \sim \text{Poisson}(e^{\alpha_t}), \quad \alpha_t = \phi_0 + \phi_1 \alpha_{t-1} + \epsilon_t, \quad \{\epsilon_t\} \sim \text{IID } N(0, \sigma^2), \tag{1.1}$$

where ϕ_0, ϕ_1, and σ^2 are parameters. The model reduces to a classic GLM when $\phi_1 = 0$ and $\sigma^2 = 0$. The observation equation stipulating a Poisson distribution can be changed to any other conditional probability mass function if desired; the log link $\exp \alpha_t$ is used to keep the Poisson parameter non-negative. If α_t is not affected by X_{t-1}, such as in the Gaussian AR(1) setup above, we call this a parameter driven model. The above setup can be viewed as a Bayesian hierarchical model. A good reference for this tactic

is Fokianos et al. (2009a).

In contrast, a count model is called an observation-driven model if the state process is affected by past observations. One example of this mimics the evolution of a GARCH time series:

$$X_t|\mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t), \qquad \lambda_t = d + a_1\lambda_{t-1} + b_1X_{t-1},$$

where d, a_1, and b_1 are nonnegative and $\mathcal{F}_{t-1}$ is the history of observations containing $X_1, \ldots, X_{t-1}$.

Generalized ARMA (GLARMA) methods also use a state space representation. See Dunsmuir et al. (2015) for more detail. GLARMA models are easy to work with, can incorporate covariates, and can have negative correlations. However, the marginal distribution of the model remains unclear. Elaborating, a Poisson distributed $\{X_t\}$ is not constructed via (1.1). Rather, it is $X_t|\alpha_t$ that is Poisson distributed. Once the randomness of α_t is taken into account, the marginal distribution of X_t can be far from Poisson Benjamin et al. (2003).

As such, many previous count modeling classes have some undesirable properties. In general, there are four features that an analyst needs in a count model: 1) general marginal distributions; 2) a general correlation structure; 3) the ability to accommodate covariates; 4) numerically stable and feasible methods for likelihood inference. All of the above mentioned approaches do not permit all of these features; indeed, some permit only one or none.

Recently, a different attack to the problem was taken in Jia et al. (2023). These methods satisfy all four of the above tenets simultaneously. Jia et al. (2023) use a Gaussian copula/transformation to produce a count time series with flexible correlation structures and any desired marginal distribution. Indeed, copula count modeling is not new in the literature. In operations research, Cario and Nelson (1997) and Chen (2001) call this technique the normal to anything (NORTA) approach. Copula techniques to model count data are an active research area in spatial statistics for constructing count fields having an arbitrary marginal distributions Smith and Khaled (2012); Masarotto and Varin (2012); Han and De Oliveira (2016).

The technical construction of Jia et al. (2023) is now presented. A stationary latent Gaussian sequence $\{Z_t\}$ with mean zero and unity variance is used to construct $\{X_t\}$ via the transformation

$$X_t = F_X^{-1}(\mathbf{\Phi}(Z_t)),$$

where $\mathbf{\Phi}$ is the cumulative standard normal distribution and

$$F_X^{-1}(u) = \inf\{t : F_X(t) \geq u\}, \quad u \in (0,1)$$

is the generalized inverse (quantile function) of any discrete distribution F_X. Jia et al. (2023) show that the marginal distribution of X_t is $F_X(\cdot)$, that $\{X_t\}$ is strictly stationary, and derive many other properties of the model.

Let $\gamma_X(h) = \text{Cov}(X_t, X_{t+h})$. Jia et al. (2023) derive $\gamma_Z(h) = \text{Cov}(Z_t, Z_{t+h})$ from $\gamma_X(h)$ via Hermite expansions. The Hermite polynomial basis that we use, $\{H_k(\cdot)\}_{k=0}^{\infty}$,

is defined via

$$H_k(z) = (-1)^k e^{z^2/2} \frac{d^k}{dz^k} \left(e^{-z^2/2} \right).$$

In particular, Jia et al. (2023) show that

$$\gamma_X(h) = \sum_{k=1}^{\infty} k! g_k^2 \gamma_Z^k(h)$$

and derive

$$g_k = \frac{1}{k!\sqrt{2\pi}} \sum_{n=0}^{\infty} e^{-\mathbf{\Phi}^{-1}(Q_n)^2/2} H_{k-1} \left(\mathbf{\Phi}^{-1}(Q_n) \right)$$

where $Q_n = F_X(n)$. The correlation of X_t and Z_t at lag $h \geq 0$ are related by

$$\rho_X(h) = \sum_{k=1}^{\infty} \frac{k! g_k^2}{\gamma_X(0)} \gamma_Z(h)^k =: L(\rho_Z(h)).$$

Here, $L(\cdot)$ is called the link function and serves to map the correlations of $\{Z_t\}$ to those of $\{X_t\}$. $L(\cdot)$ passes through the point $(0,0)$, $(1,1)$ and $(-1, L(-1))$; notice that $L(-1)$ is not necessarily -1. Gaussian copula models produce very flexible autocorrelations; in fact, Jia et al. (2023) prove that the correlation is the most negative possible between X_{t_1} and X_{t_2} when X_{t_1} and X_{t_2} have the same marginal distribution F_X. Additionally, copula models can easily handle covariates and any count marginal distribution can be achieved. Non-Gaussian copulas have also been studied in Biller (2009), Joe (2014), and Escarela et al. (2006).

Chapter 2

Seasonal Count Time Series

2.1 Introduction

Count time series is an active area of current research, with several recent review papers and books appearing on the topic (Fokianos, 2012; Davis et al., 2016; Weiß, 2018; Davis et al., 2021). Gaussian models, which are completely characterized by the series' mean and autocovariance structure, may inadequately describe count series, especially when the counts are small. This chapter uses a recent advance of Jia et al. (2023) to develop a very general count time series model with seasonal characteristics. Specifically, a transformation technique is used to convert a standardized seasonal correlated Gaussian process into a seasonal count time series. The modeling paradigm allows any marginal count distribution to be achieved, has the most flexible correlation structures possible (these can be positive or negative), and can be fitted via likelihood methods. Nonstationary extensions, particularly those involving covariates, are easily

achieved.

With T denoting the known period of the data, our objective is to model a time series $\{X_t\}$ in time t that has a count marginal distribution and periodic properties with period T. A seasonal notation uses $X_{dT+\nu}$ to denote the series during the νth season of cycle d. Here, $\nu \in \{1, 2, \ldots, T\}$ is the seasonal index and $d \in \{0, 1, 2, n/T - 1\}$. We assume that there are n total observations, taken at the times $1, 2, \ldots, n/T$. To avoid trite work with edge effects, we assume that n/T is a whole number.

We seek to construct count series having the cumulative distribution function $F_\nu(x) = P[X_{dT+\nu} \leq x]$ for each cycle d — this stipulation imposes a periodic marginal distribution on the series. In fact, our constructed series will be strictly periodically stationary: for each $k \geq 1$ and all times $t_1 < t_2 < \ldots < t_k$, the joint distribution of $(X_{t_1}, \ldots, X_{t_k})'$ coincides with that of $(X_{t_1+T}, \ldots, X_{t_k+T})'$. We use notations such as $\{X_t\}$ and $\{X_{dT+\nu}\}$ interchangeably, the latter being preferred when seasonality is emphasized.

Some previous seasonal count time series models are now reviewed. The most widely used seasonal count time series models to date develop periodic versions of discrete integer-valued autoregressions (PINAR models) — see Monteiro et al. (2010); Santos et al. (2019), and Bentarzi and Aries (2020). For example, a first order PINAR series $\{X_t\}$ obeys the difference equation

$$X_{dT+\nu} = p(\nu) \circ X_{dT+\nu-1} + \epsilon_{dT+\nu}.$$

Here, $p(\nu) \in [0,1]$ for each season ν and $\circ$ denotes the classical thinning operator: for an independent and identically distributed (IID) sequence of zero-one Bernoulli trials $\{B_i\}_{i=1}^{\infty}$ that have probability of success p and a count-valued random variable C that is independent of $\{B_i\}_{i=1}^{\infty}$, $p \circ C := \sum_{i=1}^{C} B_i$. The noises $\{\epsilon_{dT+\nu}\}$ are periodic independent and identically distributed (IID) count-valued random variables having finite second moments.

The PINAR model class has drawbacks. Even in the stationary case, PINAR models cannot produce some marginal distributions. Joe (2016) quantifies the issue in the stationary case, showing that only marginal distributions in the so-called discrete self-decomposable family can be achieved. Another issue is that PINAR models must have non-negative correlations. Negatively correlated count series do arise (Kachour and Yao, 2009; Livsey et al., 2018; Jia et al., 2023). Likelihood inference for PINAR and INAR models can also be computationally intensive.

A different method for constructing seasonal count series uses a periodic renewal point processes as in Fralix et al. (2012) and Livsey et al. (2018). Here, a zero-one binary sequence $\{B_t\}_{t=1}^{\infty}$ is constructed to be periodically stationary and $\{B_{1,t}\}, \{B_{2,t}\}, \ldots$ denote IID copies of $\{B_t\}_{t=1}^{\infty}$. The periodic count series is constructed via the superposition

$$X_t = \sum_{i=1}^{N_t} B_{i,t}.$$

Here, $\{N_t\}_{t=1}^{\infty}$ is a periodic IID sequence of count valued random variables independent of the $\{B_{i,t}\}$. For example, to obtain a correlated sequence $\{X_t\}$ with Poisson

marginal distributions, $\{N_t\}$ is taken to be independent in t with Poisson marginal distributions, with $N_{dT+\nu}$ having the mean $\lambda_\nu > 0$. Then it is easy to see that $X_{dT+\nu}$ is Poisson distributed with mean $\lambda_\nu P(B_\nu = 1)$. Fralix et al. (2012); Lund and Livsey (2016), and Livsey et al. (2018) show how to produce the classical count marginal distributions (Poisson, binomial, and negative binomial) with this setup and consider $\{B_t\}$ constructed by clipping Gaussian processes.

While binary-based models typically have negative correlations whenever $\{B_t\}$ does, it can be difficult to achieve some marginal distributions. A prominent example of this is the often sought generalized Poisson marginal. Perhaps worse, likelihood methods of parameter inference appear intractable — current parameter inference methods use Gaussian pseudo-likelihoods, which only employ the series' mean and covariance structure. See Davis et al. (2021) for additional detail.

Before proceeding, a clarification needs to be made. The models constructed here posit a particular count marginal distribution for the data *a priori*. This differs from dynamic linear modeling goals, where count models are often built from conditional specifications. For a time-homogeneous first-order autoregressive (AR(1)) example, a dynamic linear model might employ the state space setup $X_t|\alpha_t \sim \text{Poisson}(e^{\alpha_t})$, where $\alpha_t = \beta\alpha_{t-1} + \eta_t$, $|\beta| < 1$, and $\{\eta_t\}$ is zero mean Gaussian noise. Such a setup produces a conditional Poisson distribution, not a series with a Poisson marginal distribution. In fact, the marginal distribution in the above Poisson state space setup is Poisson log-normal (Asmussen and Foss, 2014) and can be far from Poisson.

Additional work on periodic count series is contained in Moriña et al. (2011);

Monteiro et al. (2015); Bentarzi and Bentarzi (2017); Aknouche et al. (2018); Santos et al. (2021); Aknouche et al. (2018), and Ouzzani and Bentarzi (2019). Most of these references take one of the above approaches. Motivated by Jia et al. (2023), this chapter presents a different approach.

The rest of this chapter proceeds as follows. The next section reviews periodic time series methods, focusing on periodic autoregressive moving-average (PARMA) and seasonal autoregressive moving-average (SARMA) difference equation structures. Section 2.3 clarifies our model and its properties. Section 2.4 narrates parameter estimation methods and Section 2.5 studies these techniques via simulation. Section 2.6 analyzes a twenty year segment of weekly rainy day counts in Seattle, Washington. Section 2.7 concludes with comments and remarks.

2.2 Periodic Time Series Background

This section briefly reviews periodic (seasonal) time series. Our future count construction uses a series $\{Z_t\}$, standardized to $E[Z_t] \equiv 0$ and $\text{Var}(Z_t) \equiv 1$, and having Gaussian marginal distributions. While the mean of $\{Z_t\}$ is zero, periodic features in the autocorrelation function (ACF) of $\{Z_t\}$, which we denote by $\rho_Z(t,s) = \text{Cov}(Z_t, Z_s)$, will become paramount.

We call $\{Z_t\}$ a PARMA(p,q) series if it obeys the periodic ARMA(p,q) differ-

ence equation

$$Z_{dT+\nu} = \sum_{k=1}^{p} \phi_k(\nu) Z_{dT+\nu-k} + \eta_{dT+\nu} + \sum_{k=1}^{q} \theta_k(\nu) \eta_{dT+\nu-k}. \tag{2.1}$$

Here, $\{\eta_t\}$ is a zero mean white noise sequence with the periodic variance $\mathrm{Var}(\eta_{dT+\nu}) = \sigma^2(\nu) > 0$. The autoregressive order is p and the moving-average order is q, which are taken constant in the season ν for simplicity. The autoregressive and moving-average coefficients are $\phi_1(\nu), \ldots, \phi_p(\nu)$ and $\theta_1(\nu), \ldots, \theta_q(\nu)$, respectively, during season ν. We tacitly assume that model parameters are identifiable from the autocovariance function of the series. This may require more than the classical causality and invertibility conditions (Reinsel, 2003). Gaussian PARMA solutions are strictly stationary with period T as long as the autoregressive polynomial does not have a root on the complex unit circle — see Lund and Basawa (1999) for quantification. Not all PARMA parameters are free due to the restriction $\mathrm{Var}(X_t) \equiv 1$; the following example delves further into the matter.

Example 3.1 A PAR(1) series with period T obeys the recursion

$$Z_t = \phi(t) Z_{t-1} + \eta_t, \tag{2.2}$$

where $\{\eta_t\}$ is zero mean white noise with $\mathrm{Var}(\eta_t) = \sigma^2(t)$. The quantities $\phi(t)$ and $\sigma^2(t)$ are assumed periodic in t with period T. This difference equation is known to have a unique (in mean squared) and causal solution whenever there is a stochastic contraction

over an entire cycle, namely $|\prod_{\nu=1}^{T} \phi(\nu)| < 1$ (Lund and Basawa, 1999).

To impose $\mathrm{Var}(Z_t) \equiv 1$, take a variance on both sides of (2.2) and set $\mathrm{Var}(Z_t) = \mathrm{Var}(Z_{t-1}) = 1$ to infer that $\sigma^2(t) = 1 - \phi^2(t)$, which we tacitly assume is positive for all t. This uses $\mathrm{Cov}(Z_{t-1}, \eta_t) = 0$, which follows from causality. The covariance structure of $\{Z_t\}$ can now be extracted as

$$\rho_Z(t, s) = \prod_{i=0}^{t-s-1} \phi(t - i).$$

for $s < t$. ♣.

Another class of periodic models in use today are the SARMA series. SARMA series are actually time-stationary models, but have comparatively large autocorrelations at lags that are multiples of the period T. The most basic SARMA(p, q) series $\{Z_t\}$ obeys a difference equation driven at lags that are multiples of the period T:

$$Z_t = \sum_{k=1}^{p} \phi_k Z_{t-kT} + \eta_t + \sum_{k=1}^{q} \theta_k \eta_{t-kT}, \tag{2.3}$$

where $\{\eta_t\}$ is zero mean independent noise with a constant variance. In this setup, $\rho_Z(t, s) = 0$ unless $t - s$ is a whole multiple of the period T. As such, many authors allow $\{\eta_t\}$ to have additional correlation, specifically a zero mean ARMA(p^*, q^*) series. This results in a model that can have non-zero correlations at any lag; however, the model is still stationary and does not have any true periodic features. Since the model is stationary, we write $\rho_Z(t, s) = \rho_Z(t - s)$.

Example 3.2 A SAR(1) series with period T and AR(1) $\{\eta_t\}$ obeys the difference equation pair

$$Z_t = \phi Z_{t-T} + \eta_t; \quad \eta_t = \alpha\eta_{t-1} + \epsilon_t, \tag{2.4}$$

where $\{\epsilon_t\}$ is zero mean white noise with variance σ_ϵ^2, $|\phi| < 1$, and $|\alpha| < 1$. Combining these two difference equations results in a stationary and causal AR($T+1$) model for $\{Z_t\}$.

Imposing $\mathrm{Var}(Z_t) \equiv 1$ and taking a variance in the first equation in (2.4) gives

$$1 = \phi^2 + \mathrm{Var}(\eta_t) + 2\phi\mathrm{Cov}(Z_{t-T}, \eta_t).$$

To proceed, use equation (2.4)'s causal solutions $\eta_t = \sum_{k=0}^{\infty} \alpha^k \epsilon_{t-k}$ and $Z_{t-\ell} = \sum_{m=0}^{\infty} \phi^m \eta_{t-mT-\ell}$ to get

$$\mathrm{Cov}(\eta_t, Z_{t-\ell}) = \sigma_\epsilon^2 \frac{\alpha^\ell}{(1-\alpha^2)(1-\phi\alpha^T)} \tag{2.5}$$

for any $\ell > 0$. Combining the last two equations, we see that taking

$$\sigma_\epsilon^2 = \frac{(1-\phi^2)(1-\alpha^2)(1-\phi\alpha^T)}{1+\phi\alpha^T} \tag{2.6}$$

indices $\mathrm{Var}(Z_t) \equiv 1$.

To extract the covariance structure of $\{Z_t\}$, multiply both sides of (2.4) by

Z_{t-h} and take expectations to get the Yule-Walker type equations

$$\begin{aligned} \rho_Z(0) &= \phi\rho_Z(T) + E(Z_t\eta_t) \\ &\vdots \\ \rho_Z(T) &= \phi\rho_Z(0) + E(Z_{t-T}\eta_t). \end{aligned}$$

This system can be rewritten in vector form as

$$\begin{bmatrix} 1 & 0 & \cdots & 0 & -\phi \\ 0 & 1 & \cdots & -\phi & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & -\phi & \cdots & 1 & 0 \\ -\phi & 0 & \cdots & 0 & 1 \end{bmatrix} \begin{bmatrix} \rho_Z(0) \\ \rho_Z(1) \\ \vdots \\ \rho_Z(T-1) \\ \rho_Z(T) \end{bmatrix} = \begin{bmatrix} E(\eta_t Z_t) \\ E(\eta_t Z_{t-1}) \\ \vdots \\ E(\eta_t Z_{t-T+1}) \\ E(\eta_t Z_{t-T}) \end{bmatrix}. \tag{2.7}$$

The Appendix shows that the inverse of the matrix in the above linear system exists. From this, (2.5), (2.6), and some algebraic manipulations detailed in the Appendix, one extracts

$$\rho_Z(h) = \frac{\alpha^h + \phi\alpha^{T-h}}{1 + \phi\alpha^T}, \quad 0 \le h \le T.$$

To identify the model correlations at lags $h > T$, multiply the first equation in (2.4) by Z_{t-h} for $h > T$ and take expectations to get the recursion $\rho_Z(h) = \phi\rho_Z(h-T) +$

$E(\eta_t Z_{t-h})$. This can be solved with (2.5) to get (see the Appendix)

$$\rho_Z(h) = \phi^a \frac{\alpha^b + \phi\alpha^{T-b}}{1+\phi\alpha^T} + \sum_{k=0}^{a-1} \phi^k \alpha^{h-Tk} \frac{1-\alpha^2}{1+\phi\alpha}, \quad h > T,$$

where $a = \lfloor h/T \rfloor$ and $b = h - aT$. ♣.

PARMA and SARMA methods are compared in detail in Lund (2011). PARMA models are usually more applicable since the immediate past of the process is typically more influential than past process lags at multiples of the period T. Applications in the environment (Vecchia, 1985; Bloomfield et al., 1994; Lund et al., 1995a; Tesfaye et al., 2004) tend to be PARMA; SARMA structures are useful in economics (Franses, 1994; Franses and Paap, 2004; Hurd and Miamee, 2007). PARMA reviews are Gardner and Franks (1975); Lund and Basawa (1999), and Gardner et al. (2006); statistical inference for PARMA models is studied in Lund and Basawa (2000); Basawa and Lund (2001); Basawa et al. (2004); Lund et al. (2006); Shao and Ni (2004), and Shao (2006). SARMA inference is addressed in Chatfield and Prothero (1973).

2.3 Methodology

Our methods extend the work in Jia et al. (2023) with Gaussian transformations (copulas) to the periodic setting. Let $\{X_t\}$ denote the time series to be constructed, which takes values in the count support set $\{0, 1, 2, \ldots\}$. Our construction works with a latent Gaussian series $\{Z_t\}$ with zero mean and a unit variance at all times. Then X_t

is obtained from Z_t via

$$X_{dT+\nu} = F_\nu^{-1}\left(\Phi(Z_{dT+\nu})\right), \tag{2.8}$$

where $\Phi(\cdot)$ is the cumulative distribution function (CDF) of the standard normal distribution and $F_\nu(\cdot)$ is the desired marginal distribution for X_t during season ν. Here, F_ν^{-1} is the quantile function

$$F_\nu^{-1}(u) = \inf\left\{x : F_\nu(x) \geq u\right\}. \tag{2.9}$$

As Jia et al. (2023) shows, this construction leaves $X_{dT+\nu}$ with the marginal distribution F_ν for every d and ν. This model is very flexible: any marginal distribution can be achieved for any desired season ν, even continuous ones. The marginal distribution F_ν can have the same form or be different over distinct seasons ν. Any marginal distribution whatsoever can be achieved; when count distributions are desired, the quantile definition in (2.9) is the version of the inverse CDF that produces the desired marginals.

2.3.1 Properties of the Model

Toward ARMA and PARMA model order identification, if $\{Z_t\}$ is an m-dependent series, then Z_{t_1} and Z_{t_2} are independent when $|t_1 - t_2| > m$ since $\{Z_t\}$ is Gaussian. By (2.8), X_{t_1} and X_{t_2} are also independent and $\{X_t\}$ is also m-dependent. From the characterization of stationary moving averages (Proposition 3.2.1 in Brockwell and Davis (1991a)) and periodic moving-averages in Shao and Lund (2004), we see that if $\{Z_t\}$ is a periodic moving average of order q, then $\{X_t\}$ is also a periodic moving

average of order q. Unfortunately, analogous results for autoregressions do not hold. For example, if $\{Z_t\}$ is a periodic first order autoregression, $\{X_t\}$ may not be a periodic autoregression of any order.

We now derive the covariance structure of $\{X_t\}$ via Hermite expansions. Let $\gamma_X(t,r) = \mathrm{Cov}(X_t, X_r)$ be the covariance of $\{X_t\}$ at the times t and r, where $r \leq t$. Then $\gamma_X(t,r)$ can be related to $\rho_Z(t,r)$ via Hermite expansions. To do this, let $G_\nu(x) = F_\nu^{-1}(\Phi(x))$ and write the Hermite expansion of $G_\nu(\cdot)$ as

$$G_\nu(z) = g_0(\nu) + \sum_{k=1}^{\infty} g_k(\nu) H_k(z). \tag{2.10}$$

Here, $g_k(\nu)$ is the kth Hermite coefficient for season ν, whose calculation is described below, and $H_k(z)$ is the kth Hermite polynomial defined by

$$H_k(z) = (-1)^k e^{z^2/2} \frac{d^k}{dz^k} \left(e^{-z^2/2} \right). \tag{2.11}$$

The first three Hermite polynomials are $H_0(x) \equiv 1$, $H_1(x) = x$, and $H_2(x) = x^2 - 1$. Higher order polynomials can be found via the recursion $H_k(x) = xH_{k-1}(x) - H'_{k-1}(x)$, which follows from (2.11). When $G_\nu(\cdot)$ is discontinuous, the partial sums in (2.10) are continuous, but converge to a discrete limit.

The polynomials H_k and H_j are orthogonal with respective to the standard Gaussian measure if $k \neq j$: $E[H_k(Z)H_j(Z)] = 0$ for a standard normal Z unless $j = k$

(in which case $E[H_k(Z)^2] = k!$). The Hermite coefficients are computed from

$$g_k(\nu) = \frac{1}{k!}\int_{-\infty}^{\infty} G_\nu(t)H_k(t)\phi(t)dt, \quad k = 1, 2, \ldots, \tag{2.12}$$

where $\phi(t) = \Phi'(t) = e^{-t^2/2}/\sqrt{2\pi}$ is the standard normal density.

Lemma 2.1 in Han and De Oliveira (2016) shows that

$$\gamma_X(t, r) = \sum_{k=1}^{\infty} k! g_k(s(t))g_k(s(r))\rho_Z(t, r)^k, \tag{2.13}$$

where $s(t) = t - T\lfloor (t+1)/T \rfloor$ denotes the season corresponding to time t. Let $\sigma_X^2(t) = \gamma_X(t, t) = \sum_{k=1}^{\infty} k! g_k^2(s(t))$ denote the variance of X_t. Then the ACF of $\{X_t\}$ is

$$\rho_X(t, r) = \frac{\gamma_X(t, r)}{\sigma_X(t)\sigma_X(r)} = \sum_{k=1}^{\infty} \frac{k! g_k(s(t))g_k(s(r))}{\sigma_X(t)\sigma_X(r)}\rho_Z(t, r)^k = \sum_{k=1}^{\infty} \ell_k \rho_Z(t, r)^k := L(\rho_Z(t, r)), \tag{2.14}$$

which is a power series in $\rho_Z(t, r)$ with kth coefficient

$$\ell_k := \frac{k! g_k(s(t))g_k(s(r))}{\sigma_X(t)\sigma_X(r)}. \tag{2.15}$$

The quantity $L(\cdot)$ is called a link function and ℓ_k a link coefficient. When $\{Z_t\}$ is stationary and F_ν does not depend on ν, Jia et al. (2023) show that $L(0) = 0$, $L(1) = 1$, and $L(-1) = \text{Corr}(G(Z_0), G(-Z_0))$. It is not true that $L(-1) = -1$ in any case nor is $L(1) = 1$ in the periodic case; indeed, stationary or periodically stationary count processes with arbitrarily positive or negative correlations may not exist. For

example, the pair $(Z, -Z)$, where Z is standard normal has correlation -1, but two Poisson random variables, both having mean λ, whose correlation is -1, do not exist.

The model produces the most flexible correlation structures possible in a pairwise sense. Specifically, consider two distinct seasons ν_1 and ν_2 and suppose that F_{ν_1} and F_{ν_2} are the corresponding marginal distributions for these seasons. Then Theorems 2.1 and 2.5 in Whitt (1976) show that the bivariate random pair (X_{ν_1}, X_{ν_2}) having the marginal distributions F_{ν_1} and F_{ν_2}, respectively, with the largest correlation possible has form $X_{\nu_1} = F_{\nu_1}^{-1}(U)$ and $X_{\nu_2} = F_{\nu_2}^{-1}(U)$, where U is a uniform[0,1] random variable. To achieve this largest correlation, one simply takes $\{Z_t\}$ to have unit correlation at these times; that is, take $Z_{\nu_1} = Z_{\nu_2}$. Since $\Phi(Z_{\nu_1})$ is distributed as uniform[0,1], the claim follows. For negative correlations, the same results in Whitt (1976) also show that the most negatively correlated pair that can be produced has the form $X_{\nu_1} = F_{\nu_1}^{-1}(U)$ and $X_{\nu_2} = F_{\nu_2}^{-1}(1-U)$. This is produced with a Gaussian series having $\mathrm{Corr}(Z_{\nu_1}, Z_{\nu_2}) = -1$, which is obtained by selecting $Z_{\nu_2} = -Z_{\nu_1}$. Then $\Phi(Z_{\nu_1})$ is again uniform[0,1] and $\Phi(Z_{\nu_2}) = \Phi(-Z_{\nu_1}) = 1 - \Phi(Z_{\nu_1})$, since $\Phi(-x) = 1 - \Phi(x)$ for all real x.

The previous paragraph implies that one cannot obtain more general autocorrelation functions for count series than what has been constructed above — they do not exist. Negatively correlated count series do arise (Kachour and Yao, 2009; Livsey et al., 2018; Jia et al., 2023) and can be described with this model class. In the stationary case where the marginal distribution F_ν is constant over all seasons ν, a series $\{X_t\}$ with $\mathrm{Cov}(X_t, X_{t+h}) = 1$ for all h is achieved by taking $Z_t \equiv Z$, where Z is standard normal. This unit correlation property will not carry over to our periodic setting. For example, a

random pair (X_{ν_1}, X_{ν_2}) having a Poisson marginal with mean λ_{ν_1} during season ν_1 and a Poisson marginal with mean λ_{ν_2} during season ν_2 with a unit correlation do not exist when $\lambda_{\nu_1} \neq \lambda_{\nu_2}$. This said, the model can produce any correlation structures within "the range of achievable correlations". As such, the model class here is quite flexible.

The amount of autocorrelation that $\{X_t\}$ inherits from $\{Z_t\}$ is now discussed. An implication of the result below, which establishes monotonicity of the link function by showing that its derivative is positive, is that the larger the autocorrelations are in $\{Z_t\}$, the larger the autocorrelations are in $\{X_t\}$. We state the result below and prove it in the Appendix.

Proposition 3.1: *For a fixed t and r, let $L(\cdot)$ denote the link function in (2.14). Then for $u \in (-1, 1)$, the derivative of the link is positive and has form*

$$L'(u) = \frac{\sum_{j_1=0}^{\infty}\sum_{j_2=0}^{\infty} e^{-\frac{1}{2(1-u^2)}\left[\Phi^{-1}(C_{j_1}(s(t))^2+\Phi^{-1}(C_{j_2}(s(r))^2-2u\Phi^{-1}(C_{j_1}(s(t))\Phi^{-1}(C_{j_2}(s(r))\right]}}{\sqrt{\sigma_X(t)\sigma_X(r)}2\pi\sqrt{1-u^2}}. \tag{2.16}$$

Here,

$$C_j(\nu) = \mathbb{P}[X_\nu \leq j] = F_\nu(j) \tag{2.17}$$

denotes the cumulative probabilities of X_ν at season ν. Equation (2.13) shows that

$$|\rho_X(t,r)| \leq |\rho_Z(t,r)|,$$

which can also be deduced by the Cauchy-Schwarz inequality. As such, the amount of autocorrelation in the count series $\{X_t\}$ is "less than" that for the latent Gaussian $\{Z_t\}$

series. For feel, Figure 2.1 plots $L(u)$ for several parameter combinations with Poisson and binomial distributions. The plots show that $L(u)$ is non-decreasing in u and that the autocorrelation of the count model is always less than the underlying Gaussian process. The "correlation loss" from $\{Z_t\}$ to $\{X_t\}$ is minor once both λ are larger than unity for the Poisson case and both p are larger than 0.1 for the binomial case. In contrast to the stationary analysis in Jia et al. (2023), the link function may not have $L(1) = 1$ when the two marginal distributions are different.

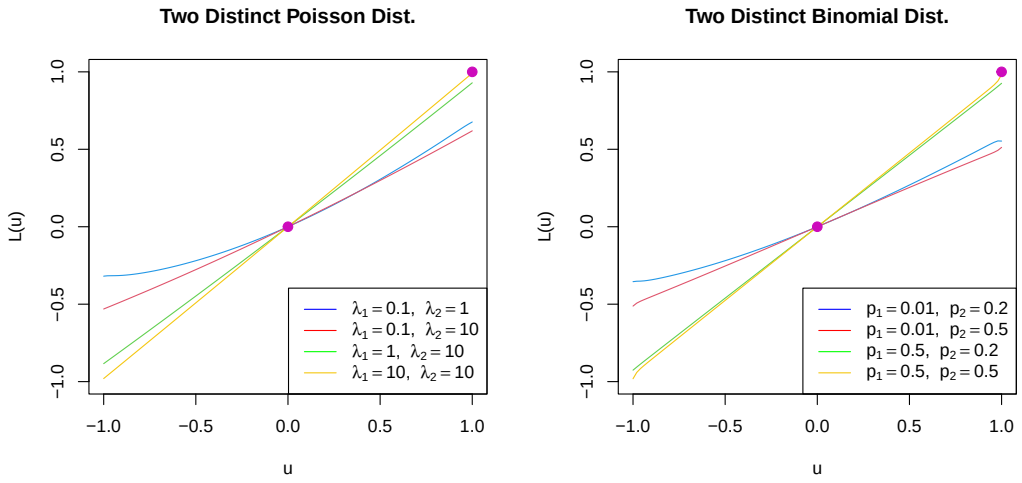

Figure 2.1: Left: the link function constructed by two distinct Poisson distributions; right: the link function constructed by two binomial Poisson distributions with number of trials equal to 7. Pink dots are points $(0, 0)$ and $(1, 1)$.

2.3.2 Calculation and Properties of the Hermite Coefficients

An important numerical task entails calculating $g_k(\nu)$, which only depends on $F_\nu(\cdot)$ by (2.12). To do this, rewrite $G_\nu(z)$ in the form

$$G_\nu(z) = \sum_{j=0}^{\infty} j\mathbf{1}_{[C_{j-1}(\nu)\le\Phi^{-1}(z)<C_j(\nu)]} = \sum_{j=1}^{\infty} j\mathbf{1}_{[\Phi^{-1}(C_{j-1}(\nu)),\Phi^{-1}(C_j(\nu)))}(z), \tag{2.18}$$

where the convention $C_{-1} = 0$ is made. We also take $\Phi^{-1}(0) = -\infty$ and $\Phi^{-1}(1) = \infty$. Then for $k \ge 1$, integration by parts yields

$$\begin{aligned}
g_k(\nu) &= \frac{1}{k!}\sum_{j=0}^{\infty} n\mathbb{E}\left[\mathbf{1}_{[\Phi^{-1}(C_{j-1}(\nu)),\Phi^{-1}(C_j(\nu)))}(Z_0)H_k(Z_0)\right] \\
&= \frac{1}{k!}\sum_{j=0}^{\infty}\frac{j}{\sqrt{2\pi}}\int_{\Phi^{-1}(C_{j-1}(\nu))}^{\Phi^{-1}(C_j(\nu))} H_k(z)e^{-z^2/2}dz \\
&= \frac{1}{k!}\sum_{j=0}^{\infty}\frac{j}{\sqrt{2\pi}}\int_{\Phi^{-1}(C_{j-1}(\nu))}^{\Phi^{-1}(C_j(\nu))} (-1)^k\left(\frac{d^k}{dz^k}e^{-z^2/2}\right)dz \\
&= \frac{1}{k!}\sum_{j=0}^{\infty}\frac{j}{\sqrt{2\pi}}(-1)^k\left(\frac{d^{k-1}}{dz^{k-1}}e^{-z^2/2}\right)\Bigg|_{z=\Phi^{-1}(C_{j-1}(\nu))}^{z=\Phi^{-1}(C_j(\nu))} \\
&= \frac{1}{k!}\sum_{j=0}^{\infty}\frac{j}{\sqrt{2\pi}}(-1)e^{-z^2/2}H_{k-1}(z)\Bigg|_{z=\Phi^{-1}(C_{j-1}(\nu))}^{z=\Phi^{-1}(C_j(\nu))}.
\end{aligned}$$

Simplifying this telescoping sum gives

$$g_k(\nu) = \frac{1}{k!\sqrt{2\pi}}\sum_{j=0}^{\infty} e^{-[\Phi^{-1}(C_j(\nu))]^2/2}H_{k-1}(\Phi^{-1}(C_j(\nu))). \tag{2.19}$$

Notice that the summands in (2.19) are zero whenever $\Phi^{-1}(C_j(\nu)) = \pm\infty$. Lemma 2.1 in Jia et al. (2023) shows that the expansion converges whenever $\mathbb{E}[X_t^p] < \infty$ for some $p > 1$. This condition automatically holds for time series, which implicitly require a finite second moment. For count distributions with a finite support, $C_j(\nu)$ becomes unity for large j. For example, a Binomial marginal distribution with 7 trials is considered in our later application. Here, the summation can be reduced to $j \in \{0, 1, \ldots, 7\}$. For count distributions on a countably infinite support, approximating (2.19) requires truncation of an infinite series. This is usually not an issue: numerically, $C_j(\nu)$ quickly converges to unity as $j \to \infty$ for light tailed distributions — or equivalently, $e^{-\Phi^{-1}(C_j(\nu))^2/2} H_{k-1}(\Phi^{-1}(C_j(\nu))) \to 0$. In addition to (2.19), $g_k(\nu)$ can also be approximated by Gaussian quadrature; see *gauss.quad.prob* in the package *statmod* in *R*. However, the approximation in (2.19) is more appealing in terms of simplicity and stability (Jia et al., 2023).

2.4 Parameter Inference and Forecasting

2.4.1 Parameter Inference

This subsection develops likelihood methods of inference for the model parameters via particle filtering and sequential Monte Carlo techniques. With many count time series model classes, likelihoods are intractable (Davis et al., 2021). Accordingly, researchers have defaulted to moment and composite likelihood techniques. However, if the count distributional structure truly matters, likelihood methods should "feel"

this structure and return superior parameter estimates. Gaussian pseudo-likelihood estimates, which are based only on the mean and autocovariance of the series, are developed in Jia et al. (2023) for stationary series; they present an example where Gaussian pseudo-likelihood estimates perform well and an example where they perform poorly.

For notation, let $\boldsymbol{\theta}$ contain all parameters in the T marginal distributions $F_1, \ldots, F_T$ and $\boldsymbol{\eta}$ contain all parameters governing the evolution of $\{Z_t\}$. The data $\{x_1, x_2, \ldots, x_n\}$ denote our realization of the series $\{X_t\}_{t=1}^n$.

The likelihood function $\boldsymbol{\mathcal{L}}(\boldsymbol{\theta}, \boldsymbol{\eta})$ is simply a high dimensional multivariate normal probability. To see this, use (2.8) to get

$$\boldsymbol{\mathcal{L}}(\boldsymbol{\theta}, \boldsymbol{\eta}) = \mathbb{P}(X_1 = x_1, \cdots, X_n = x_n) = \mathbb{P}\left(Z_1 \in (a_1, b_1], \cdots, Z_n \in (a_n, b_n]\right) \tag{2.20}$$

for some numbers $\{a_i\}_{i=1}^n$ and $\{b_i\}_{i=1}^n$ (these are clarified below but are not important for the discussion here). The covariance matrix of $(Z_1, \ldots, Z_n)$ depends only on $\boldsymbol{\eta}$, not on $\boldsymbol{\theta}$. Unfortunately, evaluating a high dimensional multivariate normal probability is numerically challenging for large n. While methods to handle this problem exist (Kazianka and Pilz, 2010; Kazianka, 2013; Bai et al., 2014), they often contain substantial estimation bias.

An alternative approach, which is the one taken here, uses simulation methods to approximate the likelihood. General methods in this category include the quasi-Monte Carlo methods of Genz and Bretz (2002) and the prominent Geweke–Hajivassiliou–Keane (GHK) simulator of Geweke (1991) and Hajivassiliou et al. (1996). The performance of

these methods, along with an additional "data cloning" approach, are compared in Han and De Oliveira (2020), where the author shows that estimators from these methods are similar, but that the GHK methods are much faster, having a numerical complexity as low as order mn. Here, m is the pre-selected number of sample paths to be simulated (the number of particles). As we will subsequently see, the GHK simulator works quite well for large m.

Jia et al. (2023) develop a sequential importance sampling method that uses a modified GHK simulator. In essence, importance sampling is used to evaluate integrals by drawing samples from an alternative distribution and averaging their corresponding weights. Suppose that we seek to estimate $\int_{\mathcal{D}} f(\boldsymbol{x})d\boldsymbol{x}$. Then

$$\int_{\mathcal{D}} f(\boldsymbol{x})d\boldsymbol{x} = \int_{\mathcal{D}} \frac{f(\boldsymbol{x})}{q(\boldsymbol{x})} q(\boldsymbol{x})d\boldsymbol{x},$$

where $f(\boldsymbol{x})/q(\boldsymbol{x})$ is called the weight and the proposed distribution q is called the importance distribution. Without loss of generality, we assume that $q(\boldsymbol{x}) > 0$ whenever $\boldsymbol{x} \in \mathcal{D}$ and that $q(\boldsymbol{x}) = 0$ for $\boldsymbol{x} \in \mathcal{D}^c$. Then the importance sampling estimate of the integral is the law of large numbers justified average

$$\int_{\mathcal{D}} \frac{f(\boldsymbol{x})}{q(\boldsymbol{x})} q(\boldsymbol{x})d\boldsymbol{x} \approx \frac{1}{m} \sum_{k=1}^{m} \frac{f(\boldsymbol{x}^{(k)})}{q(\boldsymbol{x}^{(k)})},$$

where $\{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(m)}\}$ are m IID samples drawn from the proposed distribution q. With

the notation $z_{1:n} = (z_1, \ldots, z_n)$, notice that the likelihood in (2.20) has form

$$\int_{\{z_t \in (a_t, b_t], t=1,\ldots,n\}} \boldsymbol{\phi}_{\boldsymbol{\eta}}(z_{1:n})\, dz_1 \ldots dz_n = \int_{\{z_t \in (a_t, b_t], t=1,\ldots,n\}} \frac{\boldsymbol{\phi}_{\boldsymbol{\eta}}(z_{1:n})}{q(z_{1:n})} q(z_{1:n}) dz_1 \ldots dz_n. \tag{2.21}$$

Observe that $\{a_i\}_{i=1}^n$ and $\{b_i\}_{i=1}^n$ only depend on $\boldsymbol{\theta}$ and the data $\{x_1, x_2, \ldots, x_n\}$; specifically,

$$a_t = \Phi^{-1}(C_{x_t - 1}(s(t))) \quad \text{and} \quad b_t = \Phi^{-1}(C_{x_t}(s(t))),$$

where $C_n(\nu)$ is defined in (2.17) and $s(t)$ is the season at time t. Here, it is best to choose a proposed distribution q such that 1) $q(z_{1:n}) > 0$ for $z_t \in (a_t, b_t]$ and $q(z_{1:n}) = 0$ otherwise; 2) the weight $\boldsymbol{\phi}_{\boldsymbol{\eta}}(z_{1:n})/q(z_{1:n})$ is easy to compute; and 3) $\{Z_t\}$ can be efficiently drawn from q. Our GHK simulator satisfies all three conditions.

To develop our GHK sampler further, we take advantage of the latent Gaussian structure in the PARMA or SARMA series $\{Z_t\}$. In simple cases, $\{Z_t\}$ may even be a Markov chain. The GHK algorithm samples Z_t, depending on the its previous history $Z_{t-1}, \ldots, Z_1$ and X_t, from a truncated normal density. Specifically, let $p_{\boldsymbol{\eta}(t)}(z_t | z_{t-1}, \cdots, z_1, x_t)$ denote the truncated normal density of Z_t given the history $Z_{t-1}, \ldots, Z_1$ and $X_t = x_t$. Then

$$p_{\boldsymbol{\eta}(t)}(z_t | z_{t-1}, \ldots, z_1, x_t) = \frac{1}{r_t} \left[\frac{\phi(\frac{z_t - \hat{z}_t}{r_t})}{\Phi(\frac{b_t - \hat{z}_t}{r_t}) - \Phi(\frac{a_t - \hat{z}_t}{r_t})} \right], \quad a_t < z_t < b_t, \tag{2.22}$$

where $\hat{z}_t$ and r_t are the one-step-ahead mean and standard deviation of Z_t conditioned on $z_1, z_2, \ldots, z_n$. Again, a_t and b_t only depend on x_t. Here, we choose the importance

sampling distribution

$$q_{\boldsymbol{\eta}}(z_{1:n}|x_{1:n}) = p_{\boldsymbol{\eta}(1)}(z_1|x_1)\prod_{t=2}^{n} p_{\boldsymbol{\eta}(t)}\left(z_t|z_{t-1},\ldots,z_1,x_t\right). \tag{2.23}$$

Elaborating further, let $\mathcal{N}(\mu,\sigma^2;a,b)$ denote a normal random variable with mean μ and variance σ^2 restricted to lie in $(a,b]$, where $a < b$. Then X_1 is first drawn from $\mathcal{N}(0,1;a_1,b_1)$. Thereafter, $Z_2, Z_3, \ldots, Z_n$ are sequentially sampled from the distribution in (2.22). The proposed importance sampling distribution is efficient to sample, has the desired distributional support, and induces an explicit expression for the weights:

$$\frac{\boldsymbol{\phi}_{\boldsymbol{\eta}(t)}\left(z_{1:n}\right)}{q_{\boldsymbol{\eta}(t)}(z_{1:n}|x_{1:n})} = \frac{p_{\boldsymbol{\eta}(1)}(z_1)}{p_{\boldsymbol{\eta}(1)}(z_1|x_1)}\prod_{t=2}^{n}\frac{p_{\boldsymbol{\eta}(t)}\left(z_t\middle|z_{t-1},\ldots,z_1\right)}{p_{\boldsymbol{\eta}(t)}\left(z_t\middle|z_{t-1},\ldots,z_1,x_t\right)}.$$

Using (2.22) gives

$$\frac{p_{\boldsymbol{\eta}}\left(z_t\middle|z_{t-1},\ldots,z_1\right)}{p_{\boldsymbol{\eta}}\left(z_t\middle|z_{t-1},\ldots,z_1,x_t\right)} = \Phi\left(\frac{b_t-\hat{z}_t}{r_t}\right) - \Phi\left(\frac{a_t-\hat{z}_t}{r_t}\right).$$

Therefore,

$$\frac{\boldsymbol{\phi}_{\boldsymbol{\eta}}\left(z_{1:n}\right)}{q(z_{1:n})} = \left[\Phi\left(b_1\right) - \Phi\left(a_1\right)\right]\prod_{t=2}^{n}\left[\Phi\left(\frac{b_t-\hat{z}_t}{r_t}\right) - \Phi\left(\frac{a_t-\hat{z}_t}{r_t}\right)\right].$$

Define the initial weight $w_1 = \Phi(b_1) - \Phi(a_1)$. We recursively update the weights via

$$w_t = w_{t-1}\left[\Phi\left(\frac{b_t - \hat{z}_t}{r_t}\right) - \Phi\left(\frac{a_t - \hat{z}_t}{r_t}\right)\right]$$

at time t during the sequential sampling procedure. At the end of the sampling, we obtain

$$w_n = \frac{\boldsymbol{\phi}_{\boldsymbol{\eta}}(z_{1:n})}{q_{\boldsymbol{\eta}}(z_{1:n}|x_{1:n})}.$$

In the classic GHK simulator, $\hat{Z}_t$ and r_t^2 are obtained from a Cholesky decomposition of the covariance matrix of $\{Z_t\}$. Here, they are based on the PARMA or SARMA model for $\{Z_t\}$.

Our full sequential importance sampling procedure is summarized below.

1 Initialize the process by sampling Z_1 from the $\mathcal{N}(0, 1; C_{x_1}(s(1)), C_{x_1}(s(1)))$ distribution. Define the weight w_1 by

$$w_1 = \Phi^{-1}(C_{x_1}(s(1))) - \Phi^{-1}(C_{x_1-1}(s(1))) \tag{2.24}$$

2 Now iterate steps 2 and 3 over $t = 2, 3, \ldots, n$. Conditioned on $Z_1, \ldots, Z_{t-1}$, generate

$$Z_t \overset{\mathcal{D}}{=} \mathcal{N}\left(\hat{Z}_t, r_t; \Phi^{-1}(C_{x_t}(s(t))), \Phi^{-1}(C_{x_t-1}(s(t)))\right). \tag{2.25}$$

For example, in the PAR(1) model, $\hat{Z}_t = \phi(t)Z_{t-1}$ for $t \geq 1$, with the startup condition $\hat{Z}_1 = 0$; $r_t = 1 - \phi^2(t)$ for $t > 1$ with the startup condition $r_1 = 1$.

3 Define the weight w_t for $t > 1$ via

$$w_t = w_{t-1} \left[\Phi\left(\frac{\Phi^{-1}(C_{x_t}(s(t))) - \hat{Z}_t}{r_t} \right) - \Phi\left(\frac{\Phi^{-1}(C_{x_t-1}(s(t))) - \hat{Z}_t}{r_t} \right) \right] \quad (2.26)$$

The above procedure generates a fair draw of a single "particle path" $\{Z_t\}$ with the property that $\{X_t\}_{t=1}^n$ generated from $\{Z_t\}_{t=1}^n$ yields the observations $x_1, \ldots, x_n$. Repeating this process m independent times gives m simulated process trajectories. Let $\{\mathbf{Z}^{(1)}, \ldots, \mathbf{Z}^{(m)}\}$ denote these trajectories and denote their corresponding weights at time n by $\{w_n^{(k)}\}_{k=1}^m$.

The importance sampling estimate is given by

$$\hat{\mathcal{L}}^{\mathrm{GHK}}(\boldsymbol{\theta}, \boldsymbol{\eta}) = \frac{1}{m} \sum_{k=1}^{m} w_n^{(k)}.$$

A large m provides more accurate estimation.

The popular "L-BGSF-B" gradient step and search method is used to optimize the estimated likelihood $\hat{\mathcal{L}}^{\mathrm{GHK}}(\boldsymbol{\theta}, \boldsymbol{\eta})$; other optimizers may also work. However, $\hat{\mathcal{L}}^{\mathrm{GHK}}(\boldsymbol{\theta}, \boldsymbol{\eta})$ is "noisy" due to the sampling. One popular fix to this smooths the estimated likelihood by generating a set of random quantities in the particle filtering through transformation and keeps them constant across the computations for different sets of parameters. This method, called common random numbers (CRNs), makes the simulated likelihood $\hat{\mathcal{L}}^{\mathrm{GHK}}(\boldsymbol{\theta}, \boldsymbol{\eta})$ relatively smooth in its parameters; see Kleinman et al. (1999) and Glasserman and Yao (1992) for more on CRNs. In practice, the

CRN point estimator behaves similarly to those for regular likelihoods; moreover, the Hessian-based covariance matrix, which is based on the derivative of $\hat{\mathcal{L}}^{\text{GHK}}(\boldsymbol{\theta}, \boldsymbol{\eta})$, behaves much better numerically when CRNs are used. As the next section demonstrates, this procedure will yield standard errors that are very realistic.

Particle filtering methods are known to suffer from a phenomena called particle degeneracy when n gets large. Methods to combat particle degeneracy are discussed in Djurić and Bugallo (2013), Naesseth et al. (2015), and Rebeschini and Van Handel (2015) and typically involve some sort of resampling of the particles. As our simulations in the next section did not degrade for series lengths of up to a thousand, we will not detour into this issue here.

2.4.2 Residuals and Model Diagnostics

Turning to residuals, many definitions are possible. A simple residual used here starts by computing $\hat{Z}_t = E[Z_t|X_t]$:

$$\mathbb{E}[Z_t|X_t] = \frac{\exp(-\boldsymbol{\Phi}^{-1}(C_{x_t-1})^2/2) - \exp(-\boldsymbol{\Phi}^{-1}(C_{x_t})^2/2)}{\sqrt{2\pi}(C_{x_t} - C_{x_t-1})} \tag{2.27}$$

We then use the underlying ARMA, PARMA, or SARMA difference equation to turn the $\{\hat{Z}_t\}$ into one-step-ahead prediction residuals. For example, with an AR(1) $\{Z_t\}$, the residual at time t, denoted by R_t, is the customary

$$R_t = \hat{Z}_t - \hat{\phi}\hat{Z}_{t-1}, \quad t > 1.$$

No adjustments are made for variances since it is assumed that $\text{Var}(Z_t) \equiv 1$.

As the conditional expectation in (2.27) does not contain $X_{t-1}, \ldots, X_1$ in the predicting variables (only X_t), more exotic residual definitions can be formulated and pursued. This said, these residuals are easy to compute and are sufficient for our purposes.

For additional model diagnostics, the probability integral transform (PIT) can be used as a tool to evaluate model fitness. PIT methods, proposed in Dawid (1984), check the statistical consistency between probabilistic forecasts and the observations. Under the ideal scenario that the observations are drawn from the prediction distribution and the predictive distribution is continuous, PIT residuals are uniformly distributed over $[0, 1]$. PIT histograms tend to be U-shaped when the observations are over-dispersed. Unfortunately, the above themes do not hold for discrete count data. To remedy this, Czado et al. (2009) propose a nonrandomized PIT residual where uniformity still holds.

Quantifying this, write the conditional cumulative distribution function of X_t as

$$P_t(y) := \mathbb{P}\left(X_t \leq y | X_1 = x_1, \ldots, X_{t-1} = x_{t-1}\right), y \in \{0, 1, \ldots\}. \tag{2.28}$$

Then the nonrandomized mean PIT residual is defined as $\bar{F}(u) = n^{-1} \sum_{t=1}^{n} F_t(u|x_t)$,

where

$$F_t(u|y) = \begin{cases} 0, & \text{if } u \leq P_t(y-1) \\ \dfrac{u - P_t(y-1)}{P_t(y) - P_t(y-1)}, & \text{if } P_t(y-1) < u < P_t(y) \\ 1, & \text{if } u \geq P_t(y) \end{cases} . \tag{2.29}$$

The quantity $P_t(y)$ can be approximated during the particle filtering algorithms; specifically,

$$\hat{P}_t(y) = \sum_{i=0}^{y} w_{i,t}(\hat{Z}_t), \tag{2.30}$$

where

$$w_{i,t}(z) = \Phi\left(\frac{\Phi^{-1}(C_i\left(s(t)\right)) - z}{r_t}\right) - \Phi\left(\frac{\Phi^{-1}(C_{i-1}\left(s(t)\right)) - z}{r_t}\right).$$

The weight $w_{i,t}(z)$ can be obtained at time t during the particle filtering algorithm.

2.4.3 Forecasting

Particle filtering methods can also be used to forecast. From Jia et al. (2023), the expectation of the one-step-ahead prediction of the next data point X_{n+1} from $X_1, \ldots, X_n$ is

$$\mathbb{E}\left[X_{n+1} \mid \mathcal{H}_n\right] = \frac{\int_{\mathbb{R}} \int_{\{z_t \in (a_t, b_t], t=1,\ldots,n)\}} G_{\boldsymbol{\theta}, s(t+1)}(z_{t+1}) \boldsymbol{\phi}_{\boldsymbol{\eta}}(z_{1:(n+1)}) dz_1 \ldots dz_{n+1}}{\int_{\{z_t \in (a_t, b_t], t=1,\ldots,n)\}} \boldsymbol{\phi}_{\boldsymbol{\eta}}(z_{1:n}) dz_1 \ldots dz_n}, \tag{2.31}$$

where $\mathcal{H}_n = \{X_1 = x_1, \ldots, X_n = x_n\}$ is the process history. In the particle filtering, the denominator of (2.31) is the same as that in (2.21), which is used in obtaining likelihood estimators. For the numerator in (2.31), m independent "particle paths"

$\{\mathbf{Z}^{(1)}, \ldots, \mathbf{Z}^{(m)}\}$ were drawn and can be used to make the one-step-ahead samples $Z_{n+1}^{(k)}$ for $k \in \{1, \ldots, m\}$. These are then used to approximate the mean of $G_{\boldsymbol{\theta},s(n+1)}^{(k)}(Z_{n+1}^{(k)})$. Specifically, the forecast of X_{n+1} via (2.31) is

$$\hat{\mathbb{E}}[X_{n+1} \mid \mathcal{H}_n] = \frac{\frac{1}{m}\sum_{k=1}^{m} G_{\widehat{\boldsymbol{\theta}},s(n+1)}^{(k)}\left(Z_{n+1}^{(k)}\right) w_n^{(k)}}{\frac{1}{m}\sum_{k=1}^{m} w_n^{(k)}} = \frac{\sum_{k=1}^{m} G_{\widehat{\boldsymbol{\theta}},s(n+1)}^{(k)}\left(Z_{n+1}^{(k)}\right) w_n^{(k)}}{\sum_{k=1}^{m} w_n^{(k)}}. \tag{2.32}$$

Here, $\hat{\boldsymbol{\theta}}$ contains parameter estimates for the marginal distributions and are computed from the data history $\mathcal{H}_n$.

Following the same arguments, forecasting j steps ahead can be done by sampling $Z_{n+1}, \ldots, Z_{n+j}$ sequentially. For example, Z_{n+1} is drawn from the distribution of $Z_{n+1} \mid (Z_{1:n}, \hat{\boldsymbol{\eta}})$, Z_{n+2} is drawn from the distribution of $Z_{n+2} \mid (Z_{1:(n+1)}, \hat{\boldsymbol{\eta}})$, etc.. The formula for the j-step-ahead prediction is

$$\hat{\mathbb{E}}[X_{n+j} \mid \mathcal{H}_n] = \frac{\sum_{k=1}^{m} G_{\widehat{\boldsymbol{\theta}},s(n+j)}^{(k)}\left(Z_{n+j}^{(k)}\right) w_n^{(k)}}{\sum_{k=1}^{m} w_n^{(k)}}. \tag{2.33}$$

Our next section illustrates forecasting capabilities further.

2.5 Simulations

This section presents a simulation study to evaluate the performance of our estimation methods. Periodic time series models often have a large number of parameters. One way of consolidating these parameters into a parsimonious tally involves placing Fourier parametric constraints on the model parameters (Lund et al., 2006; Anderson

et al., 2007), as is done below.

2.5.1 Poisson Marginals

Our first simulation examines the classical Poisson count distribution with the PAR(1) $\{Z_t\}$ in Example 2.1. Here, F_ν is taken as a Poisson marginal with mean $\lambda(\nu)$, where the first-order Fourier constraint

$$\lambda(\nu) = a_1 + a_2 \cos\left(\frac{2\pi(\nu - a_3)}{T}\right)$$

is imposed to consolidate the T mean parameters into three. Here, $|a_2| < a_1$ is imposed to keep $\lambda(\nu)$ non-negative. The periods $T = 10$ and $T = 50$ are studied, the latter taken to roughly correspond to our future application to weekly rainy day counts. Our $\{Z_t\}$ process obeys

$$Z_t = \phi(t)Z_{t-1} + \epsilon_t\sqrt{1 - \phi(t)^2},$$

with the AR coefficient $\phi(\nu)$ also being constrained by a first-order Fourier series that induces a causal model:

$$\phi(\nu) = b_1 + b_2 \cos\left(\frac{2\pi(\nu - b_3)}{T}\right). \tag{2.34}$$

These specifications ensure that $\{Z_t\}$ is a standard normal process ($E[Z_t] \equiv 0$ and $\mathrm{Var}(Z_t) \equiv 1$). The parameters chosen must have $\lambda(\nu)$ positive for each ν and the PAR(1) model must be causal. A six-parameter scheme that obeys these constraints is

$a_1 = 10, a_2 = 5, a_3 = 5; b_1 = 0.5, b_2 = 0.2$, and $b_3 = 5$, which is now studied.

The first row in Figure 2.2 displays quantities for this series when $T = 10$ and $n = 100$. The leftmost plot displays the raw data. Sample means (denoted by $\bar{X}_\nu$) and standard deviations (denoted by S_ν, which were computed with a denominator of 9 for each season) of the series were then computed and used to seasonally standardize the series, forming

$$U_{nT+\nu} := \frac{X_{nT+\nu} - \bar{X}_\nu}{S_\nu}. \tag{2.35}$$

The middle and rightmost plots in 2.2 display the sample correlations and partial autocorrelations of $\{U_t\}$ (while $\{U_t\}$ is not necessarily stationary, we provide the plots for feel). Note the large correlations at lag one in both plots.

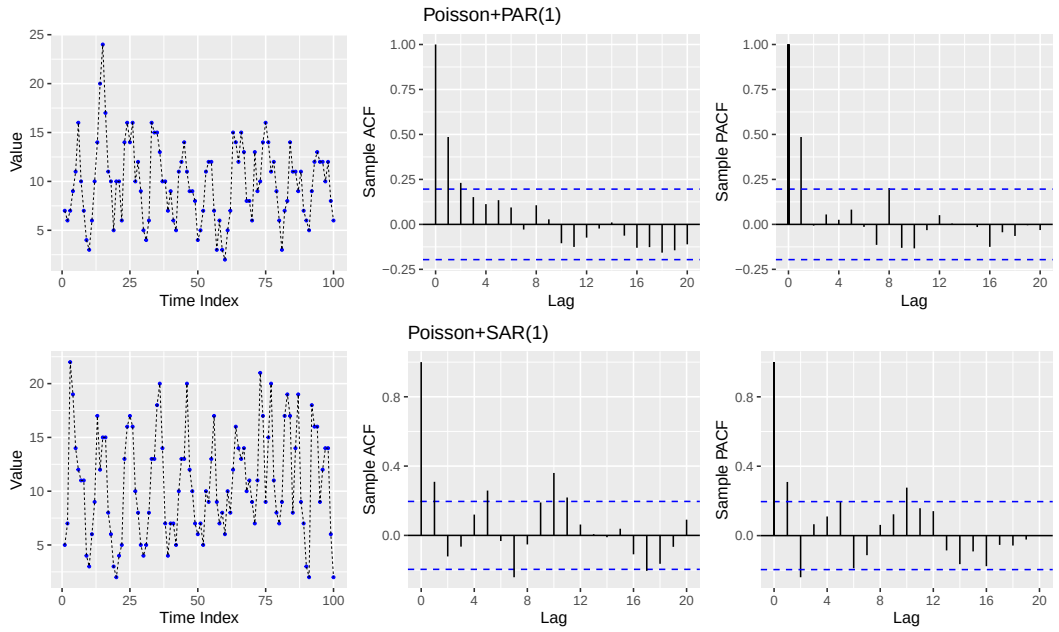

Figure 2.2: Top: A Poisson+PAR(1) realization with $n = 100$ and $T = 10$ and the sample autocorelations and partial autocorrelations of its seasonally standardized versions. Bottom: the same quantities for a Poisson+SAR(1) realization.

Moving to particle filtering parameter estimation, for each simulated series, $m = 500$ independent particles were used along with the series lengths $n = 100, 300$ and 1000. CRN techniques are used to the ensure that the likelihood is relatively smooth with respect to its parameters. This is an essential step — see Masarotto et al. (2017) and Han and De Oliveira (2018) for more on CRNs. Identifiability issues with the phase shift Fourier parametrizations arise since $a\cos(\pi/2 - b) = -a\cos(b - \pi/2)$; because of this, we impose $a_3, b_3 \in [0, T)$. Finally, the popular quasi-Newton method L-BFGS-B is implemented to optimize the likelihoods (Steenbergen, 2006). The true model parameters were used as initial guesses in our optimizations. Similar results were obtained when other initial values were used. It takes, on average, 15s, 45s, and 270s in the coding language R on a Macbook Pro to complete one simulation analysis for time series of lengths of $n = 100, 300$, and 1000, respectively.

Figure 2.3 shows boxplots of parameter estimators aggregated from 500 independent series of various lengths and periods. The sample means of the parameter estimators are all close to their true values. When $T = 50$ and $n = 100$, there are only two complete cycles of data to estimate parameters. The phase shift parameters seem to be the hardest to accurately estimate. For standard errors of these estimators, Table 2.1 reports two values: 1) sample standard deviations of the parameter estimators over the 500 runs (denominator of 499), and 2) the average (over the 500 runs) of standard errors obtained by inverting the Hessian matrix at the maximum likelihood estimate for each run (denominator of 500). Additional simulations (not shown here) with larger sample sizes with $T = 50$ show that any biases recede as n increases; we did not do any

resampling with the particle filtering methods.

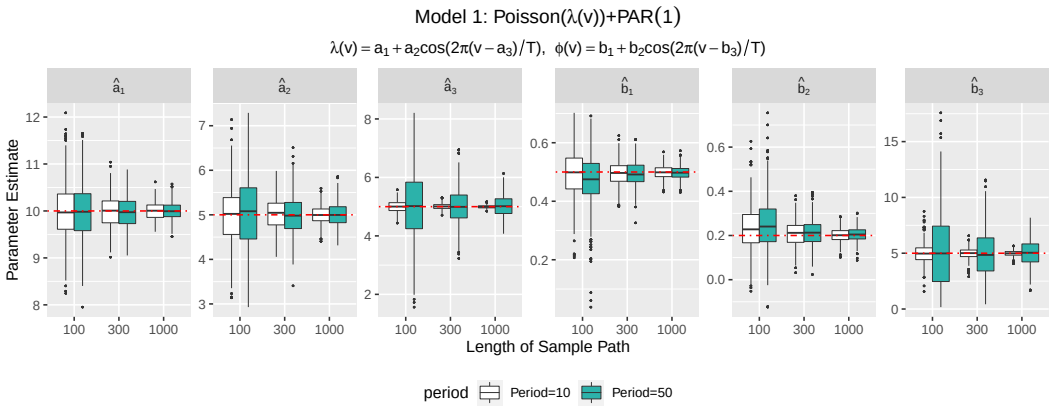

Figure 2.3: Box plots of parameter estimators for a Poisson marginal distribution with a PAR(1) $\{Z_t\}$. All estimators appear approximately unbiased — the dashed lines demarcate true parameter values.

We next consider the same Poisson marginal case, but now change $\{Z_t\}$ to the SAR(1) series in Example 2.2. The a_1, a_2, and a_3 parameters chosen for this simulation are the same as above. The SAR(1) parameters chosen are $\phi = 0.5$ and $\alpha = 0.3$. The bottom row in figure 2.2 shows a simulated trajectory of this series along with sample autocorrelations and partial autocorrelations of its seasonally standardized values as computed from (2.35). Notice the large autocorrelations at lag 10, which coincides with the period. Figure 2.4 shows boxplots of the parameter estimators akin to those in Figure 2.3. The overall performance is again very good — interpretations of the results are similar to those for the above PAR(1) model. Table 2.2 shows our two types of standard errors and again reveals nice agreement.

Model 1								
n	T		$\hat{a_1}$	$\hat{a_2}$	$\hat{a_3}$	$\hat{b_1}$	$\hat{b_2}$	$\hat{b_3}$
100	10	mean	9.98375	5.00087	5.00443	0.49213	0.23474	4.98639
		SD	0.60068	0.63494	0.18490	0.07818	0.09896	0.90434
		$\hat{E}(I'(\theta)^2)$	0.57199	0.60025	0.17375	0.07720	0.10282	0.90060
100	50	mean	9.98127	5.05051	5.01370	0.46899	0.25285	5.24743
		SD	0.58975	0.79520	1.14622	0.08861	0.11470	3.50616
		$\hat{E}(I'(\theta)^2)$	0.59190	0.76871	1.11326	0.08268	0.11469	4.51175
300	10	mean	9.98652	5.01761	5.00714	0.49608	0.20812	4.98741
		SD	0.33694	0.36520	0.10039	0.04173	0.05550	0.48356
		$\hat{E}(I'(\theta)^2)$	0.32824	0.35012	0.09958	0.04151	0.05569	0.45228
300	50	mean	9.97370	4.98609	5.00089	0.49386	0.20998	4.89220
		SD	0.33449	0.45038	0.62028	0.04188	0.05726	2.09052
		$\hat{E}(I'(\theta)^2)$	0.34588	0.45153	0.65046	0.04156	0.05584	2.33228
1000	10	mean	9.99549	4.99661	4.99529	0.49862	0.20155	4.99082
		SD	0.18216	0.20010	0.05275	0.02250	0.03078	0.23884
		$\hat{E}(I'(\theta)^2)$	0.17989	0.19258	0.05480	0.02216	0.02971	0.23532
1000	50	mean	9.99779	5.00452	5.01565	0.49721	0.20500	5.05659
		SD	0.19145	0.25633	0.35310	0.02239	0.03188	1.15546
		$\hat{E}(I'(\theta)^2)$	0.19065	0.24916	0.35300	0.02220	0.02975	1.16491

Table 2.1: Standard errors for the parameter estimators for a Poisson marginal distribution with a PAR(1) $\{Z_t\}$. The results show the sample standard deviation (SD) of the parameter estimators from 500 independent series, and the average of the 500 standard errors obtained by inverting the Hessian matrix ($\hat{E}(I'(\theta)^2)$) at the maximum likelihood estimate over these same runs.

2.5.2 A Markov Chain Induced Marginal Distribution

Another marginal distribution that we consider is derived from a two-state Markov chain (TSMC) model. This distribution will fit our weekly rainy day counts well in the next section. Consider a Markov transition matrix $\mathbf{Q}$ on two states with form

$$\mathbf{Q} = \begin{bmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{bmatrix}.$$

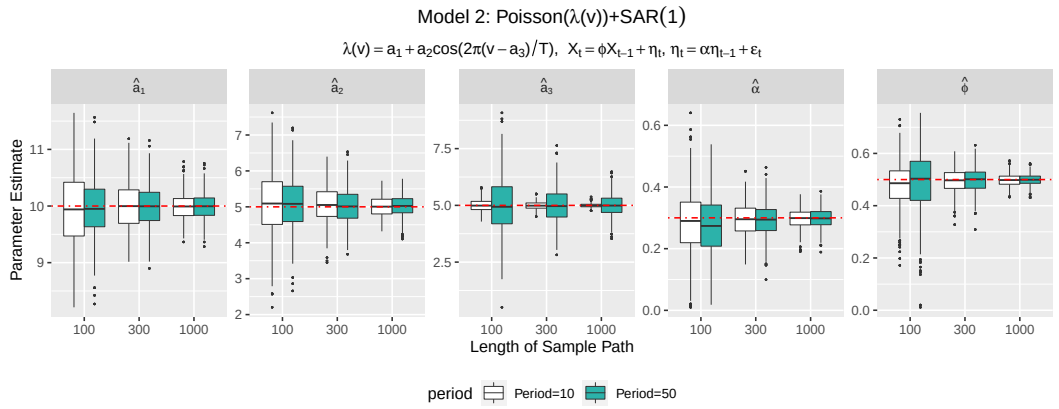

Figure 2.4: Box plots of parameter estimators for the Poisson marginal distribution with a SAR(1) $\{Z_t\}$. All estimators appear approximately unbiased — the dashed lines demarcate true parameter values.

Here, $\alpha \in (0,1)$ is interpreted as the probability that day $t+1$ is dry given that day t is dry; analogously, $\beta \in (0,1)$ is the probability that day $t+1$ is rainy given that day t is rainy. Let $\{M_t\}_{t=0}^{7}$ be a Markov chain with these transition probabilities. The marginal distribution that we consider for $\{X_t\}$ has the form

$$\mathbb{P}(X_t = k) = \mathbb{P}_{M_0}\left(\sum_{t=1}^{7} M_t = k\right), \quad k \in \{0,1,2,3,4,5,6,7\},$$

where $M_0 \in \{0,1\}$. Here, $M_0 = 0$ signifies that the day before the week started was dry and $M_0 = 1$ signifies that the day before the week started was rainy. The exact probability distribution of X_t is given in Theorem 4.1 of Minkova and Omey (2014), but is unwieldy and not particularly important to give here. This marginal distribution tends to be over-dispersed (Minkova and Omey, 2014).

To allow for periodicites in the TSMC structure, we parametrize α and β as

Model 2							
n	T		$\hat{a_1}$	$\hat{a_2}$	$\hat{a_3}$	ϕ	α
100	10	mean	9.94650	5.07987	4.99206	0.47876	0.28482
		SD	0.66096	0.87087	0.26175	0.08455	0.10242
		$\hat{E}(I'(\theta)^2)$	0.65459	0.80021	0.25874	0.08242	0.09947
100	50	mean	9.96430	5.07165	4.99623	0.49003	0.27121
		SD	0.51019	0.71742	1.16749	0.11527	0.09630
		$\hat{E}(I'(\theta)^2)$	0.50809	0.68786	1.10774	0.10345	0.09874
300	10	mean	9.99120	5.06420	4.99278	0.49535	0.29260
		SD	0.41347	0.50127	0.15764	0.04399	0.05155
		$\hat{E}(I'(\theta)^2)$	0.40803	0.50201	0.15827	0.04269	0.05426
300	50	mean	9.99133	5.01493	5.01603	0.49874	0.29189
		SD	0.37005	0.49076	0.78444	0.04674	0.05392
		$\hat{E}(I'(\theta)^2)$	0.36755	0.49687	0.79666	0.04543	0.05403
1000	10	mean	9.99023	5.01476	4.99412	0.49789	0.29760
		SD	0.23726	0.27841	0.08863	0.02371	0.03056
		$\hat{E}(I'(\theta)^2)$	0.22981	0.28191	0.08943	0.02281	0.02930
1000	50	mean	9.99384	5.01830	4.99809	0.49863	0.29815
		SD	0.22794	0.29628	0.47654	0.02124	0.03018
		$\hat{E}(I'(\theta)^2)$	0.22244	0.30013	0.47752	0.02323	0.02926

Table 2.2: Standard errors for the parameter estimators for the Poisson marginal distribution with a SAR(1) $\{Z_t\}$. The results show the sample standard deviation (SD) of the parameter estimators from 500 independent series, and the average of the 500 standard errors obtained by inverting the Hessian matrix ($\hat{E}(I'(\theta)^2)$) at the maximum likelihood estimate over these same runs.

short Fourier series again:

$$\alpha(\nu) = c_1 + c_2 \cos\left(\frac{2\pi(\nu - c_3)}{T}\right); \quad \beta(\nu) = d_1 + d_2 \cos\left(\frac{2\pi(\nu - d_3)}{T}\right).$$

Our first TSMC simulation considers the PAR(1) $\{Z_t\}$ in (2.34). This is a nine parameter model. The parameter values considered are $c_1 = 0.4, c_2 = 0.2, c_3 = 5; d_1 = 0.5, d_2 = 0.2, d_3 = 0.3, b_1 = 0.2, b_2 = 0.1$, and $b_3 = 5$, which induce a causal $\{Z_t\}$ and legitimate Markov chain transitions (all transitions have non-negative probabilities).

Plot of simulated trajectories and its respective ACF and PACF are given by top row in figure (2.5).

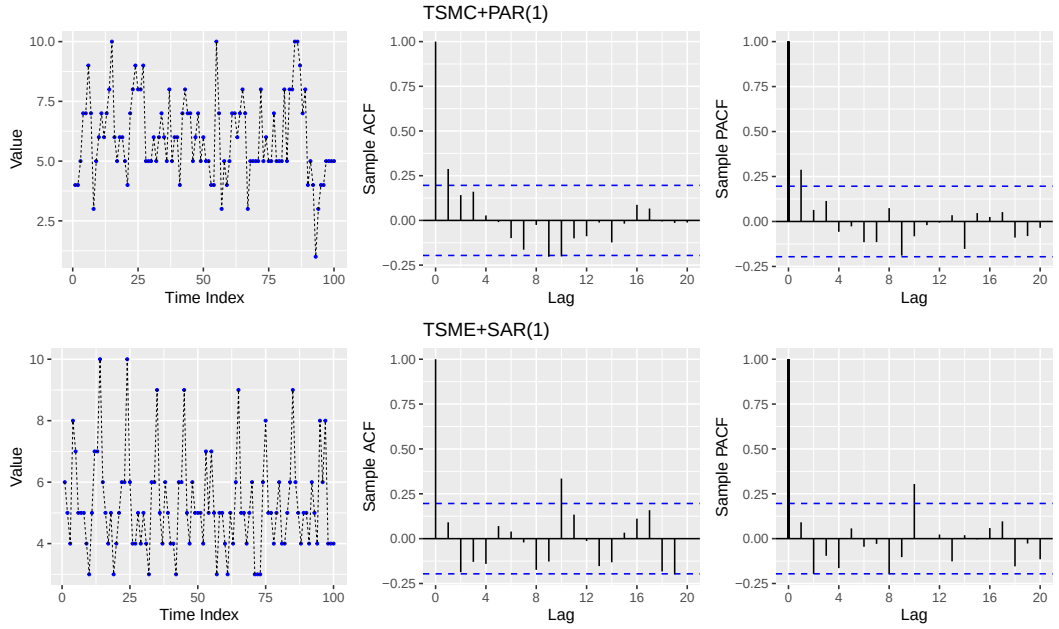

Figure 2.5: Top: A TSMC+PAR(1) realization with $n = 100$ and $T = 10$ and sample autocorelations and partial autocorrelations of its seasonally standardized $\{U_t\}$. Bottom: the same quantities for a TSMC+SAR(1) realization.

Figure 2.6 shows boxplots of estimated parameters over 500 independent series of various lengths and periods. Table 2.3 shows standard errors computed from the two methods previously described. For the most part, the results are satisfying. Some of the phase shift parameter's "Hessian inverted" standard errors are larger than the sample standard deviation standard errors. The phase shift parameter is the argument where its associated cosine wave is maximal and lies in $[0, T]$. Owing to its larger possible range, this parameter will have more variability than say parameters supported in (-1,1). Also, when $n = 100$ and $T = 50$, there are only two complete cycles from which to

estimate the location of this maximum — this will be statistically difficult. Additional simulations (not reported) show that these discrepancies recede as the sample size gets larger.

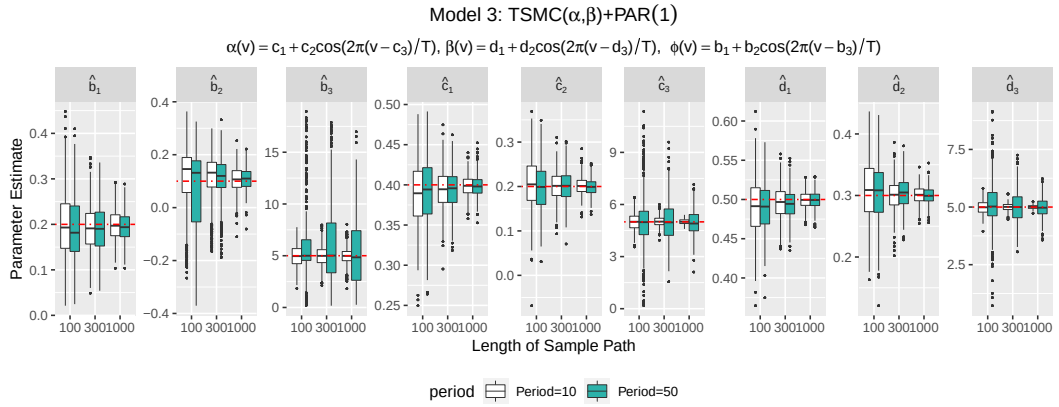

Figure 2.6: Box plots of parameter estimators for the TSMC marginal distribution with a PAR(1) $\{Z_t\}$. All estimators appear approximately unbiased — the dashed lines demarcate true parameter values.

Finally, we consider the TSMC marginal distribution with a SAR(1) $\{Z_t\}$. The parameters $\phi = 0.5$ and $\alpha = 0.3$ were used in this simulation study. The bottom row in Figure 2.5 shows a simulated trajectory and its respective sample autocorrelations and partial autocorrelations. Figure 2.7 shows boxplots of the estimated parameters over 500 independent series of various lengths and periods. Table 2.4 shows standard errors computed by our two methods. Again, the performance is good — the interpretation of the results is analogous to that given above.

Figures 2.8 demonstrate the forecasting capabilities of our methods. Sample series were first simulated from the Poisson+PAR(1) and TSMC+PAR(1) models. The parameters chosen are those in Sections 5.1 and 5.2. The first $n = 100$ data points were

Model 3											
n	T		$\hat{c_1}$	$\hat{c_2}$	$\hat{c_3}$	$\hat{d_1}$	$\hat{d_2}$	$\hat{d_3}$	$\hat{b_1}$	$\hat{b_2}$	$\hat{b_3}$
100	10	mean	0.389	0.205	4.997	0.490	0.307	4.985	0.199	0.105	4.963
		SD	0.041	0.060	0.472	0.036	0.048	0.295	0.076	0.134	1.221
		$\hat{E}(I'(\theta)^2)$	0.043	0.059	0.550	0.035	0.045	0.280	0.108	0.178	2.275
100	50	mean	0.391	0.199	5.009	0.490	0.302	5.109	0.192	0.072	5.994
		SD	0.040	0.057	1.876	0.032	0.050	1.137	0.072	0.153	4.153
		$\hat{E}(I'(\theta)^2)$	0.044	0.065	3.210	0.036	0.049	1.603	0.112	0.195	4.718
300	10	mean	0.394	0.200	5.013	0.496	0.301	5.002	0.194	0.114	4.969
		SD	0.024	0.033	0.306	0.021	0.026	0.159	0.052	0.084	1.095
		$\hat{E}(I'(\theta)^2)$	0.025	0.033	0.291	0.020	0.026	0.159	0.060	0.087	1.376
300	50	mean	0.395	0.201	4.989	0.495	0.304	5.017	0.191	0.108	5.989
		SD	0.024	0.035	1.269	0.020	0.026	0.733	0.053	0.085	4.130
		$\hat{E}(I'(\theta)^2)$	0.025	0.034	1.506	0.020	0.027	0.823	0.060	0.097	8.054
10^3	10	mean	0.399	0.201	5.008	0.499	0.301	5.004	0.198	0.106	4.995
		SD	0.013	0.020	0.151	0.011	0.015	0.085	0.032	0.045	0.794
		$\hat{E}(I'(\theta)^2)$	0.014	0.018	0.155	0.011	0.014	0.086	0.033	0.046	0.888
10^3	50	mean	0.398	0.198	4.949	0.499	0.300	4.985	0.195	0.109	5.337
		SD	0.014	0.018	0.743	0.011	0.014	0.432	0.032	0.042	3.477
		$\hat{E}(I'(\theta)^2)$	0.014	0.018	0.799	0.011	0.015	0.445	0.033	0.047	4.111

Table 2.3: Standard errors for the parameter estimators for the TSMC marginal distribution with a PAR(1) $\{Z_t\}$. The results show the sample standard deviation (SD) of the parameter estimators from 500 independent series, and the average of the 500 standard errors obtained by inverting the Hessian matrix ($\hat{E}(I'(\theta)^2)$) at the maximum likelihood estimate over these same runs.

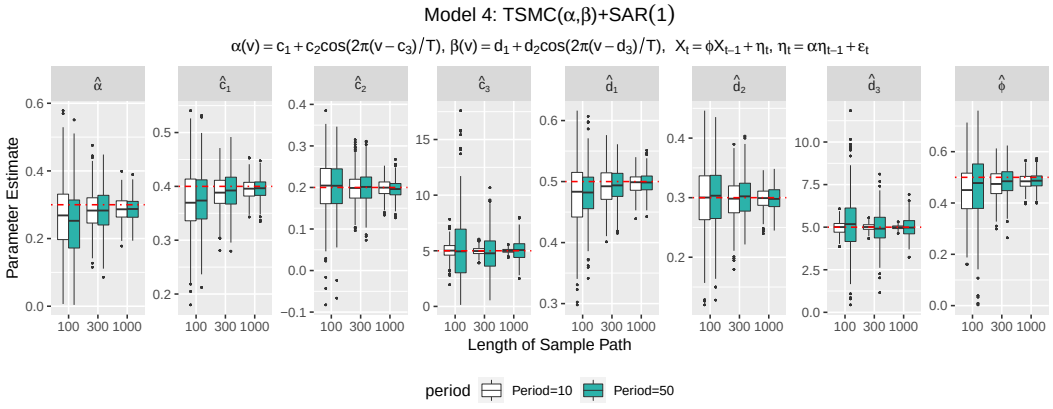

Figure 2.7: Box plots of parameter estimators for the TSMC marginal distribution with a SAR(1) $\{Z_t\}$. All estimators appear approximately unbiased — the dashed lines demarcate true parameter values.

Model 4										
n	T		$\hat{c_1}$	$\hat{c_2}$	$\hat{c_3}$	$\hat{d_1}$	$\hat{d_2}$	$\hat{d_3}$	ϕ	α
100	10	mean	0.372	0.200	4.995	0.480	0.299	5.010	0.447	0.260
		SD	0.061	0.068	0.605	0.050	0.056	0.356	0.107	0.111
		$\hat{E}(I'(\theta)^2)$	0.055	0.063	0.624	0.046	0.053	0.335	0.102	0.109
100	50	mean	0.378	0.199	4.890	0.482	0.299	5.106	0.469	0.250
		SD	0.049	0.060	2.807	0.044	0.048	1.564	0.123	0.104
		$\hat{E}(I'(\theta)^2)$	0.049	0.062	3.602	0.040	0.050	1.572	0.121	0.107
300	10	mean	0.388	0.201	5.006	0.493	0.302	5.004	0.477	0.289
		SD	0.033	0.038	0.317	0.030	0.032	0.202	0.056	0.063
		$\hat{E}(I'(\theta)^2)$	0.033	0.038	0.335	0.028	0.032	0.193	0.054	0.060
300	50	mean	0.390	0.199	5.030	0.492	0.302	5.016	0.483	0.281
		SD	0.033	0.040	1.689	0.027	0.031	1.061	0.055	0.059
		$\hat{E}(I'(\theta)^2)$	0.032	0.038	1.706	0.027	0.032	0.960	0.056	0.060
1000	10	mean	0.395	0.198	4.992	0.497	0.299	4.998	0.484	0.286
		SD	0.019	0.021	0.183	0.016	0.019	0.112	0.028	0.035
		$\hat{E}(I'(\theta)^2)$	0.018	0.021	0.181	0.016	0.018	0.107	0.029	0.032
1000	50	mean	0.395	0.196	5.060	0.498	0.299	4.989	0.486	0.286
		SD	0.019	0.022	0.898	0.017	0.020	0.553	0.030	0.034
		$\hat{E}(I'(\theta)^2)$	0.018	0.022	0.947	0.015	0.019	0.555	0.029	0.032

Table 2.4: Standard errors for the parameter estimators for the TSMC marginal distribution with a SAR(1) $\{Z_t\}$. The results show the sample standard deviation (SD) of the parameter estimators from 500 independent series, and the average of the 500 standard errors obtained by inverting the Hessian matrix ($\hat{E}(I'(\theta)^2)$) at the maximum likelihood estimate over these same runs.

used to fit the model and a period of $T = 10$ additional observations were both simulated and forecasted from the fist 100 series values. In the Figure 2.8 plots, the forecasts track the simulated data reasonably well. Confidence intervals for the predictions could be found by finding the quantile of the sample $z_{n+j}^{(k)}$, $k \in \{1, \cdots, m\}$ of Section 4.3, but we do not pursue this issue further here.

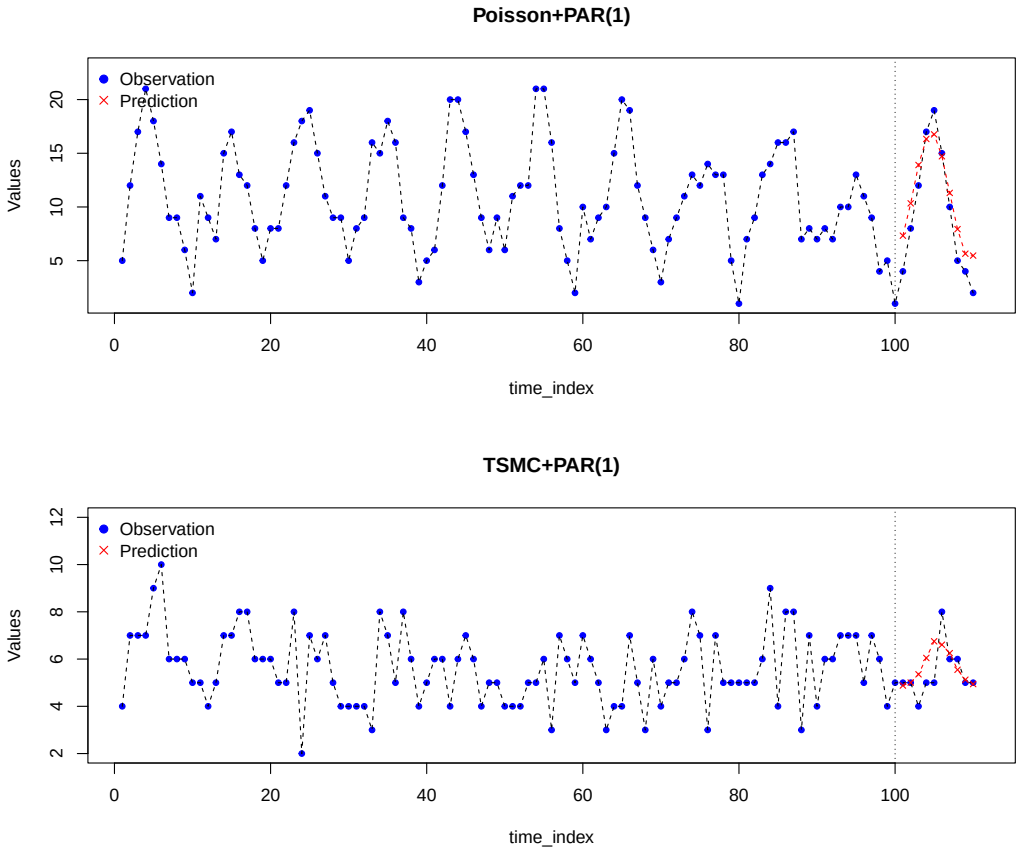

Figure 2.8: Top: A forecasting simulation for the Poisson+PAR(1) model. The parameters chosen are $a_1 = 10, a_2 = 5, a_3 = 5$; and $b_1 = 0.5, b_2 = 0.2, b_3 = 5$ and $T = 10$. Bottom: A forecasting simulation for the TSMC+PAR(1) model. The parameters chosen are $c_1 = 0.4, c_2 = 0.2, c_3 = 5$, $d_1 = 0.5, d_2 = 0.2, d_3 = 0.3$, $b_1 = 0.2, b_2 = 0.1, b_3 = 5$, and $T = 10$. The number of TSMC trials chosen is 10. In both plots, the blue dots are the observations and the red dots are the predictions. The last 10 data points were not used to fit the model. The forecasted observations reasonably track the simulated data.

2.6 Application

This section applies our techniques to a series of weekly rainy day counts in the Seattle, Washington area recorded from 01-Jan-2000 to 31-Dec-2019. The data were collected at the Seattle Tacoma airport weather station and are available at `http:`

//www.ncdc.noaa.gov. Here, any day receiving a non-zero amount of precipitation is counted as a rainy day. As such, $X_t \in \{0, 1, 2, 3, 4, 5, 6, 7\}$ for each t. For convenience, we only analyze the first 364 days in a year, inducing a period of $T = 52$ in the series. Any inaccuracies incurred by neglecting these days is minimal. Figure 2.9 summarizes our data. The top plot graphs the weekly rainy day counts from the last four years of the series only (for visual clarity), from the first week in 2016 to the 52nd week in 2019. One sees a clear seasonal cycle with summer weeks experiencing significantly less rain than winter weeks. The middle plot in the figure displays the sample mean and variance of the weekly counts over the entire 20 year data period, aggregated by week of year. For example, the mean and variance for the 1st week in January are the sample means and variance (denominator of 19) over all 20 1st weeks occurring from 2000 to 2019. The sample mean and variance have roughly sinusoidal structures and are minimal during the summer months. The bottom plots in the figure show sample autocorrelations (ACF) and partial autocorrelations (PACF) of the series. The pattern in the ACF is indicative of a periodic mean in the series that has not been removed.

Several marginal distributions for this series merit exploration. The binomial distribution with seven trials is a classic structure for such data. However, this distribution is underdispersed (variance is smaller than the mean), which does not jibe with the data patterns seen in the middle plot of Figure 2.9. Another distribution considered is the TSMC distribution of the last section. This distribution can be overdispersed, and as we will subsequently see, fits our series quite well. Other two ideal marginal distributions considered are the generalized Poisson marginal truncated to the support

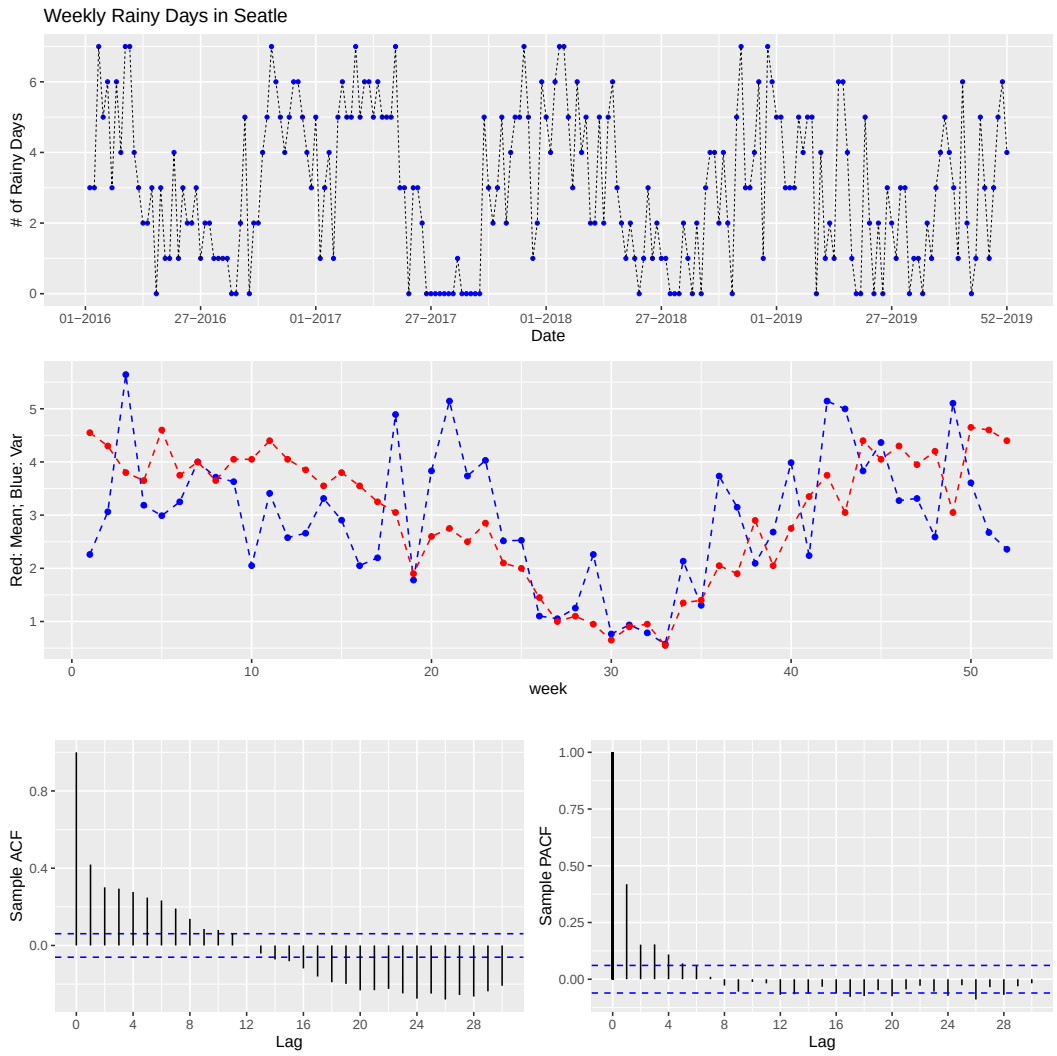

Figure 2.9: Top: The Seattle weekly rainy day counts from 2016-2019 only; Middle: Weekly sample means and variances for the rainy day counts from 2000-2019; Bottom left and bottom right: Sample ACF and PACF of all observations.

set $\{0, 1, 2, 3, 4, 5, 6, 7\}$ and the beta-binomial distribution. For clarity, the generalized

Poisson marginal we use has distribution

$$\begin{aligned}
\mathbb{P}(Y=k) &= \frac{e^{-(\lambda+\eta k)}\lambda(\lambda+\eta k)^{k-1}}{k!}, \quad k=0,1,\ldots; \\
\mathbb{E}(Y) &= \mu = \frac{\lambda}{1-\eta}; \\
\mathrm{Var}(Y) &= \sigma^2 = \frac{\lambda}{(1-\eta)^3}
\end{aligned}$$

for a count variable Y, with $\lambda > 0$ and $\eta \in [0,1)$. When $\eta = 0$, Y is Poisson(λ) and is equi-dispersed. The beta-binomial distribution we use has the form

$$\begin{aligned}
\mathbb{P}(Y=k) &= \binom{r}{k} \frac{\boldsymbol{B}(k+\alpha, r-k+\beta)}{\boldsymbol{B}(\alpha,\beta)}, \quad k=0,1,\ldots,r; \\
\mathbb{E}(Y) &= \frac{r\alpha}{\alpha+\beta}, \\
\mathrm{Var}(Y) &= \frac{r\alpha\beta(\alpha+\beta+r)}{(\alpha+\beta)^2(\alpha+\beta+1)},
\end{aligned}$$

where $\boldsymbol{B}(x,y) = \Gamma(x)\Gamma(y)/\Gamma(x+y)$ is the beta function and $r = 7$ denotes the number of trials. First order Fourier cosine constraints are placed on the mean and variance pair $(\mu(\nu), \sigma^2(\nu))$ for both cases and then mapped back to parameter pair $(\lambda(\nu), \alpha(\nu))$ and $(\beta(\nu), \alpha(\nu))$.

For structures of $\{Z_t\}$, we consider PAR(1), AR(1), and SAR(1) models (see Section 2). The PAR(1) structure uses the first order Fourier cosine consolidation in (2.34) for $\{Z_t\}$. The AR(1) $\{Z_t\}$ is simply a standard AR(1) series with a unit variance. The SAR(1) form for $\{Z_t\}$ is the two parameter model in Example 2.2. For parameters

in the marginal distributions, the success probabilities are

$$p(\nu) = a_1 + a_2 \cos\left(\frac{2\pi(a_3 - \nu)}{T}\right)$$

in the binomial fits;

$$\alpha(\nu) = a_1 + a_2 \cos\left(\frac{2\pi(a_3 - \nu)}{T}\right), \quad \beta(\nu) = b_1 + b_2 \cos\left(\frac{2\pi(b_3 - \nu)}{T}\right),$$

for the TSMC fits,

$$\mu(\nu) = a_1 + a_2 \cos\left(\frac{2\pi(a_3 - \nu)}{T}\right), \qquad \sigma^2(\nu) = b_1 + b_2 \cos\left(\frac{2\pi(b_3 - \nu)}{T}\right),$$
$$\lambda(\nu) = 1 - \sqrt{\frac{\mu(\nu)}{\sigma^2(\nu)}}, \qquad \alpha(\nu) = \mu(\nu)\sqrt{\frac{\mu(\nu)}{\sigma^2(\nu)}}$$

in the truncated generalized Poisson fit, and

$$\mu(\nu) = a_1 + a_2 \cos\left(\frac{2\pi(a_3 - \nu)}{T}\right), \qquad \sigma^2(\nu) = b_1 + b_2 \cos\left(\frac{2\pi(b_3 - \nu)}{T}\right),$$
$$\beta(\nu) = \frac{r[r - \mu(\nu)]\sigma^2(\nu) - r\mu(\nu)[r - \mu(\nu)]^2}{\mu(\nu)r[r - \mu(\nu)] - r^2\sigma^2(\nu)}, \qquad \alpha(\nu) = \frac{\beta(\nu)\mu(\nu)}{r - \mu(\nu)}$$

in the beta-binomial fits. We have used the letters a and b and subscripts in all marginal distribution parameterizations. Do not confuse these with the parameters in the $\{Z_t\}$ process of the last section; indeed we will see that a stationary $\{Z_t\}$ suffices in our latent model.

Table 2.5 displays BIC and AIC scores for various fitted $\{Z_t\}$ structures and

marginal distributions. The best marginal distribution is the TSMC; the truncated generalized Poisson and marginal distribution is a close second and the beta-binomial distribution is third. The generalized Poisson marginal distribution (non-truncated) is known to be a very flexible count distribution (Ver Hoef and Boveng, 2007) that fits many observed series well. Of note is that an AR(1) latent $\{Z_t\}$ is preferred to either a PAR(1) or SAR(1) structure. This does not mean that the end fitted model is non-periodic; indeed, the parameters in the marginal distribution F_ν will depend highly on the week of year ν. However, any seasonality in the PAR(1) and SAR(1) $\{Z_t\}$ do not make an appreciable difference — a stationary AR(1) $\{Z_t\}$ is sufficient.

Marginal Distribution	Model	WN	AR(1)	PAR(1)	SAR(1)
Binomial	AIC	4278.628	**4227.775**	4229.708	4229.458
	BIC	4293.469	**4247.563**	4259.39	4254.193
Two State Markov Chain (TSMC)	AIC	3888.114	**3853.589**	3856.244	3855.624
	BIC	3917.796	**3888.218**	3900.766	3895.200
Truncated Overdispersed Poisson	AIC	3885.032	**3853.840**	3856.995	3855.672
	BIC	3914.714	**3888.469**	3901.518	3895.248
Beta-Binomial	AIC	3902.095	**3864.795**	3866.964	3906.130
	BIC	3931.777	**3899.424**	3911.487	3945.706

Table 2.5: AIC and BIC statistics for models with binomial, TSMC, truncated Poisson, and beta-binomial marginal distributions. The lowest AIC/BIC for each marginal distribution are bolded. The TSMC marginal distribution with an AR(1) $\{Z_t\}$ is judged optimal.

Table 2.6 shows the estimated parameters in the fitted model, along with our initial guesses for the parameters (the routines converged to the same parameter estimates for all initial parameter guesses tried — this was not an issue in our likelihood optimizations). Based on asymptotic normality, which is expected but has not been proven, all parameters except b_3 appear to be significantly non-zero. Here, a zero b_3 is plausible: $b_3 = 0$ implies that the maximal variability of the weekly rainy day counts

start at the beginning of the calendar year, which roughly corresponds to the meteorological height of winter. Standard errors were estimated by inverting the Fisher information matrix at the likelihood estimates. For completeness, Tables 2.7 and 2.8 show initial guesses, parameter estimates, and standard errors for the truncated generalized Poisson and beta-binomial fits. The interpretation of these results are similar to that for Table 2.6.

Table 2.6: Estimates and standard errors of the TSMC AR(1) model. The L-BFGS-B algorithm was used to optimize particle filtering likelihoods.

Parameters	a_1	a_2	a_3	b_1	b_2	b_3	ϕ
Initial Guess	0.500	0.000	3.000	0.500	0.150	3.000	0.000
Point Estimates	0.737	-0.163	4.687	0.648	0.132	1.660	0.198
Standard Error	0.011	0.014	0.674	0.013	0.018	1.039	0.032

Table 2.7: Estimates and standard errors of the generalized Poisson-AR(1) fit. The L-BFGS-B algorithm is used to optimize particle filtering likelihoods.

Parameters	a_1	a_2	a_3	b_1	b_2	b_3	ϕ
Initial Guess	5.000	3.000	4.500	7.000	5.000	3.000	0.000
Point Estimates	3.999	2.975	3.977	8.155	6.926	3.955	0.195
Standard Error	0.263	0.298	0.448	1.586	1.685	0.926	0.033

Table 2.8: Estimates and standard errors of the beta-binomial AR(1) model. The L-BFGS-B algorithm was used to optimize particle filtering likelihoods.

Parameters	a_1	a_2	a_3	b_1	b_2	b_3	ϕ
Initial Guess	4.000	1.000	3.000	3.000	0.500	0.300	0.000
Point Estimates	2.996	1.504	3.427	3.109	0.503	0.339	0.205
Standard Error	0.066	0.085	0.495	0.124	0.174	2.548	0.032

Moving to a residual analysis, Figure 2.10 shows diagnostics for the TSMC marginal with an AR(1) $\{Z_t\}$. The top left plot shows the raw residuals and the bottom left and right plots show sample ACFs and PACFs of these residuals. No major issues are seen. The top right graph shows a QQ plot of these residuals for a standard normal

distribution. Some departure from normality is noted in the two tails of the plot.

Figure 2.11 shows PIT histograms of the residuals for the binomial, TSMC, and beta-binomial fits with AR(1) $\{Z_t\}$. There are obvious departures from uniformity for the binomial marginal — this marginal distribution does not seem to describe the data well. The histogram for the two-state Markov chain marginal is roughly uniform; hence, we have a good fitting model and the slight lack of normality in $\{Z_t\}$ in Figure 2.10 does not appear overly problematic. The beta-binomial PIT histogram also looks reasonable.

Finally, to get a feel for the correlation level in our fitted model. Let us define the average autocorrelation at lag h over all seasons: $\hat{\rho}_X(h) = \sum_{\nu=1}^{52} \rho_X(\nu, \nu+h)/52$, where $\rho_X(\nu, \nu+h)$ is as in 2.14, Table 2.6 shows that the average autocorrelation of the count model is slightly lower than that of the latent model. No periodic pattern is observed in the ACF since the latent model is simply AR(1). This result suggests that the connection between two variables can be attributed to seasonal elements in the data.

lag h	0	1	2	3	4	5	6	7
$\hat{\phi}^h$	1	0.1950	0.0380	0.0074	0.0014	0.0003	0.0001	0.0000
$\hat{\rho}_X(h)$	0.9687	0.1811	0.0351	0.0068	0.0013	0.0003	0.0001	0.0000

Table 2.9: Average autocorrelation of the fitted TSMC AR(1) model.

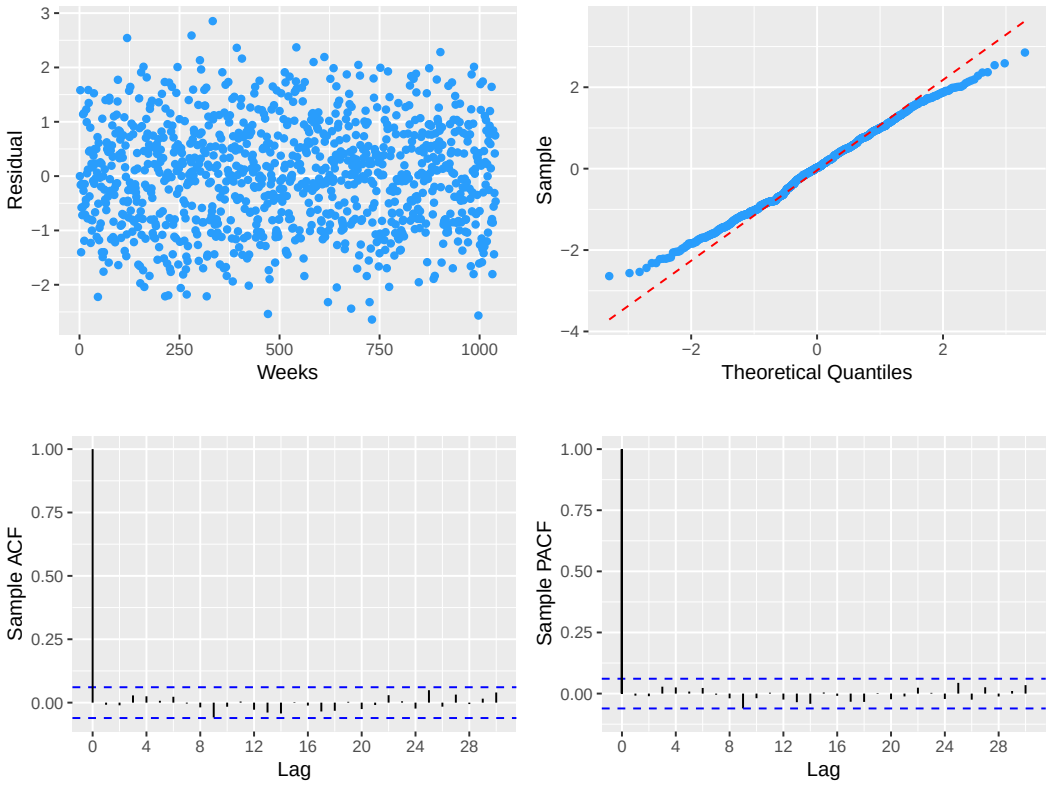

Figure 2.10: Top left: TSMC + AR(1) residuals. Top right: A QQ plot of these residuals. Bottom left: The sample ACF of these residuals. Bottom right: The sample PACF of these residuals.

2.7 Concluding Comments

The above chapter constructs a very general model for seasonal count time series through a latent Gaussian process transformation. Any marginal distribution can be achieved and the correlations are as flexible as possible and can be negative. Estimation of the model parameters through likelihood techniques can be conducted by particle filtering techniques. The methods were shown to to work well on simulated data and capably handled a periodic count sequence supported on $\{0, 1, 2, 3, 4, 5, 6, 7\}$. There, we found that the latent Gaussian process did not need to be periodic, but the

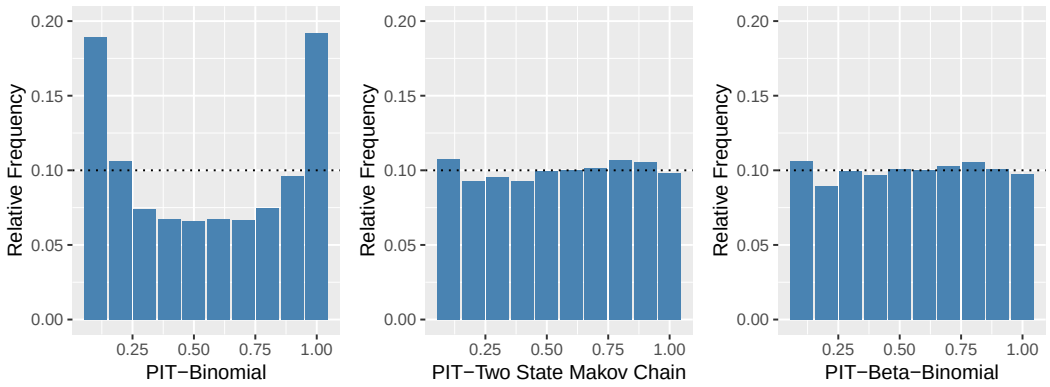

Figure 2.11: Left: a binomial marginal PIT histogram. Middle: a TSMC marginal PIT histogram. Right: a beta-binomial marginal PIT histogram. The binomial marginal does not fit as well as the TSMC and beta-binomial marginals.

marginal distribution of the data contained periodicities. The fitted model was very parsimonious, containing only seven parameters.

Extensions of the above techniques to handle covariates merit exploration. For this, we suggest allowing $\boldsymbol{\theta}$ to depend on the covariates as in Jia et al. (2023). Modifying the latent Gaussian process to handle covariates generally causes trouble. Another extension of the methods would consider periodic series supported on all integers: $\{\ldots, -2, -1, 0, 1, 2, \ldots\}$; see Barreto-Souza and Bourguignon (2015) and Cui et al. (2021) and the references within for more on the stationary case. Multivariate versions of the methods are also worth studying.

Chapter 3

Poisson Count Time Series

3.1 Introduction

Gaussian time series have a long and storied development in time series modeling (Box et al., 2015; Brockwell and Davis, 1991a; Shumway et al., 2000). Indeed, most time series connoisseurs regard Gaussian theory as essentially complete now. Less developed, but now currently heavily researched, are methods that describe autocorrelated series for count data; that is, the series X_t at time t is supported on some subset of the non-negative integers $\{0, 1, \ldots\}$. This chapter reviews, compares, and contrasts several popular methods that produce Poisson distributed series, arguably the quinessential count distribution. In particular, discrete and integer autoregressions, superpositioning methods, and copula methods are considered here.

Some caveats are worth mentioning at the onset. First, techniques exist that produce count models having a conditional Poisson distribution. One such technique,

which is essentially the GLARMA paradigm of Davis et al. (2005); Dunsmuir et al. (2015), starts with a nonnegative $\{\lambda_t\}$ process that is stationary and posits that the conditional distribution of $X_t|\lambda_t$ is Poisson with mean λ_t. While this can often lead to a convenient autoregressive representation of the counts (Fokianos et al., 2009b), X_t will not be left with a Poisson marginal distribution. Indeed, should $X|\Lambda = \lambda$ be a Poisson distributed with mean λ, then the marginal distribution of this structure must be overdispersed:

$$\text{Var}(X) = E[\text{Var}(X|\Lambda)] + \text{Var}(E[X|\Lambda]) = E[\Lambda] + \text{Var}(\Lambda) > E[\Lambda] = E[X].$$

We refer the reader to the review in Davis et al. (2021) and the references within for more on these issues. In particular, this chapter focuses on models having a true Poisson count marginal distribution.

As a second caveat, some results for stationary Gaussian time series do not hold in the Poisson setting. For one example, if $\{\gamma(h)\}_{h=-\infty}^{\infty}$ is a symmetric ($\gamma(h) = \gamma(-h)$ for all integers h) and non-negative definite sequence on the integers, then there exists a Gaussian distributed sequence $\{X_t\}$ with $\text{Cov}(X_t, X_{t+h}) = \gamma(h)$. No such result carries over to the Poisson case. Indeed, $(-1)^h$ is symmetric and non-negative definite. While a Gaussian sequence with this autocovariance exists — take $X_t = (-1)^t Z$, where Z is standard normal — it is not possible to achieve this in the Poisson setting. To see this, it is enough to show that one cannot have two Poisson variables X_1 and X_2 having the same mean λ and correlation -1 (the reader is challenged to prove this).

The rest of this chapter proceeds as follows. The next section reviews methods that generate count series having a Poisson marginal distribution. There, discrete and integer autoregressions, superpositioning methods, and copula techniques are considered. The pros and cons of these model classes are outlined; much of this material constitutes a review. Section 3.3 moves to estimation issues. There, likelihood estimation techniques are developed if possible. Unfortunately, the joint distribution needed in the likelihood is not tractable for some model classes. Particle filtering and quasi-likelihood techniques such as linear prediction arise here. Simulations show that the methods work quite well. Section 3.4 analyzes series of annual North Atlantic Basin strong hurricanes and no-hitter games pitched in major league baseball. Section 3.5 concludes the chapter with comments.

3.2 Methods

This section reviews methods producing a stationary series with Poisson marginal distributions. Some of this material has appeared elsewhere, but some new insights are offered in our discourse.

As some of the models classes reviewed cannot have negative autocorrelations, flexibility and completeness of the autocovariances becomes an issue. Before proceeding, we first investigate the most negatively correlated Poisson variables existing, providing some intuition en route.

Let $F_\lambda(\cdot)$ be the Poisson cumulative distribution function (CDF) with mean

λ:

$$F_\lambda(n) = \sum_{k=0}^{n} \frac{e^{-\lambda}\lambda^k}{k!}, \quad n = 0, 1, \ldots$$

The most negatively correlated pair of random variables X and Y, both having the marginal cumulative distribution function (CDF) F_λ, are known to have form $X = F_\lambda^{-1}(U)$ and $Y = F_\lambda^{-1}(1-U)$, where U is uniformly distributed over [0,1] and F_λ^{-1} is the inverse CDF:

$$F_\lambda^{-1}(u) = \inf\{t : F_\lambda(t) \geq u\}$$

(Whitt, 1976) (this inverse version coincides with the quantile function). Such an (X, Y) pair can be produced from a Gaussian copula via

$$X = F_\lambda^{-1}(\Phi(Z)), \quad Y = F_\lambda^{-1}(\Phi(-Z)).$$

Here, $\Phi(\cdot)$ is the standard normal CDF and Z is a standard normal random variable. This is because $U := \Phi(Z)$ is uniformly distributed over [0,1] by the probability transformation theorem and $\Phi(-Z) = 1 - \Phi(Z)$ has the same distribution as $1 - U$ due to the symmetry of the standard normal distribution about zero.

To obtain an expression for the most negative correlation possible, let $c_n = F_\lambda(n)$ denote the Poisson λ CDF at index n and note that the inverse has form

$$F_\lambda^{-1}(u) = \sum_{n=1}^{\infty} n 1_{[c_{n-1}, c_n)}(u),$$

where $1_A(x)$ denotes an indicator function over the set A. Converting this to a tail sum gives

$$F_\lambda^{-1}(u) = \sum_{n=1}^{\infty}\sum_{k=1}^{n} 1_{[c_{n-1},c_n)}(u) = \sum_{k=1}^{\infty}\sum_{n=k}^{\infty} 1_{[c_{n-1},c_n)}(u) = \sum_{k=1}^{\infty} 1_{[c_{k-1},1)}(u). \tag{3.1}$$

Simple algebraic manipulations now give

$$E[F_\lambda^{-1}(U)F_\lambda^{-1}(1-U)] = \sum_{k=1}^{\infty}\sum_{\ell=1}^{\infty} E[1_{c_{k-1},1)}(U)1_{c_{\ell-1},1)}(1-U)] = \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty}(1-c_\ell-c_k)1_{[c_\ell+c_k<1]},$$

An expression for the most negative autocorrelation, which we denote by $\mathrm{NB}(\lambda)$, now follows simply as

$$\mathrm{NB}(\lambda) = \frac{\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty}(1-c_\ell-c_k)1_{[c_\ell+c_k<1]} - \lambda^2}{\lambda}. \tag{3.2}$$

A plot of $\mathrm{NB}(\lambda)$ as a function of λ is provided in Figure 3.1. As $\lambda \to \infty$, this correlation tends to -1; however, for small λ, there are significant restrictions on what correlations can be made. An interesting feature of Figure 3.1 lies with the slight non-monotonicity of $\mathrm{NB}(\lambda)$ in λ for some $\lambda \leq 3$. This is not computational roundoff; indeed, the Hermite coefficients g_k below, discussed in Subsection 2.4, are not monotonic in λ. This fact can be inferred from the plots in the supplementary material in Jia et al. (2023).

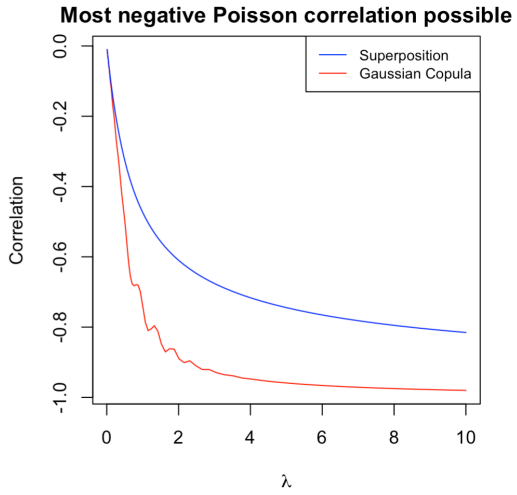

Figure 3.1: Lower Curve: Plots of the most negative achievable autocorrelation, NB(λ), for various λ. Bottom Curve: Plot of the most negative autocorrelations possible with the superpositioned model class of Section 3.2.3.

3.2.1 Discrete Autoregressions

Discrete autoregeressions (DARs), the original attempt to devise stationary series having a particular marginal distribution (Jacobs and Lewis, 1978a,b,c), work by mixing past series values. In the Poisson case, the construction begins with the sequence $\{A_t\}_{t=1}^{\infty}$ of IID Poisson variables with mean $\lambda > 0$. A sequence $\{B_t\}_{t=1}^{\infty}$ of IID Bernoulli trials is needed that is independent of $\{A_t\}$ and has success probability $P(B_t = 1) = p$.

In the first-order case, the construction starts by taking $X_1 = A_1$. For $t \geq 2$, series values are mixed via

$$X_t = B_t X_{t-1} + (1 - B_t) A_t.$$

Here, if $B_t = 1$, X_t is taken as X_{t-1}; should $B_t = 0$, $X_t = A_t$ is a "new independent

Poisson draw". Schemes extending the paradigm to higher autoregressive (AR) orders r are achievable with additional Bernoulli sequences, but the DAR class has drawbacks that make it unsuitable for use. Foremost, DAR models cannot have any negative autocorrelations. In the first order case, one can show that $\text{Corr}(X_t, X_{t+h}) = p^h$ for $h \geq 0$, which cannot be negative since $p \in (0, 1)$ must be a probability. Perhaps worse, series values are often repeated: $P[X_{t+1} = X_t] \geq p$. In the heavily correlated case where p is close to unity, the series becomes almost constant. Because of these properties, DAR series were essentially abandoned. See Möller and Weiß (2020) for recent attempts to remedy these issues. We will not consider this class further.

3.2.2 Integer Autoregressions

Steutel and van Harn (1979) introduced binomial thinning in an attempt to mimic AR recursions to construct count series. If X is a count-valued variable, define $\alpha \circ X = \sum_{i=1}^{N} B_i$, where $\{B_i\}$ are IID Bernoulli trials with success probability $\alpha \in (0, 1)$ that are independent of X; $\circ$ is called a binomial thinning operator.

Integer autoregressions (INARs) are based on thinning operators. In the first-order case, a strictly stationary series with Poisson marginal distributions with mean λ is governed by the difference equation

$$X_t = \alpha \circ X_{t-1} + \epsilon_t, \tag{3.3}$$

where $\{\epsilon_t\}$ is IID with a Poisson marginal distribution with mean $\lambda(1 - \alpha)$ (McKenzie,

1985; Alzaid and Al-Osh, 1990; Weiß, 2018).

A well-known property of solutions to (3.3) is that any discrete self-decomposable marginal distribution (these include Poisson, negative binomial, and generalized Poisson) can be produced by this recursion; this said, our focus remains on Poisson marginals. The autocorrelation function of an INAR(1) series can be shown to have the form

$$\text{Corr}\left(X_t, X_{t+h}\right) = \alpha^h; \tag{3.4}$$

in particular, negative autocorrelations cannot be produced since $\alpha \in (0, 1)$.

Higher order schemes, dubbed INAR(r) for order r, have been investigated; however, producing series with Poisson marginals in these schemes has been problematic. Be wary of issues with the literature here; specifically, the methods in Alzaid and Al-Osh (1990) and Du and Li (1991) will not achieve Poisson marginals; see the discussion in Scotto et al. (2015). To circumvent this problem, Zhu and Joe (2003) propose CINAR models. A CINAR(r) series of order r follows the recursion

$$X_t = D_{t,1}\left(\alpha \circ X_{t-1}\right) + \cdots + D_{t,r}\left(\alpha \circ X_{t-r}\right) + \epsilon_t. \tag{3.5}$$

The IID time t "decision vector" is $\boldsymbol{D}_t = (D_{t,1}, \ldots, D_{t,r}) \sim \text{Mult}(1; \phi_1, \ldots, \phi_r)$ and is independent of $\{\epsilon_t\}$ and $\{X_s\}_{s<t}$. The decision vector $\boldsymbol{D}_t$ chooses which of the past r series values is used to thin, enabling the scheme to keep a Poisson marginal distribution. Here, the innovation ϵ_t and the thinning of X_{t-j} for the chosen $j \in \{1, \ldots, r\}$ are always

conducted independently.

Zhu and Joe (2003) show that the marginal distribution of any CINAR(r) series must also be self-decomposable. As with the Poisson INAR(1) model in (3.3), the CINAR(r) model has a marginal Poisson distribution with mean λ when $\{\epsilon_t\}$ is IID Poisson with mean $\lambda(1-\alpha)$, regardless of the order of r. Weiß (2008) derives the autocovariance of a CINAR(r) series from (3.5); from the resulting expression, one can show that this covariance must be non-negative. As such, CINAR(r) models cannot have any negative autocorrelations. Again, this model class fails to span the range of all possible autocovariances.

3.2.3 Superposition Techniques

Poisson distributions can be built by adding IID copies of Bernoulli random variables. Indeed, if $\{B_i\}_{i=1}^{\infty}$ are IID Bernoulli variables with success probability $p = P[B_t = 1]$ and N is Poisson, independent of $\{B_i\}_{i=1}^{\infty}$ and with mean λ, then $\sum_{i=1}^{N} B_i$ has a Poisson distribution with mean $p\lambda$. Blight (1989) and Cui and Lund (2009) use this construction to produce correlated count series having Poisson marginal distributions.

Elaborating, suppose that $\{B_{t,i}\}_{t=1}^{\infty}$ are IID copies of the autocorrelated Bernoulli trial sequence $\{B_t\}$. Clarifying, for each fixed i, $\{B_{t,i}\}_{t=1}^{\infty}$ is autocorrelated in time t — say $\mathrm{Cov}(B_{t,i}, B_{t+h,i}) := \gamma_B(h)$ — but $\{B_{t,i}\}_{t=1}^{\infty}$ and $\{B_{t,j}\}_{t=1}^{\infty}$ are independent when $i \neq j$. A series with Poisson marginals can be built via the superpositioned form

$$X_t = \sum_{i=1}^{N_t} B_{t,i}.$$

where $\{N_t\}$ is IID Poisson(λ). Then X_t has a Poisson distribution with mean λp. The autocovariances of $\{X_t\}$ are

$$\gamma_X(h) := \text{Cov}(X_t, X_{t+h}) = E[\min(N_1, N_2)]\gamma_B(h) \tag{3.6}$$

for $h > 0$, where N_1 and N_2 are **independent** Poisson variables with mean λ. Note that $\gamma_X(h)$ will be negative whenever $\gamma_B(h)$ is negative; hence, this model class can produce negatively correlated series. One can show that

$$E[\min(N_1, N_2)] := \kappa(\lambda) = 2\lambda \left[1 - e^{-4\lambda}\{I_0(4\lambda) + I_1(4\lambda)\}\right], \tag{3.7}$$

where the I_js are modified Bessel functions:

$$I_j(x) = \sum_{n=0}^{\infty} \frac{(x/2)^{2n+j}}{n!(n+j)!}, \quad j = 0, 1$$

(Jia et al., 2021). The above construct essentially builds the correlated Poisson series $\{X_t\}$ from the independent Poisson series $\{N_t\}$.

Several ways to construct a correlated sequence of Bernoulli trials exist. One way uses a stationary renewal sequence built from the IID lifetimes $\{L_i\}_{i=1}^{\infty}$ supported on $\{1, 2, \ldots\}$ and an initial delay L_0 supported on $\{0, 1, 2, \ldots\}$ as follows. Define the random walk $S_n = L_0 + L_1 + \cdots + L_n$ for $n \geq 0$ and set $B_t = 1$ when a renewal occurs at time t (i.e., when $S_n = t$ for some $n \geq 0$) and zero otherwise. When the initial delay L_0 is chosen as the first derived distribution of one of the L_is for $i \geq 1$ (a generic copy

of these is denoted by L), viz.

$$P[L_0 = k] = \frac{P[L > k]}{\mu_L}, \quad k = 0, 1, 2, \ldots,$$

the Bernoulli sequence is stationary in that $E[B_t] \equiv \frac{1}{\mu_L}$ and

$$\gamma_B(h) = \frac{1}{\mu_L}\left(u_h - \frac{1}{\mu_L}\right). \tag{3.8}$$

Here, the notation uses $\mu_L = E[L]$ and u_h as the probability of a time h renewal in a non-delayed renewal process ($L_0 = 0$). When L is aperiodic and $E[L] < \infty$ (which we henceforth assume), $u_h \longrightarrow \mu_L^{-1}$ as $t \longrightarrow \infty$ by the elementary renewal theorem (Smith, 1958). One can show that $\{X_t\}$ has long memory (absolutely non-summable autocovariances over all lags) when $E[L^2] = \infty$ (Lund et al., 2016); Jia et al. (2021) derive further properties of superpositioned series and investigate non-Poisson count marginal distributions.

Another way to produce a stationary but correlated Bernoulli sequence clips a Gaussian sequence as in Kedem (1980). Elaborating, let $\{Z_t\}$ be a standard stationary Gaussian sequence with $E[Z_t] \equiv 0$, $\mathrm{Var}(Z_t) \equiv 1$, and $\mathrm{Corr}(Z_t, Z_{t+h}) = \rho_Z(h)$. Define Bernoulli trials as

$$B_t = 1_A(Z_t),$$

where A is some fixed set. In this case, autocovariances are

$$\gamma_B(h) = P(Z_t \in A \cap Z_{t+h} \in A) - P(Z_t \in A)^2. \tag{3.9}$$

As an example, when $A = (0, \infty)$, $P(Z_t \in A) = 1/2$ and classic bivariate normal orthant probability calculations give

$$\gamma_B(h) = \frac{\arcsin(\rho_Z(h))}{2\pi}.$$

Notice that the autocovariances in (3.8) and (3.9) can be negative. Specifically, for a renewal $\{B_t\}$, $\gamma_B(h) < 0$ whenever $u_h < \mu_L^{-1}$; for a Gaussian clipped $\{B_t\}$ with $A = (0, \infty)$, $\gamma_B(h) < 0$ whenever $\rho_Z(h) < 0$.

While the autocovariance function in (3.6) can be decisively negative, it does not achieve the minimum possible in (3.2). To see this, note from (3.6) and (3.7) that

$$\rho_X(h) := \text{Corr}(X_t, X_{t+h}) = \frac{\kappa(\lambda/p)}{\lambda}\gamma_B(h).$$

The smallest $\gamma_B(h)$ that can be made from a binary pair of random variables can be shown to be -p^2 for any lag h. Thus, the most negative lag h correlation that can be built from this model for any h is $-p^2\kappa(\lambda/p)/\lambda$.

Figure 3.1 displays this most negative correlation. Again, the correlation approaches -1 as λ increases and there are significant restrictions for λ close to zero. To find these values, a numerical search was used to find the $p \in (0, 1)$ that minimizes

$-p^2 J(\lambda/p)/\lambda$ for each fixed λ. These most negative correlations are uniformly worse (bigger) than the optimal ones identified earlier. This brings us to our best Poisson model, which will be able to achieve the full spectrum of achievable autocorrelations.

3.2.4 Gaussian Copulas

A recent class of very parsimonious and general count models developed in Jia et al. (2023) has been demonstrated to have remarkable flexibility. This model starts with a stationary standard Gaussian sequence $\{Z_t\}$ and transforms it to the desired count. Standardized means that $E[Z_t] \equiv 0$, $\mathrm{Var}(Z_t) \equiv 1$, and $\gamma_Z(h) = \rho_Z(h) = \mathrm{Corr}(Z_t, Z_{t+h})$. The construction transforms Z_t at time t via

$$X_t = F_\lambda^{-1}(\Phi(Z_t)). \tag{3.10}$$

Here, $\Phi(\cdot)$ is the standard normal CDF. By the probability integral transformation theorem, $\Phi(Z)$ has a uniform distribution over [0,1] and X_t has a Poisson marginal distribution with mean λ.

The form of the autocovariance function of $\{X_t\}$ is unwieldy, but can be quantified through several expansions. With $G(x) = F_\lambda^{-1}(\Phi(x))$, arguing as in (3.1) gives

$$G(z) = \sum_{n=1}^{\infty} n 1_{[F_\lambda(n-1) \le \Phi(z) < F_\lambda(n)]} = \sum_{\ell=0}^{\infty} 1_{[F_\lambda(\ell-1),1)}(\Phi(z)) = \sum_{\ell=0}^{\infty} 1_{[\Phi^{-1}(F_\lambda(\ell)),\infty)}(z).$$

One expression for the autocovariances now follows as

$$\gamma_X(h) = \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} P\left(Z_t \in [\Phi^{-1}(F_\lambda(j)), \infty) \cap Z_{t+h} \in [\Phi^{-1}(F_\lambda(k)), \infty)\right),$$

which is computationally intensive to evaluate.

A more tractable expansion works through the Hermite polynomials $\{H_k(x)\}_{k=0}^{\infty}$ defined by

$$H_k(x) = (-1)^k e^{z^2/2} \frac{d^k}{dz^k}\left(e^{-z^2/x}\right).$$

The first three Hermite polynomials are $H_0(x) \equiv 1$, $H_1(x) = x$, and $H_2(x) = x^2 - 1$. Higher order polynomials obey the recursion

$$H_k(x) = xH_{k-1}(x) - H'_{k-1}(x), \quad k \geq 1.$$

Expanding G in a Hermite basis, viz.

$$G(x) = \sum_{\ell=0}^{\infty} g_\ell H_k(x)$$

produces the key relationship

$$\gamma_X(h) = L(\rho_Z(h)), \quad |u| \leq 1,$$

where $L(\cdot)$, called a link function in Jia et al. (2023), has the power series expansion

$$L(u) = \sum_{k=1}^{\infty} \frac{k! g_k^2}{\gamma_X(0)} u^k = \sum_{k=1}^{\infty} \ell_k u^k,$$

where $\ell_k = k! g_k^2 / \gamma_X(0)$. It is known that $L(u)$ is differentiable in u, that $L(0) = 0$, and that $L(1) = 1$ (Jia et al., 2023). The quantity ℓ_k is called a link coefficient. Figure 3.2 plots ℓ_k for a few values of k as a function of λ. While these coefficients behave erratically, they decrease quickly as k and/or λ increase. The value $\lim_{u \downarrow -1} L(u)$ is the most negative pairwise correlation that can be made. The inequality $|\gamma_Z(h)| \leq |\gamma_X(h)|$, established in (Jia et al., 2023) (see also Tong (2014)), shows that correlation will always be lost in the transformation from Z_t to X_t, but many times, this loss is not substantial.

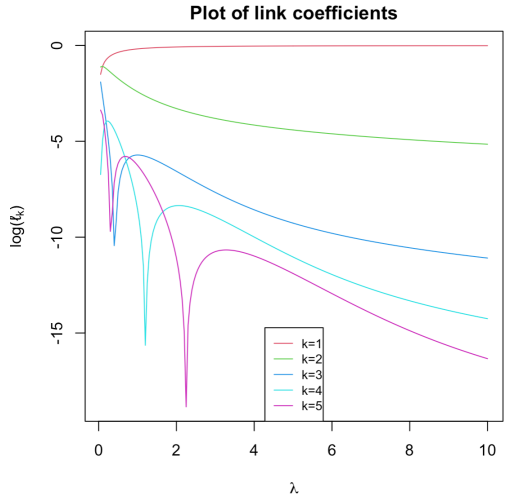

Figure 3.2: Plot of $\log(\ell_k)$ versus λ for $k \in \{1, 2, 3, 4, 5\}$.

With the conventions $\Phi^{-1}(0) = -\infty$ and $\Phi^{-1}(1) = \infty$, the Hermite coefficients

can be computed as

$$g_k = \frac{1}{k!}\left[1 - e^{-\lambda} + \sum_{\ell=1}^{\infty} H_{\ell-1}(\Phi^{-1}(F_\lambda(\ell)))\phi(F_\lambda(\ell))\right].$$

Other forms for g_k are derived in Jia et al. (2023). Figure 3.3 plots $L(u)$ against u for various values of λ. Notice that when $\lambda = 10$, $L(u) \approx u$ and very little autocorrelation is lost in the transformation of Z_t to X_t.

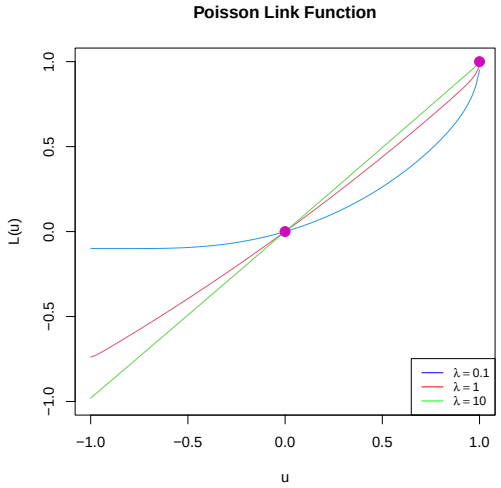

Figure 3.3: The link function $L(u)$ for $\lambda \in \{0.1, 1, 10\}$.

3.3 Inference

In the stationary case where $X_t \sim \text{Poisson}(\lambda)$, the traditional estimate of λ is the sample mean

$$\hat{\lambda} = \frac{X_1 + \cdots + X_n}{n}.$$

The variance of this estimate is

$$\text{Var}(\hat{\lambda}) = \frac{\lambda}{n}\left[1 + 2\sum_{j=1}^{n-1}(1 - j/n)\gamma_X(j)\right],$$

which is consistent whenever $\{X_t\}$ has short memory autocorrelations ($\sum_{j=1}^{\infty}|\rho_X(j)| < \infty$). One can get estimators of λ with smaller variances than the sample mean with generalized least squares methods, but any improvements are negligible as $n \to \infty$ (see the discussion in Chipman (1979); Lee and Lund (2004)).

In practical modeling scenarios, $\{X_t\}$ is usually non-stationary, possibly due to trends, periodicities, covariates, etc. To study estimation for Poisson count series with such structures, we need to develop time-varying versions of our techniques. Hence, our immediate goal is to develop time-varying models where $X_t \sim \text{Poisson}(\lambda_t)$ and there is correlation between observations.

To develop models where $X_t \sim \text{Poisson}(\lambda_t)$, we first revisit INAR(1) models. Here, complications arise. To see this, if $X_1 \sim \text{Poisson}(\lambda_1)$ and the process obeys (3.3), then we must have $\epsilon_2 \sim \text{Poisson}(\lambda_2 - p\lambda_1)$. Unfortunately, there is no guarantee that $\lambda_2 - p\lambda_1$ is non-negative, suggesting that the INAR(1) paradigm might not be a good way to handle time-varying dynamics. Should it be known that λ_t is nondecreasing in t, then one could explore this model class further; see Bentarzi and Souakri (2023) for additional comments on process existence. Because of this issue, we move to other methods.

3.3.1 Time-varying Superpositioned Series

In the superpositioned model class, time-varying models having the desired marginal properties are easy to construct. For this, let $\{N_t\}$ be a sequence of independent Poisson variables with $N_t \sim \text{Poisson}(\lambda_t/p)$. Here, $p = P[B_{t,i} = 1]$ is the success probability of the Bernoulli trials in the construction. Then it is easy to see that $X_t \sim \text{Poisson}(\lambda_t)$ as required. In this case, the derivation associated with 3.6 gives

$$\text{Cov}(X_t, X_{t+h}) = \mathbb{E}(\min(N_t, N_{t+h}))\gamma_B(h),$$

where N_t and N_{t+h} are independent Poisson variables with parameters λ_t and λ_{t+h}, respectively. Note that

$$\begin{aligned}
\mathbb{E}(\min(N_t, N_{t+h})) &= \sum_{n=1}^{\infty} P\left(\min(N_t, N_{t+h}) \geq n\right) \\
&= \sum_{n=1}^{\infty} P\left(N_t \geq n \cap N_{t+h} \geq n\right) \\
&= \sum_{n=1}^{\infty} P\left(N_t \geq n\right) P\left(N_{t+h} \geq n\right) \\
&= \sum_{n=1}^{\infty} [1 - F_{\lambda_t/p}(n-1)][1 - F_{\lambda_{t+h}/p}(n-1)],
\end{aligned}$$

where $F_{\lambda_t}(n-1)$ and $F_{\lambda_{t+h}}(n-1)$ are the Poisson CDFs at the times t and $t+h$. There does not seem to be a simplification of this formula as in (3.7) unless $\lambda_t = \lambda_{t+h}$.

To estimate parameters in superpositioned schemes, we will use linear prediction methods. Unfortunately, the model's likelihood function and conditinal expecta-

tions $E[X_t|X_1,\ldots,X_{t-1}]$ appear to be intractable. Also, how to simulate the likelihood accurately, as we will do for the Gaussian copula case below with particle filtering methods, is also unclear. A bivariate composite likelihood is indeed tractable as an alternative to linear prediction; however, we will see that linear prediction works quite well.

Linear prediction works by first calculating $\mathrm{Cov}(X_t, X_s) := \Gamma_X(t,s)$ for each $1 \le t, s \le n$. Estimators are found by minimizing the simple sum of squares

$$\sum_{t=1}^{n}(X_t - \hat{X}_t)^2, \tag{3.11}$$

where

$$\hat{X}_t = \lambda_t + \sum_{j=1}^{t-1} w_{j,t}(X_j - \lambda_j)$$

is the best one-step-ahead predictor of X_t made from linear combinations of a constant and $X_1,\ldots,X_{t-1}$. The coefficients $\{w_{j,t}\}$ satisfy the prediction equations

$$\begin{bmatrix} \Gamma_X(1,1) & \Gamma_X(1,2) & \cdots & \Gamma_X(1,t-2) & \Gamma_X(1,t-1) \\ \Gamma_X(2,1) & \Gamma_X(2,2) & \cdots & \Gamma_X(2,t-2) & \Gamma_X(2,t-1) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \Gamma_X(t-2,1) & \Gamma_X(t-2,2) & \cdots & \Gamma_X(t-2,t-2) & \Gamma_X(t-2,t-1) \\ \Gamma_X(t-1,1) & \Gamma_X(t-1,2) & \cdots & \Gamma_X(t-1,t-2) & \Gamma_X(t-1,t-1) \end{bmatrix} \begin{bmatrix} w_{1,t} \\ w_{2,t} \\ \vdots \\ w_{t-2,t} \\ w_{t-1,t} \end{bmatrix} = \begin{bmatrix} \Gamma_X(1,t) \\ \Gamma_X(2,t) \\ \vdots \\ \Gamma_X(n-2,t) \\ \Gamma_X(n-1,t) \end{bmatrix}. \tag{3.12}$$

To obtain parameter estimators, the sum of squares is numerically minimized in the parameters appearing in $\{\lambda_t\}$. In this scheme, we only minimize the sum of

squares in (3.11) about the mean parameters appearing in λ_t; the parameters appearing in the covariance structure of $\{B_t\}$ are held to their true values during this optimization. Future work will consider how to estimate these parameters in tandem; here, a Cochrane-Orcutt recursion seems developable (Cochrane and Orcutt, 1949). One complication is that some parameters in $\{B_t\}$ (namely p) arise in both the mean and autocovariance structure of the linear predictors. One may also wish to consider weighted least squares to accommodate the changing variances of the series.

Solving the linear system in (3.12) requires a $\mathcal{O}(t^3)$ computational cost; as such, the computational burden can be expensive for large n. We recommend doing a Cholesky decomposition of the covariance matrix on the left hand side of (3.12) and then using backwards/forward substitution to obtain $\{w_{t,k}\}$. The classic Durbin-Levinson recursion is not suitable here since $\{X_t\}$ is not stationary. In our future computations, the "Nelder–Mead" optimization method was used to minimize the sum of squares in (3.11).

3.3.2 Time-varying Gaussian Copula Series

In the Gaussian copula case, the process construction carries through as before; specifically, we set

$$X_t = F_{\lambda_t}^{-1}(\Phi(Z_t)). \tag{3.13}$$

The Hermite expansion of the time homogeneous case is simply allowed to vary with time now.

For notation, let $\boldsymbol{\theta}$ contain all parameters appearing in $\{\lambda_t\}_{t=1}^n$ and $\boldsymbol{\eta}$ denote all parameters governing $\{Z_t\}$. We do not suggest trying to incorporate time dependence into the dynamics of $\{Z_t\}$ as process existence issues then arise. The covariance matrix of $(Z_1, \ldots, Z_n)'$ depends only on $\boldsymbol{\eta}$ (and not on $\boldsymbol{\theta}$).

The model's likelihood function, denoted by $\boldsymbol{\mathcal{L}}(\boldsymbol{\theta}, \boldsymbol{\eta})$, is simply a high dimensional multivariate normal probability. To see this, use (3.13) with the data $X_1, \ldots, X_n$ to get

$$\boldsymbol{\mathcal{L}}(\boldsymbol{\theta}, \boldsymbol{\eta}) = \mathbb{P}(X_1 = x_1, \cdots, X_n = x_n) = \mathbb{P}\left(Z_1 \in (a_1, b_1], \cdots, Z_n \in (a_n, b_n]\right), \qquad (3.14)$$

where $\{a_t\}_{t=1}^n$ and $\{b_t\}_{t=1}^n$ are

$$a_t = \Phi^{-1}(F_{\lambda_t}(x_t - 1)), \quad b_t = \Phi^{-1}(F_{\lambda_t}(x_t)).$$

While this probability is not intractable, it is infeasible to accurately evaluate for large n. A likelihood can, however, be quite accurately simulated by particle filtering methods (Douc et al., 2014). Particle filtering simulation methods can be used to reliably approximate the model's likelihood and even compute standard errors. The current preferred methods of multivariate normal probability evaluation are argiably the Geweke–Hajivassiliou–Keane (GHK) simulators of Geweke (1991) and Hajivassiliou et al. (1996). Here, we develop an adaptive version of this simulator.

Particle filtering methods, which are classic importance sampling techniques,

aim to evaluate integrals by drawing samples from an alternative distribution and averaging their corresponding weights. Should we need to estimate the integral $\int f(\boldsymbol{x})d\boldsymbol{x}$ over some domain D, then we use

$$\int_D f(\boldsymbol{x})d\boldsymbol{x} = \int_D \frac{f(\boldsymbol{x})}{q(\boldsymbol{x})} q(\boldsymbol{x})d\boldsymbol{x},$$

where $f(\boldsymbol{x})/q(\boldsymbol{x})$ is the weight and q is called the importance distribution. The importance sampling estimate of the integral is

$$\int_D \frac{f(\boldsymbol{x})}{q(\boldsymbol{x})} q(\boldsymbol{x})d\boldsymbol{x} \approx \frac{1}{m}\sum_{k=1}^{m} \frac{f(\boldsymbol{x}^{(k)})}{q(\boldsymbol{x}^{(k)})},$$

where $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(m)}$ are m IID samples drawn from q. We require that q satisfies $q(z_{1:n}) > 0$ for $z_t \in (a_t, b_t]$ and $q(z_{1:n}) = 0$ otherwise; our notation uses $z_{1:k} = (z_1, \ldots, z_k)$ and $x_{1:k} = (x_1, \ldots, x_k)$.

We take advantage of the Markov chain properties of the latent AR $\{Z_t\}$. The GHK algorithm samples Z_t, depending on the its previous history $Z_{t-1}, \ldots, Z_1$ and X_t, from a truncated normal density. Specifically, let $p_{\boldsymbol{\eta}(t)}\left(z_t | z_{1:t-1}; x_t\right)$ denote the truncated normal density of Z_t given the history $Z_1, \ldots, Z_{t-1}$ and X_t. Then

$$p_{\boldsymbol{\eta}(t)}\left(z_t | z_{t-1}, \ldots, z_1, x_t\right) = \frac{1}{r_t}\left[\frac{\phi(\frac{z_t - \hat{z}_t}{r_t})}{\Phi(\frac{b_t - \hat{z}_t}{r_t}) - \Phi(\frac{a_t - \hat{z}_t}{r_t})}\right], \quad a_t < z_t < b_t, \tag{3.15}$$

where $\hat{z}_t$ and r_t are the one-step-ahead mean and standard deviation of Z_t conditioned on $Z_{1:t-1}$. Note that a_t and b_t only depend on x_t. Here, we choose the importance

sampling distribution

$$q_{\boldsymbol{\eta}}(z_{1:n}|x_{1:n}) = p_{\boldsymbol{\eta}(1)}(z_1|x_1) \prod_{t=2}^{n} p_{\boldsymbol{\eta}(t)}\left(z_t|z_{1:t-1}; x_t\right). \tag{3.16}$$

After some cancellation, we arrive at

$$\frac{\boldsymbol{\phi}_{\boldsymbol{\eta}}\left(z_{1:n}\right)}{q(z_{1:n})} = \left[\Phi\left(b_1\right) - \Phi\left(a_1\right)\right] \prod_{t=2}^{n} \left[\Phi\left(\frac{b_t - \hat{z}_t}{r_t}\right) - \Phi\left(\frac{a_t - \hat{z}_t}{r_t}\right)\right].$$

Here, $\phi_{\boldsymbol{\theta}}(z_{1:n})$ denotes the multivariate normal distribution with mean zero and covariance matrix that of $Z_{1:n}$. See Kong and Lund (2023) for derivation details.

Define the initial weight $w_1 = \Phi(b_1) - \Phi(a_1)$. The weights are recursively updated via

$$w_t = w_{t-1}\left[\Phi\left(\frac{b_t - \hat{z}_t}{r_t}\right) - \Phi\left(\frac{a_t - \hat{z}_t}{r_t}\right)\right]$$

at time t during the sequential sampling procedure. At the end of the sampling, we obtain

$$w_n = \frac{\boldsymbol{\phi}_{\boldsymbol{\eta}}(z_{1:n})}{q_{\boldsymbol{\eta}}(z_{1:n}|x_{1:n})}.$$

In the classic GHK simulator, $\hat{Z}_t$ and r_t are obtained from the covariance matrix of $\{Z_t\}$. When $\{Z_t\}$ is the causal autoregression of order r, viz.,

$$Z_t = \phi_1 Z_{t-1} + \cdots + \phi_p Z_{t-r} + \epsilon_t$$

where $\{\epsilon_t\}$ is Gaussian white noise with variance σ_ϵ^2 that makes $\mathrm{Var}(Z_t) = 1$, the one-

step-ahead predictors and their mean squared errors obey

$$\hat{Z}_t = \phi_1 Z_{t-1} + \ldots \phi_r Z_{t-r}, \quad t > r$$

and $r_t = \sigma_\epsilon$ for $t > r$. See Brockwell and Davis (1991a) for computing these quantities when $t \leq r$.

The above procedure generates a fair draw of a single "particle path" $\{Z_t\}$ with the property that $\{X_t\}_{t=1}^n$ generated from $\{Z_t\}_{t=1}^n$ yields the observations $x_1, \ldots, x_n$. Repeating this process m independent times gives m simulated process trajectories. Let $\{\mathbf{Z}^{(1)}, \ldots, \mathbf{Z}^{(m)}\}$ denote these trajectories and denote their corresponding weights at time n by $\{w_n^{(k)}\}_{k=1}^m$.

The importance sampling estimate is given by

$$\hat{\mathcal{L}}(\boldsymbol{\theta}, \boldsymbol{\eta}) = \frac{1}{m} \sum_{k=1}^{m} w_n^{(k)}.$$

A large m of course provides more accurate estimation. The popular "BGSF" gradient step and search method is used to optimize the estimated likelihood $\hat{\mathcal{L}}(\boldsymbol{\theta}, \boldsymbol{\eta})$; other optimizers may also work.

Common random numbers (CRNs), techniques that use the same random quantities across differing parameter values in particle filtering, are used to produce a "smooth" estimated likelihood function. With CRNs, Hessian-based standard errors derived from the likelihood function's derivatives at the likelihood estimate are much

more reliable; see Kleinman et al. (1999) and Glasserman and Yao (1992) for more on CRNs.

3.3.2.1 A Simulation Study

This section studies parameter estimators of the superpositioned and Gaussian copula Poisson count series through simulation. To illustrate the techniques in a simple setting, our simulations consider a single trend and covariate parameter:

$$\lambda_t = \exp\{\mu + \beta_1 t + \beta_2 C_t\}.$$

More complicated scenarios are dealt with similarly. Here, C_t is the value of the covariate at time t, generated as zero-one IID Bernoulli(0.3) draws under the R seed "1234". The covariate sequence $\{C_t\}_{t=1}^{n}$ is fixed throughout all simulations. A log link has been used to keep the Poisson parameter non-negative, with e^{μ} being the baseline value of λ. The quantity β_1 is the "trend" parameter and β_2 measures the contribution of the covariate to the mean. In general, we do not look to conduct inferences about the location parameter μ. In practice, λ_t can be any non-negative function, making the model flexible.

In our simulations below, we set the parameters to $\mu = 1$, $\beta_1 = 0.01$, and $\beta_2 = 1$ and consider series lengths of 50, 100, and 300. Five hundred independent simulation replicates are studied in every simulation scenario.

For the superposition scheme, the $\{B_t\}$ process used to generate our series is

obtained from a clipped AR(1) $\{Z_t\}$ series. The AR(1) parameters are set to $\phi = 1/2$ and $\sigma^2 = 3/4$ so that a unit variance Gaussian series is clipped. Here p is forced to $1/2$ by setting $B_t = 1_{(0,\infty)}(Z_t)$. During estimation, the autocovariance parameters are fixed to their true values in the linear prediction scheme and we examine estimates of the three parameters appearing in the mean λ_t. The true model parameters were used as initial guesses in our optimizations. Similar results were obtained when other initial values were used.

Figure 3.4 displays parameter estimator boxplots for each mean parameter. The dotted red line demarcates the true parameter value. All boxplots are centered around their true parameter values and the distributional shape seems approximately normal. For standard errors of these estimators, Table 3.1 reports the sample standard deviations of the parameter estimators over the 500 runs. As expected, estimation accuracy increases as the series length increases. Overall, the estimators seem accurate.

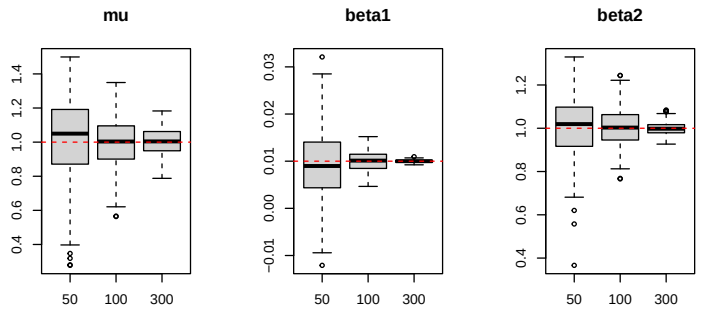

Figure 3.4: Boxplots of parameter estimators for a superpositioned Poisson count series with $\{B_t\}$ constructed by clipping an AR(1) $\{Z_t\}$. Both ϕ and p are set to $1/2$, their true values, during estimation. The other three mean parameter estimators appear approximately unbiased; dashed lines demarcate true parameter values.

Superposition Poisson AR(1) Model				
n		$\hat{\mu}$	$\hat{\beta}_1$	$\hat{\beta}_2$
50	mean	1.00805	0.00914	1.00887
	SD	0.24117	0.00733	0.14888
100	mean	0.99380	0.01003	1.00240
	SD	0.14921	0.00210	0.08668
300	mean	1.00271	0.00999	1.00012
	SD	0.08064	0.00031	0.03107

Table 3.1: Mean and standard deviation (SD) of estimators for the superpositioned Poisson count series with $\{B_t\}$ constructed via a clipped AR(1). True values of the parameters are $\mu = 1$, $\beta_1 = 0.01$, and $\beta_2 = 1$. The results report the sample mean and standard deviation (denominator of 499) of the parameter estimates from the 500 runs.

Moving to the Gaussian copula scheme, we again use AR(1) errors with $\phi = 1/2$ and $\sigma^2 = 3/4$ to generate the latent process $\{Z_t\}$. The settings for μ, β_1, and β_2 used above are repeated. In this scheme, all parameters are estimated via particle filtering methods, even the AR(1) parameter ϕ. The true model parameters were used as initial guesses in our optimizations. Similar results were obtained when other initial values were used.

Figure 3.5 shows boxplots of all estimators and series lengths. The dotted red line again indicates true parameter values. All boxplots are centered around the true parameter values and look approximately normal, with perhaps one exception being $\hat{\phi}$ under the shortest series length $n = 50$. For standard errors, Table 3.2 reports two values: 1) the sample standard deviation of the parameter estimators over the 500 runs (denominator of 499), and 2) the average (over all runs) of standard errors obtained by inverting the Hessian matrix at the maximum likelihood estimate for each run (denominator of 500). The difference between these two values are quite small,

implying that Hessian-based standard errors obtained from one sample path are indeed accurate. Standard errors again decrease with increasing n. Again, the performance appears good.

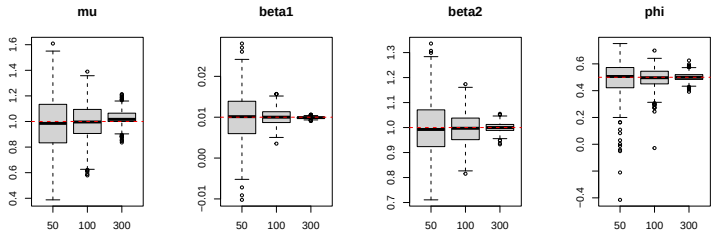

Figure 3.5: Boxplots of parameter estimators for Gaussian copula estimates of a Poisson count series with an AR(1) $\{Z_t\}$ with $\phi = 0.5$. All estimators appear approximately unbiased; dashed lines demarcate true parameter values.

Gaussian Copula Poisson AR(1) Model					
n		$\hat{\mu}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\phi}$
50	mean	0.98419	0.01012	0.99880	0.48945
	SD	0.22609	0.00610	0.11068	0.13010
	$\hat{E}(I'(\theta)^2)$	0.22640	0.00644	0.10701	0.11412
100	mean	0.99550	0.01004	0.99716	0.49222
	SD	0.13941	0.00200	0.06520	0.07693
	$\hat{E}(I'(\theta)^2)$	0.14720	0.00206	0.06642	0.07498
300	mean	1.02786	0.00988	1.00020	0.50182
	SD	0.05987	0.00025	0.01896	0.03320
	$\hat{E}(I'(\theta)^2)$	0.06410	0.00027	0.02219	0.04144

Table 3.2: Standard errors for the parameter estimators for the Poisson marginal distribution with an AR(1) $\{Z_t\}$. The results show the sample standard deviation (SD) of the parameter estimators from 500 independent series (denminator of 499), and the average of the 500 standard errors obtained by inverting the Hessian matrix ($\hat{E}[I'(\theta)^2)]$ at the maximum likelihood estimate over these same runs.

In comparing superpositioned and Gaussian copula results, we see that standard errors for the Gaussian likelihood estimators are slightly smaller than their superpositioned counterparts. This is expected: likelihood estimators are generally the

asymptotically most efficient estimators. This said, the calculations needed to produce the likelihood estimators are more intensive than those for linear prediction. Finally, we did study higher order autoregressions; results are again impressive and similar to the above. For brevity's sake, figures and tables of these simulations are not presented here.

3.4 Applications

This section considers two count series that we fit with Poisson marginal distributions: Atlantic Basin tropical storm counts and the number of no-hitters games pitched annually in Major League Baseball. Both series are comprised of small counts, where the marginal distribution becomes important. Because the superpositioned linear prediction estimation performs slightly worse than Gaussian copula likelihood estimation, we concentrate on the latter estimation technique in this section.

3.4.1 Atlantic Tropical Cyclones

Our first series contains the annual number of tropical cyclones observed in the North Atlantic Basin since 1970. This series is plotted in the top plot of Figure 3.6. Poisson marginal distributions have been previously advocated for these and other tropical cyclone counts (Robbins et al., 2011; Solow and Beet, 2008; Mooley, 1980).

There is concern that the number of North Atlantic Basin cyclones has been increasing in recent years, with researchers pointing to 1995 as a year where the North Atlantic warmed and tropical storm activity increased; see the changepoint analyses

in Robbins et al. (2011) and Fisher et al. (2020). Because of this, we will allow for a linear trend as one covariate (a changepoint mean shift structure is also worthy of consideration). A strong El-Nino, which is a measure of equitorial warming in the Pacific Ocean, is thought to impede Atlantic tropical cyclone development (Gray, 1984; Goldenberg and Shapiro, 1996) through its influence on the southern jet stream: a strong El-Nino produces a strong southern jet stream, which produces wind shear at stratospheric levels, shearing the tops off of developing thunderstorm clouds and hindering tropical cyclone development. As a second covariate, annual values of El-Nino 3, which are shown in the bottom plot of Figure 3.6, are used.

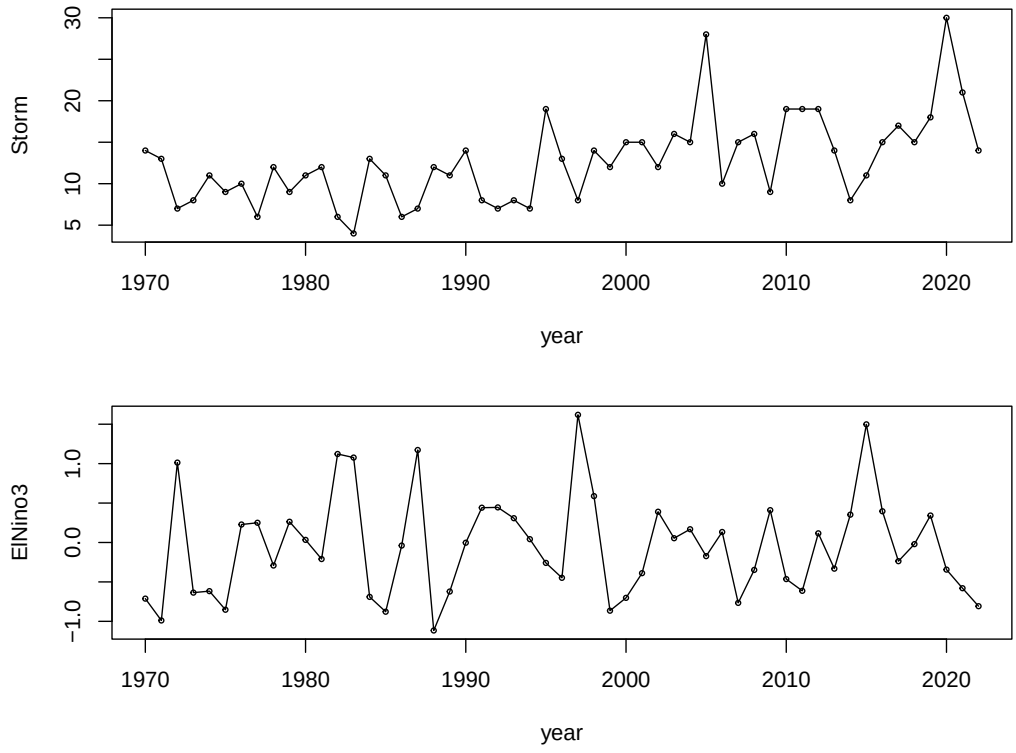

Figure 3.6: Top: the yearly number of North Atlantic Basin tropical storms recorded from 1970-2022. Bottom: the annual El-Nino covariate (ElNino3) over the same period.

While the North Atlantic Basin tropical cyclone record goes back to 1851, some of the earliest data is thought to be incomplete. Specifically, there is concern that some weak storms formed over the open Atlantic waters in the record's earliest years, lived their entire lives over open water, and were never detected. As such, we start our analysis at 1970. This is approximately when the GOES satellites were launched; no storms should have evaded detection thereafter. While one could use indicator variables as additional covariates to quantify undetected storms, we will simply start the record at 1970; see Robbins et al. (2011) and Fisher et al. (2020) for an analysis of the full record.

The level of correlation in this count series is not extreme. In fact, many authors believe that these annual counts to be approximately independent (Robbins et al., 2011). Certainly, if significant year-to-year autocorrelation existed, they would be easier to forecast a year in advance. (Annual forecasting competitions are conducted in May for this series, where Poisson regression methods are typically used with various covariates to predict counts for the upcoming June-November season. Forecasts even a year in advance have generally shown little predictive power). Our model fits below confirm that there is minimal autocorrelation in these counts.

The results are as expected. First, there is little autocorrelation in these counts. Here, we fitted white noise, AR(1), and AR(2) autocorrelation structures in the latent Gaussian process, but both AIC and BIC model selection criteria in Table 3.3 prefer the white noise model. With this white noise structure, the estimated positive trend in the model is $\hat{\beta}_1 = 0.0154$ (0.0025), which translates to a hurricane season that will be some

Model: $\lambda_t = e^{\mu+\beta_1 t+\beta_2 C_t}$								
		$\hat{\mu}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\phi}_1$	$\hat{\phi}_2$	AIC	BIC
WN	Est.	2.0699	0.0154	-0.2830	NA	NA	**283.4695**	**289.3803**
	$\sqrt{E(I'(\hat{\theta})^2)}$	0.0872	0.0025	0.0658	NA	NA		
AR1	Est.	2.0699	0.0154	-0.2831	-0.0018	NA	285.4694	293.3506
	$\sqrt{E(I'(\hat{\theta})^2)}$	0.0874	0.0025	0.0660	0.1623	NA		
AR2	Est.	2.0598	0.0158	-0.2839	-0.0456	-0.2326	286.6195	296.471
	$\sqrt{E(I'(\hat{\theta})^2)}$	0.0760	0.0022	0.0673	0.1628	0.1553		

Table 3.3: Summary of the tropical cyclone Poisson count fit. The AIC and BIC criteria prefer white noise errors and the annual El-Nino covariate appears significant.

four and a half times more active in 2070 than it was in 1970. The standard error of this estimator produces a z-score of about 6.2, indicating a significant increasing trend in the counts and trouble for coastal residents. The estimated coefficient of the El-Nino covariate is $\hat{\beta}_2 = -0.2830$ (0.0658) and is significantly negative, with a z-score of about -4.3. Indeed, an active El-Nino appears to impede tropical cyclone development.

3.4.2 Baseball No-Hitters

Our second series contains the number of annual no-hitter games pitched in major league baseball from 1893 - 2022. A no-hitter occurs when a pitcher (or multiple pitchers) do not allow the opposing team to get any hits over the course of a game. It is usually indicative of a dominant pitching performance.

There has never been more than nine no-hitters pitched in a season; some years do not see any non-hitters thrown. Figure 3.7 shows the no-hitter counts along with two explanatory covariates: the total number of games played in the major league baseball season and the height of the pitching mound. The total number of games played in a

season has changed as more teams have been added to the league; also the number of games team play in a season has varied. Strikes and the Covid-19 pandemic have forced cancellation of some games in a few sporadic years. Of course, the more games played, the more likely it is to have a no-hitter pitched. Our second covariate is the height of the mound. A higher pitching mound is thought to give pitchers an advantage. The height of the pitching mound was reduced from 15 inches to 10 inches in 1969; hence, this covariate could be viewed as a breakpoint or intervention.

Our model here takes

$$\lambda_t = \exp\{\mu + \beta_1 C_{1,t} + \beta_2 C_{2,t}\},$$

where $C_{1,t}$ is the number of games played in year t and $C_{2,t}$ is the height of the pitching mound in year t. As we will see, there is some autocorrelation in these counts.

Table 3.4 shows the results of the Gaussian copula model fit with white noise, AR(1), and AR(2) errors for $\{Z_t\}$. First, both AIC and BIC model selection statistics prefer an AR(1) $\{Z_t\}$. The estimated AR(1) coefficient here is $\hat{\phi} = 0.3199$, which is more autocorrelation than we perhaps expected (no-hitters are extreme performances and rare, which are often modeled as independent; see the peaks over threshold theory in Pickands III (1975)). While we do not consider eliminating the mean μ in the model, the estimates and standard errors for β_2 suggest that pitching mound height does not significantly influence no-hitter counts, but that more no-hitters occur when more games are played.

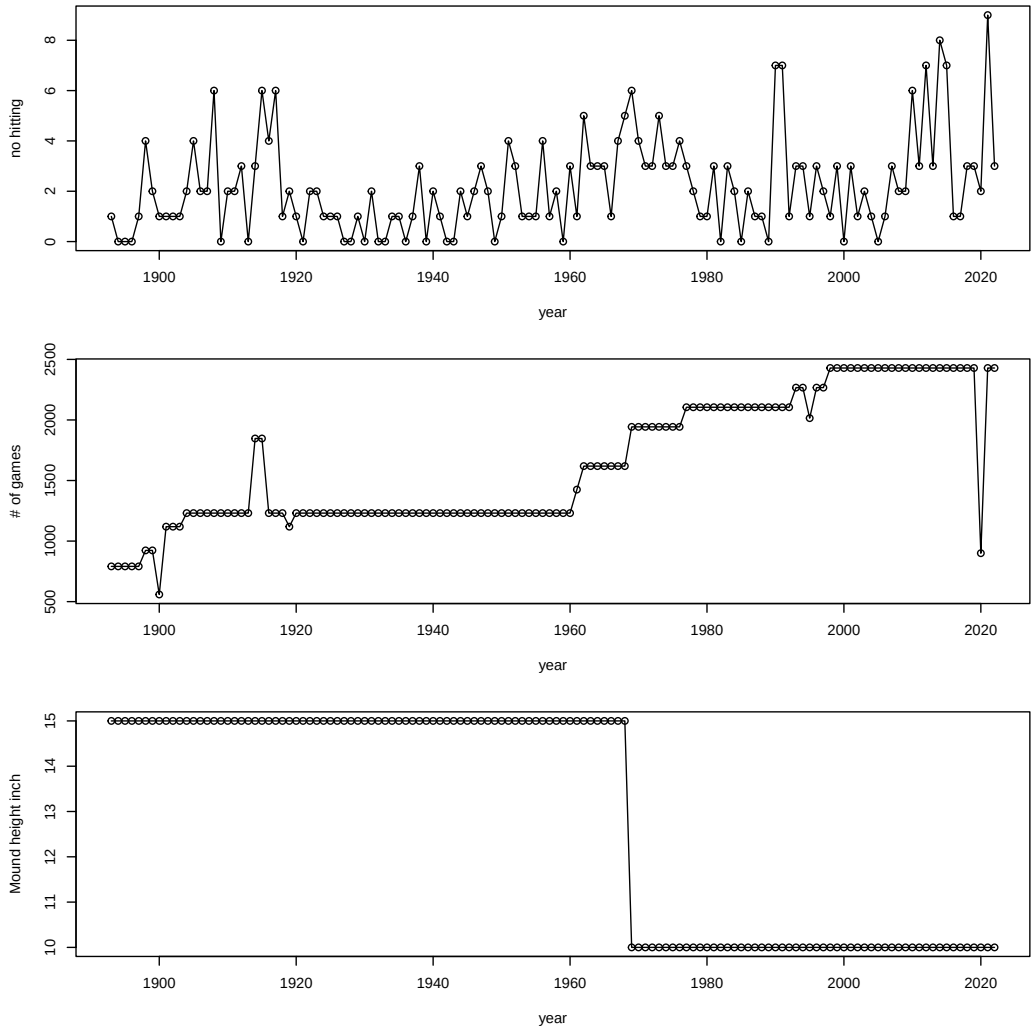

Figure 3.7: Top: Major League baseball annual no hitter counts from 1893-2022. Middle: The number of games played during each year. Bottom: The height of the pitching mound in inches during each year.

Table 3.5 refits the model with the no-hitter covariate eliminated and AR(1) errors. The estimators, standard errors, and conclusions do not change appreciably from the last table.

Model: $\lambda_t = e^{\mu+\beta_1 C_{1,t}+\beta_2 C_{2,t}}$								
		$\hat{\mu}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\phi}_1$	$\hat{\phi}_2$	AIC	BIC
WN	Est.	-1.5367	0.0008374	0.0697988	NA	NA	491.6805	500.2831
	$\sqrt{E(I'(\hat{\theta})^2)}$	1.1982	0.0002743	0.0590602	NA	NA		
AR1	Est.	-1.7639	0.0008968	0.0803000	0.3199	NA	486.4131	497.8832
	$\sqrt{E(I'(\hat{\theta})^2)}$	1.3005	0.0002958	0.0646546	0.0710	NA		
AR2	Est.	-1.6198	0.0008709	0.0726532	0.2792	0.1531	486.7636	501.1013
	$\sqrt{E(I'(\hat{\theta})^2)}$	1.3288	0.0003008	0.0664830	0.0719	0.0750		

Table 3.4: Summary of the No-hitter Poisson count fit. The AIC and BIC criteria prefer AR(1) errors and the pitching mound height covariate appears insignificant.

Model: $\lambda_t = e^{\mu+\beta_1 C_{1,t}}$						
		$\hat{\mu}$	$\hat{\beta}_1$	$\hat{\phi}_1$	AIC	BIC
AR1	Est.	-0.1851	0.0005687	0.3152	**485.9015**	**494.5041**
	$\sqrt{E(I'(\hat{\theta})^2)}$	0.2419	0.0001284	0.0706		

Table 3.5: A refit of the model in the last table with the pitching mound height covariate eliminated.

3.5 Concluding Comments

This chapter reviewed and developed methods that produce time series of Poisson distributed counts. Both stationary and non-stationary settings were considered and inference methods for some of the well-performing model classes were developed. Many of the methods have deficiencies in what they can handle. An implication of the chapter is that the Gaussian copula transformation technique is the most flexible paradigm considered as it produces the most general autocovariances structures possible, easily accommodates covariates, and likelihood methods of inference can be conducted via particle filtering methods. The popular INAR model class was judged deficient in several manners.

Additional research is needed on several fronts. First, ways to generate Poisson counts beyond those discussed here exist. Worthy of mention are structured mixture models (Zheng et al., 2022) and shot noise methods (Jang and Oh, 2021), the latter being related to our superpositioning techniques here. Given the flexibility of the Gaussian copula paradigm, it seems pedantic to investigate these classes further unless they can be shown to be flexible, parsimonious, accommodate covariates, and have analyzable likelihood functions. Second, asymptotic normality of the parameter estimators was not proven here, but needs to investigated. We are unsure how to do this when the likelihood function is intractable as in the Gaussian copula setting. Third, multivariate versions of the methods are worthy of development. Here, one needs to settle on a definition of multivariate Poisson — many are possible (Teicher, 1954; Kocherlakota and Kocherlakota, 2017; Inouye et al., 2017). Finally, extensions of the methods to the zero inflated case, which frequently arises with Poisson analyses (Lambert, 1992; Fernando et al., 2022), are worth considering.

Chapter 4

Trends in Northern Hemispheric Snow Presence

4.1 Introduction

Snow cover plays a critical role in the Earth's hydrological processes and its impact on the broader global climate is of great interest (Barnett et al., 2005; Karl et al., 2009; Goudie, 2018; Van Mantgem et al., 2009). Snow greatly influences the global energy balance due to its high albedo and insulating characteristics and is therefore a prominent indicator of climate change (Liston and Hiemstra, 2011; Mote, 2003; Lawrence and Slater, 2010; Callaghan et al., 2011). On a regional scale, the spatial consistency (patchiness) of snow cover can influence surface temperatures via horizontal variations in absorbed solar radiation. Continental-scale snow cover acts to maintain thermal stability in the Arctic and subarctic regions, possibly inducing changes in global

circulation patterns attributable to large-scale releases of carbon and methane gas (Zona et al., 2016). While the amount of water available in the snowpack is quantified in snow depths and/or snow water equivalents (SWE), areal snow presence/coverage defined by snow cover extent (SCE) is often used to estimate the location and availability of regional water resources (Mote et al., 2018; Serreze et al., 2000; Robinson et al., 1993).

Remotely sensed satellite images are common sources of SCE data; these images provide spatial and temporal observations that can be used in regional and continental-scale analyses. Satellite data is used here to estimate SCE trends, allowing us to assess SCE changes over time and space. The Satellite-derived SCE data product investigated here is binary, with snow presence being recorded as unity and snow free ground being assigned zero.

Some mid-latitude locations have sporadic snow coverage, with snow cover typically lasting only a few weeks at a time, even during the height of winter. The majority of our work lies with introducing a mathematical model and developing the statistical methods needed to analyze trends in autocorrelated and binary-valued sequences. The model is flexible enough to adapt to the data from many of our study cells.

Statistical analysis of snow data has been debated in the climate literature, especially in regard to trend and uncertainty assessment — see (Yue et al., 2002) and the references therein. Here, a flexible mathematical model and rigorous accompanying statistical methods are used to estimate trends and accurately assess their uncertainity margins. Some nuances arise in this pursuit. First, as our SCE data are recorded weekly, annual periodicity needs to be taken into account. Second, since SCE data is correlated,

with snow presence in a week making snow presence in adjacent weeks more likely, serial autocorrelation needs to be accounted for in trend uncertainity quantifications. Finally, previous authors have noted data quality issues (Bormann et al., 2018; Estilow et al., 2015) in some cells that need to be addressed, without pinpointing the specific problematic cells. We carefully address this issue below. The general pattern of results found here agrees with trends found in other studies using more rudimentary statistical approaches (Brown and Robinson, 2011a; Lemke et al., 2007; Notarnicola, 2022).

The rest of this chapter proceeds as follows. Section 2 describes the SCE data used in this study and its nuances. Section 3 introduces the mathematical model and statistical methods needed to quantify the problem, including the all-important uncertainty calculations for our trend estimates. Section 4 presents a simulation study, showing that model parameters can be accurately estimated from a half-century of weekly observations. Section 5 presents two case studies, analyzing observations from a cell in North Dakota that is actually experiencing increasing snow coverage. We also give an example of data from a cell having poor data quality. Section 6 presents results for the entire Northern Hemisphere (NH) and discusses our general findings and their implications. Section 7 concludes with comments and remarks.

4.2 Data

The data studied here were aggregated from daily satellite flyovers, with SCE values being estimated manually weekly by meteorologists for each studycell. Specifi-

cally, this study uses the Northern Hemisphere Weekly Visible Satellite Charts data from the Climate Data Record as developed by the National Oceanic and Atmospheric Administration (NOAA) (Robinson et al., 2012). The data are available at `https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.ncdc:C00756`.

The Rutgers University Snow Lab at `http://climate.rutgers.edu/snowcover/` was an integral part of the construction of the data product studied here and is a useful repository for literature and links to this and other SCE data sets. This study examines the time period August 1967 - July 2021. For cell structure, the data use NOAA's 89×89 Cartesian grid that overlies a polar stereographic projection of the NH. The product contains $88 \times 88 = 7,744$ cells with a resolution of 190.4km at 60°N. The SCE data during the first week in December 2020 is plotted in Figure 4.1.

Thorough descriptions of the data are provided in Dye (2002) and Estilow et al. (2015). Early discussion of the data's production is found in Wiesnet et al. (1987) and Robinson et al. (1993). Before June of 1999, NOAA used the first clear-sky day during each week to estimate the SCE. If the cell contains at least 50 percent snow coverage, its SCE was assigned as unity; otherwise, it is assigned zero.

With the introduction of the Interactive Multisensor Snow and Ice Mapping System (IMS), the methods used to estimate SCE changed in June 1999. These methods use different data and a refined grid partition of 24km covering the NH to estimate snow presence/absence on the 190.4km resolution grid, these changes are detailed in Estilow et al. (2015). Brown et al. (2007) did not find evidence of inhomogenities over Northern Canada before and after the 1999 change; however, Déry and Brown (2007) claim that

pre-1999 methods overestimate snow presence in mountainous regions during Spring ablation. An analysis of the 1999 changes are provided later.

There are other changes in the data construction procedure (Estilow et al., 2015) for the data product studied here. In June of 1977, the Defense Meteorological Satellite Program data supplemented the data record. Next, Geostationary Meteorological Satellite imagery was introduced to the data construction in February of 1988 and January 1989. In May of 1999, the IMS system was introduced into the data construction process. Finally, NOAA took over responsibility of data construction in June of 2008.

Ten years of observations for a cell located near Napoleon, ND (46.4309°N, 99.8852°W), from August 1967- July 1976 are displayed in Figure 4.2. This cell will be analyzed in detail in Section 5. The graph reveals the ephemeral nature of snow processes here, starting each year circa November and typically lasting through early April. Once snow cover is present, it usually stays through Spring ablation; however, years exist when snow is absent mid-winter (1967-1968 and 1973-1974, for examples).

The data in this study contain 7,744 NH cells, 3,011 of which are deemed to be over land. See the metadata for the key to this partition, or to obtain cell areas. Winter centered years are used here so that the first week of any year corresponds to the first week of August. This scaling prevents a single winter's snow record from lying within two distinct calendar years. Shifting in this manner is done for convenience only — the scaling does not influence any trends.

4.2.1 Data Preprocessing

Before beginning any analysis, each land cell was categorized into four subgroups, depending on its data. Group 1 includes all cells that reported 10 or fewer weeks of snow cover during the 1967-2020 period of record (2,808 weeks). This group also contains any cell that reported 10 or fewer weeks of bare ground over the record period. Group 1 cells primarily lie in the southerly latitudes of the NH, which rarely experience snow, or the interior Greenland icecap, which is almost always under snow cover. All 1,131 Group 1 cells were excluded as any trends computed from these records lack sufficient information/variability to fit our model (there are more model parameters than changes in snow presence/absence).

Group 2 contains 72 cells that were insufficiently fitted by our model (our model is the subject of the next section). While these cells all had more than 10 snow/bare ground weeks during the 2,808 week study period, they typically did not have many more. While one can theoretically obtain trend estimates for cells in Group 2, error margins obtained are so large that any trend estimates would essentially be meaningless. These cells were primarily located in Southern China, the Southern United States, and Coastal Greenland. While one could combine Group 1 and Group 2 together into a single "insufficient information" group, we keep the groups separate on this technical distinction: trend error margins do not exist in Group 1, and while they exist for Group 2, they are too large to make any conclusions.

Several studies (Bormann et al., 2018; Estilow et al., 2015) discuss the unre-

liable snow presence/absence estimates in mountainous regions in the pre-1999 data. Figure 4.3 plots the data from an example Group 3 cell located in the Chinese Himalayan Mountain Range near 27.9682°N, 97.7094°E. Several issues are apparent. The top plot shows that some of the earlier years in the record have no snow cover in winter weeks, but some snow cover during summer weeks. The bottom plot reveals that the pre-1999 years report very little snow cover compared to the post-1999 years. While the methodological revisions in 1999 may render the post-1999 data believable, this cell is best excluded in a trend analysis. As such, our immediate objective is to construct a quality control method to be applied to all cells before trend analysis.

Let $\{X_t\}$ denote the two-state snow presence/absence series in time. Here, $X_t = 1$ means that snow cover is present at time t and $X_t = 0$ means that snow is absent at time t. Let S_n be the number of weeks of snow on the ground during year n:

$$S_n = \sum_{\nu=1}^{T} 1_{[X_{(n-1)T+\nu}=1]},$$

where 1_A denotes the indicator of the event A and $T = 52$ is the period of the data.

As a quality control measure, a traditional cumulative sum (CUSUM) test statistic is applied to $\{S_n\}$ from each land cell not in Groups 1 or 2. The CUSUM statistic has been widely used for statistical quality control for more than 50 years (Bissell, 1969). The CUSUM method checks for structural breaks in the $\{S_n\}$ data series. The significance level for the test was set to 1×10^{-5}. If the CUSUM statistic for the cell has a p-value less than this significance level, the cell is deemed corrupted

and is classified as belonging to Group 3.

Group 3 contains 190 cells. These cells overwhelmingly reside in the mountainous regions of the NH (Rockies, Alps, Caucasus, Scandinavia, and Himalayas) and are omitted from further analysis. The discarded cells largely align with the regions discussed in Bormann et al. (2018). Our CUSUM analysis addresses a point raised in Estilow et al. (2015): "More research is needed to determine whether SCE analysis in mountainous regions (e.g., the Tibetan Plateau) shows systematic change during this time period." We concur with Bormann et al. (2018): the analysis in the Tibetan and other high mountain regions changed with the implementation of the IMS based product in May of 1999. The data before 1999 is unreliable in many high mountain regions.

Figure 4.4 depicts the Group category of all cells; there are 1,618 violet-shaded cells where our model fit was deemed reliable. These cells cover most areas of the NH where snow is seasonally persistent. A spreadsheet containing the group numbers of our cells, and all code used for this project, is available at `https://github.com/JiajieKong/Snow-Presence-Trends`.

Several previous studies of this data exist. Déry and Brown (2007) studies the data from January 1972 - December 2006. Déry and Brown (2007) report significant temporal autocorrelation in the data, at both weekly and annual scales. Autocorrelation makes some statistical methods such as Sen's slope troublesome for trend analysis as uncertainties are extremely difficult to estimate with such a non-parametric method (Yue et al., 2002). Negative trends in SCE area are reported in Déry and Brown (2007) from March through June. Figure 4.3 in Lemke et al. (2007) shows March-April snow

cover departures by subtracting the percentage coverage (by cell) of weeks with snow cover from 1988-2004 minus the same percentage coverage during 1967-1987. While it is not clear how to interpret such a statistic as any type of smooth trend, the largest reductions in that study occurred roughly between the 0°C and 5°C isotherms.

4.3 Model and Estimation

4.3.1 The Model

Our methods use a two-state Markov chain model on the states $\{0, 1\}$ to describe the series for a fixed cell. This model can accurately quantify trend uncertainty as shown below. State zero indicates lack of snow and state one signifies snow cover. The transition probability matrix of this chain from week $t-1$ to week t is parameterized as

$$\mathbf{P}(t) = \begin{bmatrix} p_{0,0}(t) & p_{0,1}(t) \\ p_{1,0}(t) & p_{1,1}(t) \end{bmatrix}.$$

Here, $p_{0,1}(t)$ is the probability that snow cover is present at time t given that it is absent at time $t-1$. The other three elements in the matrix are similarly interpreted. There are only two free quantities in $\mathbf{P}(t)$ at any t since $p_{0,0}(t) = 1-p_{0,1}(t)$ and $p_{1,0}(t) = 1-p_{1,1}(t)$.

The marginal probability distribution of X_t at time t will be denoted by $\boldsymbol{\pi}(t) = (\pi_0(t), \pi_1(t)) = (P(X_t = 0), P(X_t = 1))$. Because the chain commences with an observation in August, the startup condition $\boldsymbol{\pi}(1) = (1, 0)$ is taken, signifying that the

chain starts with bare ground. With this initial distribution, $\boldsymbol{\pi}(t)$ is computed via

$$\boldsymbol{\pi}(t) = \boldsymbol{\pi}(1) \prod_{k=2}^{t} \mathbf{P}(k). \tag{4.1}$$

For each pair of times $t_1 < t_2$ in $\{1, \ldots, N\}$, the transition matrix

$$\mathbf{P}^*(t_1, t_2) = \prod_{t=t_1+1}^{t_2} \mathbf{P}(t)$$

contains the four transition probabilities of snow cover/absence from time t_1 to time t_2.

Since $p_{0,1}(t)$ and $p_{1,0}(t)$ are probabilities, they take values in $[0, 1]$. Hence, these quantities are modeled with the logistic-type link

$$p_{0,1}(t) = \frac{1}{1 + \exp(-m_t)}, \quad p_{1,0}(t) = \frac{1}{1 + \exp(-m_t^*)},$$

where m_t and m_t^* contain seasonal effects and trend parameters. These quantities are posited to have the additive form

$$m_t = \mu_t + \alpha t, \quad m_t^* = \mu_t^* + \alpha^* t,$$

where the parameters are clarified as follows. For the weekly observations analyzed here, the period $T = 52$ weeks is forced to the data by omitting any observations that occur at the end of July (one day during non leap years and two days during leap years). This tactic results in little loss of precision, see (Lund et al., 2006) for similar tactics. The

parameters μ_t and μ_t^* contain seasonal effects that are sinusoidaly parametrized as

$$\mu_t = A_0 + A_1 \left[\cos\left(\frac{2\pi(t-\tau)}{T}\right)\right], \quad \mu_t^* = A_0^* + A_1^* \left[\cos\left(\frac{2\pi(t-\tau^*)}{T}\right)\right].$$

Observe that μ_t and μ_t^* are periodic with period $T = 52$ weeks and obey $\mu_{t+T} = \mu_t$ and $\mu_{t+T}^* = \mu_t^*$. The quantities A_0 and A_0^* govern the length of the snow season. For example, when $A_0 > 0$, the season where snow is present tends to last longer than the snow free season (and vice versa). The parameters A_1 and A_1^*, which are assumed positive for mathematical identifiability of the cosine waves, control how fast snow to bare ground transitions take place (and vice versa). The parameters τ and τ^* are phase shifts. Since $p_{0,1}(t)$ and/or $p_{1,0}(t)$ are maximized when m_t and/or m_t^* is maximized, and the cosine function is maximized when its argument is zero, $p_{0,1}(t)$ is maximized at week τ, which is typically in the late Fall or early winter, and $p_{1,0}(t)$ is maximized at week τ^*, which typically occurs in the late winter or early spring. The parameters α and α^* are linear trend parameters and govern how fast snow cover changes are happening. While the above model has a linear time trend and a simple cosine seasonal cycle, other forms of trends and seasonality could be used if needed.

Our periodic Markov chain model allows X_t to be autocorrelated in time t. Indeed, week to week SCE data exhibits correlation: if snow is present/absent at week t, it is more likely to be present/absent at week $t + 1$. Good models for snow depth processes are also correlated in time. Indeed, Woody et al. (2009b) argues for a Markov structured storage model for daily snow depths: the snow depth today is the snow depth

yesterday, plus any new snowfall, minus any meltoff or compaction between yesterday and today. Our model is not a classical Probit count time series model as these are typically used for uncorrelated data; see Chib and Greenberg (1998) for more on probit modeling. A Markov model for binary data is parsimonious in that there are only two free parameters in $\mathbf{P}(t)$ for each fixed t. While seasonal and trend features need to be incorporated into $\mathbf{P}(t)$ to handle the periodic nature of snow, the overall model is very parsimonious. Comparing further, a time homogeneous Markov model for categorical sequences taking on S distinct categories has $S(S-1)$ free parameters, which is quite large for a large S. Additional parameters would be needed to make this model periodic.

Figure 4.5 shows a simulation of ten years of a binary snow presence process. The parameters chosen for $p_{0,1}(t)$ are $A_0 = 3, A_1 = 10, \tau = 25, \alpha = 0$, and those for $p_{1,0}(t)$ are $A_0^* = 0, A_1^* = 10, \tau^* = 5, \alpha^* = 0$; specifically, there is no trend in the simulated data. One sees that each and every year, snow presence begins in the Fall and stays on the ground until Spring. Oscillations between seasonal snow presence and bare ground occur in the Fall, and snow vanishes completely during the summer. Additional simulations show that this simple Markov chain model produces a flexible suite of snow presence/absence series.

4.3.2 Parameter Estimation

Suppose that the data sample $\mathbf{X} = (X_1, \ldots, X_N)'$ is available for a cell. We assume that N is a multiple of T to avoid trite work with fractional portion of years; this said, the methods are easily modified to accommodate fractional parts of years if

needed. Let $d = N/T$ denote the total number of years of observations; we work with observations indexed as the years $1, 2, \ldots, d$

Let $\mathbf{\Theta}$ denote all model parameters contained in m_t and m_t^*. These include A_0, A_1, τ, α and their starred counterparts. The statistical likelihood of $\mathbf{\Theta}$, denoted by $L(\mathbf{\Theta}|\mathbf{X})$, can be derived from the Markov property and is

$$\ln(L(\mathbf{\Theta}|\mathbf{X})) = \sum_{t=2}^{N} \ln\left(p_{X_{t-1},X_t}(t)\right). \tag{4.2}$$

The quantities $p_{i,j}(t)$ depend on $\mathbf{\Theta}$. Numerically maximizing this likelihood is the classical statistical way of estimated the components in $\mathbf{\Theta}$; that is, likelihood estimates model parameters as those that make the observed data most likely. These estimates will be used later in assessing variability (uncertainty) margins of the trends. The data $X_1, \ldots, X_N$ is held fixed in this maximization. While explicit forms for the estimators of the components in $\mathbf{\Theta}$ do not exist, likelihood estimates can be obtained numerically. The R programming language version 4.1.2 was used for all statistical coding in this study, the numerical routine "optim" was employed for optimization in this study.

4.3.3 Trend Estimation and their Uncertainties

Trends will be phrased in the number of snow days lost/gained per decade. For example, future trends will be phrased as a loss of one day of annual snow cover over a decade. Trends are estimated directly from the data product for all Group 4

cells. The linear rate of SCE change is quantified by $\hat{\beta}$ defined by

$$\hat{\beta} = \frac{\sum_{k=1}^{d} S_k(k-\bar{k})}{\sum_{k=1}^{d}(k-\bar{k})^2} = \frac{\sum_{k=1}^{d} S_k(k-\bar{k})}{Q}, \tag{4.3}$$

where $\bar{k} = (d+1)/2$ is the average time index and the denominator can be verified as $Q = d(d+1)(d-1)/12$. While the units of β are weeks of snow cover gained/lost per year, we will scale $\hat{\beta}$ to days of snow cover gained/lost per decade for interpretability; this simply multiplies raw trends and their standard errors by 70.

Our next objective is to obtain a standard error for $\hat{\beta}$. Taking a variance in (4.3) gives

$$\mathrm{Var}(\hat{\beta}) = \frac{\sum_{k=1}^{d}\sum_{\ell=1}^{d}(k-\bar{k})(\ell-\bar{k})\mathrm{Cov}(S_k, S_\ell)}{Q^2}.$$

This computation requires $\mathrm{Cov}(S_n, S_{n+h})$ for every $h > 0$ and n in $\{1, \ldots, d-h\}$. Details for this computation are provided in the Appendix. The standard error of $\hat{\beta}$ accounts for any correlation in the SCE data.

To statistically test whether or not SCE is changing, we want to test the null hypothesis that $\beta = 0$ against the alternative that $\beta \neq 0$. Invoking asymptotic normality of the estimator $\hat{\beta}$, this is assessed through the Z-score statistic

$$Z = \frac{\hat{\beta}}{\mathrm{Var}(\hat{\beta})^{1/2}},$$

which is compared to the standard normal distribution to make conclusions. One typically reports a p-value for the test to assess significance of the trends; this is illustrated

further in Section 6.

4.4 A Simulation Study

This section studies our model and estimation procedure via simulation, illustrating the model's capabilities and how parameters are estimated.

To demonstrate the model's flexibility, Figure 4.6 provides ten year sample plots of snow presence/absence series generated by models for five sets of parameter values. Only ten years of data are shown as it becomes visually difficult to see data features with longer series (the plot becomes "compressed"). Table 4.4 lists all parameters considered. The unstarred parameters govern $p_{0,1}(t)$, which controls transitions from bare ground to snow cover; the starred parameters govern $p_{1,0}(t)$, which controls transitions from snow cover to bare ground.

Sample Simulated Series

Model	A_0	A_1	τ	α	A_0^*	A_1^*	τ^*	α^*
I	0	30	25	0	0	30	0	0
II	0	30	25	0	0	30	42	0
III	0	30	20	0	0	30	0	0
IV	-30	30	25	0	30	30	0	0
V	30	30	25	0	-30	30	0	0

Table 4.1: Model 1 is the base case: equal transitions from no snow to snow in both Fall and Spring. Model II allows for more variability in the Spring snow presences. Model III allows more variability in the Fall snow presences. Model IV is for a cell that rarely experiences snow; Model V describes a very snowy cell.

Models I - V have no trend. Models with trends will be considered below. The parameters for Model I were chosen to represent a scenario that is seasonally regular,

with snow cover becoming present in the late Fall and staying until Spring ablation. The parameters A_0 and A_0^* are set to zero, making the winter "snow season" last roughly half the year. Model II has the same parameters as Model I, except that τ^* was changed from 0 to 42, shifting the cosine wave governing $p_{1,0}(t)$ from its Model I settings. This change makes both $p_{0,1}(t)$ and $p_{1,0}(t)$ relatively large during the Spring months, which induces a Spring SCE season that oscillates more frequently between bare ground and snow cover. Model III has the same parameters as Model I, except that τ was changed from 25 to 20, making both $p_{0,1}(t)$ and $p_{1,0}(t)$ large during the Fall. This makes bare ground to snow cover oscillations more common in the Fall. While we do not illustrate it here, increasing A_1 or A_1^* makes "transitions" from winter to summer (and vice versa) shorter (sharper). The parameters in Model IV are set to a lower latitude setting where snow only occurs sporadically during the middle of winter. This was done by decreasing the A_0 parameter from 0 to -30 for $p_{0,1}(t)$ and increasing A_1^* from 0 to 30 (compared to Model I). Model V's parameters correspond to a high latitude case where snow cover is present most of the year. This was done by increasing A_1 from 0 to 30 and decreasing A_1^* from 0 to -30 (compared to Model I). These and other simulations show that the model can generate a wide range of SCE patterns.

To illustrate trend features, we choose parameters that bring Model IV above to a very snowy setting, and Model V above to a non-snowy scenario. These are done over a 1000 year time period. These scenarios are not climatologically realistic, but

were chosen to demonstrate the overall flexibility of the approach. Figure 4.7 plots

$$\frac{1}{T}\sum_{\nu=1}^{T} X_{(k-1)T+\nu}$$

against the annual index k. This quantity is the proportion of days of year k where snow cover is present. The top graphic in Figure 4.7 corresponds to Model IV, except that α was changed from zero to 0.001 and α^* is changed from zero to -0.001. Here, the proportion of snow covered days rises from almost zero to approximately 80%. The antipodal scenario is illustrated in the bottom graphic of this figure. This moves a very snowy location to one with infrequent snow cover. This was done by taking Model V's parameters, but changing α from 0 to -0.001 and α^* from 0 to 0.001.

Turning to estimation, our first simulation case studies a 100-year series ($N = 5200$), which is roughly double the length of the data studied here. The parameters chosen for this simulation are those for Model I - V above; there is no trend in these simulations. These parameters were chosen to correspond to fitted parameters in some of our cells.

The true parameters serve as the initial values in all simulation studies. Different initial values were used in each case and usually don't influence convergence of the algorithm. In a few sporadic cases, the log likelihood function is "somewhat flat" and convergence is an issue. While one should try multiple initial guesses in practice, these issues decay as the series length increases. Figure 4.8 shows boxplots of the eight parameter estimators aggregated from 1000 independent simulations for Model I - V.

The solid line in each boxplot demarcates the median of the 1000 estimators for that parameter. One sees little bias in the estimators. Specifically, the estimation procedure was able to discern that there was no trend in the series. Additional simulations (not shown here) indicate that any estimator bias recedes with increasing series length.

Estimation of the eight model parameters by likelihood appears to work well in this case.

Our second simulation moves to a case with trends. This simulation takes the same series length and parameters as the above simulation, but modifies the trend parameters to $\alpha = 0.001$ and $\alpha^* = -0.001$. All parameters are fixed for the duration of the series. Figure 4.9 shows boxplots of the estimates of each parameter for modified Model I - V and the results look quite good; importantly, trend parameters are accurately estimated. While the trend parameters are small in magnitude in this simulation, they will be converted to days of snow cover gained/lost per decade later for ease of interpretability. Overall, the model parameters are reasonably accurately estimated with 100 years of weekly data.

4.5 A Sample Cell

This section analyzes snow coverage in a cell near Napoleon, ND (46.4309°N, 99.8852°W). This cell contains a region studied in (Woody et al., 2009b).

In the ensuing analysis, our null hypothesis is that the snow presence/absence

series is not changing. This corresponds to the null hypothesis

$$\mathrm{H}_0 : \alpha = \alpha^* = 0,$$

with the alternative hypothesis being that SCE is changing.

Table 4.2 below shows the maximum likelihood estimates of the parameters in the Section 3 model along with a single standard error. All estimated parameters appear significantly non-zero except for the α parameters (one does not usually assess whether or not the phase shift parameters τ and τ^* are zero). Statistical significance is assessed using asymptotic normality. There is no statistical evidence to conclude that α is different from zero with a p-value of 0.7708, and we conclude that $p_{0,1}(t)$ is not changing. As $p_{0,1}(t)$ governs transitions from bare ground to snow cover, this implies that the snow season is starting about the same time and has not changed over the study. In contrast, α^* is concluded to be significantly negative with a p-value of 0.0001. A negative α^* makes $p_{1,0}(t)$ smaller, which makes it harder for snow to disappear when it is on the ground. This translates to a later Spring ablation.

Table 4.2: Model parameter estimates and their standard errors for a cell containing Napoleon, ND.

Parameter	A_0	A_1	τ	α
Estimate	-3.2016	4.1499	24.3492	0.0000382
Standard Error	0.2538	0.2936	0.26460	0.0001315
Parameter	A_0^*	A_1^*	τ^*	α^*
Estimate	1.7258	3.7889	49.8375	-0.0004935
Standard Error	0.3774	0.4139	0.3800	0.0001273

To assess changes in the snow presences, the $\hat{\beta}$ statistic in (4.3) is $\hat{\beta} = 0.038613$ and $\text{Var}(\hat{\beta})^{1/2} = 0.0247$. This translates to an additional 2.702 days of SCE over a decade. The test statistic for changing SCE is $Z = 1.5633$, which has a two-sided p-value of 0.1180. This p-value is insignificant for a standard 5% test, but is borderline significant for a 10% test. Conclusions may change further if one-sided alternative hypotheses are considered. The Napoleon cell is experiencing increasing (and not decreasing) SCE changes.

The top panel of Figure 4.2 displays a 10 year plot of weekly snow presence/absence values at the Napoleon cell. The bottom panel depicts data simulated from our model with the parameter estimates displayed in Table 4.2. Both series are of length 10 years, starting on August 1, 1967 and continuing through July 31 1976. Visual inspection of the top and bottom panes of Figure 4.2 indicate the simulated data appears to model the real data quite well.

A rudimentary goodness-of-fit tests can be performed by comparing the proportion of observed unit series values to average values from the model. Note that

$$\mathbb{E}(X_t|X_{t-1}) = p_{X_{t-1},1}(t).$$

Hence, the mean number of unit values can be obtained via

$$\frac{1}{N}\sum_{t=2}^{N} p_{X_{t-1},1}(t).$$

Our fitted model has a long-term average of 0.300, which is extremely close to the proportion of observed unit values, which was 0.299. We do not develop goodness-of-fit tests further here.

4.6 Results

This section reports results for the 1,618 cells where our model fit was deemed reliable. Figure 4.10 spatially portrays the trends $\hat{\beta}$ over all analyzed cells. The corresponding Z-scores for the trend statistics are displayed in Figure 4.11. In totality, 573 of the cells (35.41 %) report a positive $\hat{\beta}$ (increasing snow), while 1045 cells (64.58 %) show a negative $\hat{\beta}$. This is almost a 2 to 1 margin preference for declining to advancing snow cover. The average trend over the 1,618 analyzed cells has lost 1.522 days of snow cover per decade.

Examination of the spatial structure in Figures 4.10, 4.11, and 4.15 reveals regions of increasing and decreasing snow presence. Decreasing snow presence in the Arctic, particularly in Russia and Western Canada and Alaska, is seen, agreeing with the findings of Bormann et al. (2018); Estilow et al. (2015). Increasing snow is encountered in Eastern Canada, the Kamchatka Peninsula, and Japan. Other regions experiencing positive trends can be seen in Figure 4.10. The Figure 4.11 Z-scores are deemed significantly non-zero should they exceed 2.0 in absolute value (the exact two-sided confidence level is 0.9544). Red colored Z-scores demarcate cells where snow cover is declining with at least 97.72% confidence and blue colors depict increasing snow with at least 97.72%

confidence. Overall, a general declining snow presence is seen along coastal areas and the periphery of the continental snowpack, with some inland increases in SCE, especially within North America. This pattern could be associated with a deeper snowpack within continental interiors and a shallower or patchier snowpack along its edges, leading to more rapid retreat of the snowpack and a longer duration of its center. This coincides with the finding of the 4th IPCC report in (Lemke et al., 2007).

The left panel in Figure 4.12 shows a histogram of the trend estimates $\hat{\beta}$ over all analyzed cells. The estimated trends $\hat{\beta}$ are approximately normally distributed with a mean of -.02174 (the loss of 1.522 days of SCE per decade alluded to above). The center and right panels in Figure 4.12 show histograms of the $\hat{\alpha}$ and $\hat{\alpha}^*$ parameters, respectively, over these same cells. The average α is -0.0004168 and the average α^* is -0.0001431.

We now move to an investigation of temporal changes in the total SCE area. Figure 4.13 plots the total snow covered area in each week of the study over all analyzed cells. Areas were obtained by adding the area of all snow covered cells; cell areas are included with meta data (Robinson et al., 2012).

The seasonal cycle of SCE is evident, with winter weeks having the most prevalent snow cover. While interannual variability is apparent, changes in this series are not visually evident in a visual inspection.

Let $\{G_t\}$ denote the snow cover area lost/gained per decade by season. This series is plotted in Figure 4.13 and is now analyzed with a simple linear regression in a periodic environment. More on periodic regression analyses can be found in Lund et al.

(1995b) and Lund (2006). Our regression model for the snow cover areas at week ν is

$$G_{nT+\nu} = \mu_\nu + \beta_\nu(nT + \nu) + \epsilon_{nT+\nu}, \quad n = 0, 1, \ldots, d-1. \tag{4.4}$$

The parameter β_ν quantifies the linear rate of change in data during the νth week, for $1 \leq \nu \leq 52$; μ_ν is a location parameter for week ν. The trend slope β_ν is allowed to depend on the week of year ν, enabling us to investigate changes within a calendar year. The regression errors $\{\epsilon_t\}$ are only assumed to have a zero mean for every week ν — no other structure is needed.

The week ν trend β_ν can be estimated by ordinary least squares via

$$\hat{\beta}_\nu = \frac{\sum_{n=1}^{d}(G_{nT+\nu} - \bar{G}_\nu)(nT + \nu)}{\sum_{n=1}^{d}(nT + \nu - \bar{t}_\nu)^2} \tag{4.5}$$

(Lund et al., 1995b). Here, $\bar{t}_\nu = d^{-1}\sum_{n=1}^{d}(nT + \nu) = (d+1)T/2 + \nu$ and $\bar{G}_\nu = d^{-1}\sum_{n=1}^{d} G_{nT+\nu}$. The denominator in (4.5) can be worked out as $T^2 d(d+1)(d-1)/12$. We will not delve into standard error computations for $\hat{\beta}_\nu$; these would require estimating the correlation structure of $\{\epsilon_{nT+\nu}\}$. One could use weighted least squares estimates that take into account this correlation; however, since ordinary and weighted least squares have the same asymptotic efficiency, this is not needed. We refer the interested reader to Lund et al. (2001) and Lee and Lund (2004) for more on these issue.

Figure 4.14 plots estimates of β_ν against ν for each week of year; see Lund

et al. (1995b) for the equations to fit this model. Increasing SCE is evident in the Fall (late October through early December), with a corresponding decrease in late Winter through Summer. While increases span only a few months and include brief peaks above 0.5 million km^2, the decrease spans February - September, with losses below 0.5 million km^2 from May through July. This implies that while the snow season is experiencing a shift toward an earlier onset and ablation period, there is a more pronounced decrease in snow cover through the warm season that is not being offset by increased snow in the Fall and early Winter. Implications of this finding include a change in seasonal water availability.

As a final task of this section, we analyze possible issues induced by the methodological changes used to extract the SCE data (these are called breakpoint times or interventions). As noted in Section 2, there are five potential breakpoints in the data (Estilow et al., 2015).

Breakpoints are discontinuity features in time series that occur at known times. Breakpoints (also called interventions) often take place when measuring conditions change, such as station relocations or updates to gauge sensors. We will investigate possible breakpoints in June 1977, February 1988, January 1989, May 1999, and June of 2008, all times where the methods to extract the zero-one SCE data changed. It would require more work to find and adjust the data for undocumented breakpoint times (called changepoints when the time of the discontinuity is unknown). Future work will assess changepoint features and homogenize the data in the individual cells. A caveat: while Lu et al. (2010) is one changepoint reference for approximately normally

distributed temperature data, methods to homogenize zero-one count data have not yet to be developed (or have not matured) in the statistics literature.

This will be done for the total SCE only; a deeper analysis exploring the effects on the individual cells is omitted. To conduct this analysis, shifts are allowed during June 1977, February 1988, January 1989, May 1999, and June 2008. The setting is quantified through the regression

$$G_{nT+\nu} = \mu_\nu + \beta_\nu(nT+\nu) + \sum_{i=1}^{5} \Delta_i 1_{[nT+\nu \geq b_i]} + \epsilon_{nT+\nu}, \quad n = 0, 1, \ldots, d-1. \tag{4.6}$$

Here, $b_1 = 515, b_2 = 1069, b_3 = 1117, b_4 = 1654$, and $b_5 = 2126$ are the week time indexes of the June 1977, February 1988, January 1989, May 1999, and June 2008 breakpoint times and Δ_i is the associated shift size of the ith breakpoint time. We do not allow Δ_i to depend on ν, but could do so if desired.

Next, a backward elimination regression procedure at level 95% was conducted to eliminate insignificant breakpoint times. This procedure found the June 1977 and June 2008 breakpoints to be insignificant. The regression model was refitted with the other three breakpoints, Δ_2, Δ_3, and Δ_4. Estimates of these three shift sizes are shown in Table 4.6. The listed p-values for these shift sizes indicate high confidence that the methodological changes impacted observations, essentially making observations "snowier". In fact, the only positive trend slopes occur from October - December after the breakpoints are taken into account.

	$\hat{\Delta}_2$(Feb 1988)	$\hat{\Delta}_3$(Jan 1989)	$\hat{\Delta}_4$(May 1999)
Est.	-1.871	1.196	0.407
S.E.	0.291	0.291	0.144
p-value	1.48×10^{-10}	4.06×10^{-5}	0.0047

Table 4.3: Estimates, standard errors, and p-values for the breakpoints in February 1988, January 1989, and May 1999.

4.7 Summary and Comments

This chapter estimated Northern Hemispheric SCE trends over the last 54 (winter-centered) years. A flexible model was developed to quantify trends in periodic presence/absence data and assess their uncertainty margins. The SCE data were collected weekly and is count valued, taking the value of unity if snow is present and zero if snow cover is absent. The data is periodic, with snow being more prevalent in the winter weeks. A contribution of this chapter is the development of a model that adequately captures the data's periodicities and count structure. Uncertainty margins of the trend estimates were developed. The model is highly flexible and could be fitted to most cells in Europe, North America, and Asia that report snow. In the most of the contiguous United States, trends could be reliably assessed down to latitudes of Prescott, AZ, Carlsbad, NM, and Knoxville, TN (the exception being some questionable SCE data from cells in mountainous area).

The results show that snow cover is declining overall, by a margin of almost 2 to 1 in terms of cell numbers. Arctic localities are showing heavy snow cover loss; however, some regions are experiencing increasing snow coverage, most notably Central and Eastern Canada and the Kamchatka and Japan vicinity. Along with this general

decline, a shift in the snow season towards an earlier onset and an earlier ablation period was seen, with the onset trending toward more snow in November and the ablation period showing declines from February through late Spring and early Summer. The increased ablation in the warm season is not offset by the increased snow cover in the late Fall, possibly implying an overall change in the timing and distribution of water availability to regions that rely on spring snowmelt.

Statistical improvements can be made to this analysis. There is undoubtedly some non-zero spatial correlation between neighboring cells. Accounting for spatial correlation would potentially lower uncertainty margins in the trend estimates; correlation usually does not appreciably change trend estimates, but accounting for correlation in multiple similar cells could reduce uncertainty margins in the trends. Given the data quality issues present, the authors felt it more prudent to analyze the cells one by one and report which ones were "unusable", which a spatial analysis would not do (at least initially). It is also possible to smooth the Figure 4.10 trends and/or their Z-scores in Figure 4.11 in a spatial manner. We did not pursue this here due to length concerns.

The reader may note that our trend estimates are based on the data only and do not depend on the model (as it should be). This said, one can also extract a trend estimate from the model. One model-based trend is

$$\frac{E[S_n] - E[S_1]}{n-1}. \tag{4.7}$$

Both $E[S_n]$ and $E[S_1]$ are computed from the estimated model parameters, say com-

puted ignoring the breakpoint. Figure 4.15 shows a plot of these trends, converted to days of SCE gained/lost per decade. The graphic naturally resembles Figure 4.10. Differences in the estimated and modeled trends are shown in Figure 4.15 and are very small overall.

While most cells report what appears to be high quality data, the green-colored cells in Figure 4.4 contain suspect data. The hope is that the data from these cells can be reexamined/fixed in the future for inclusion in studies such as this.

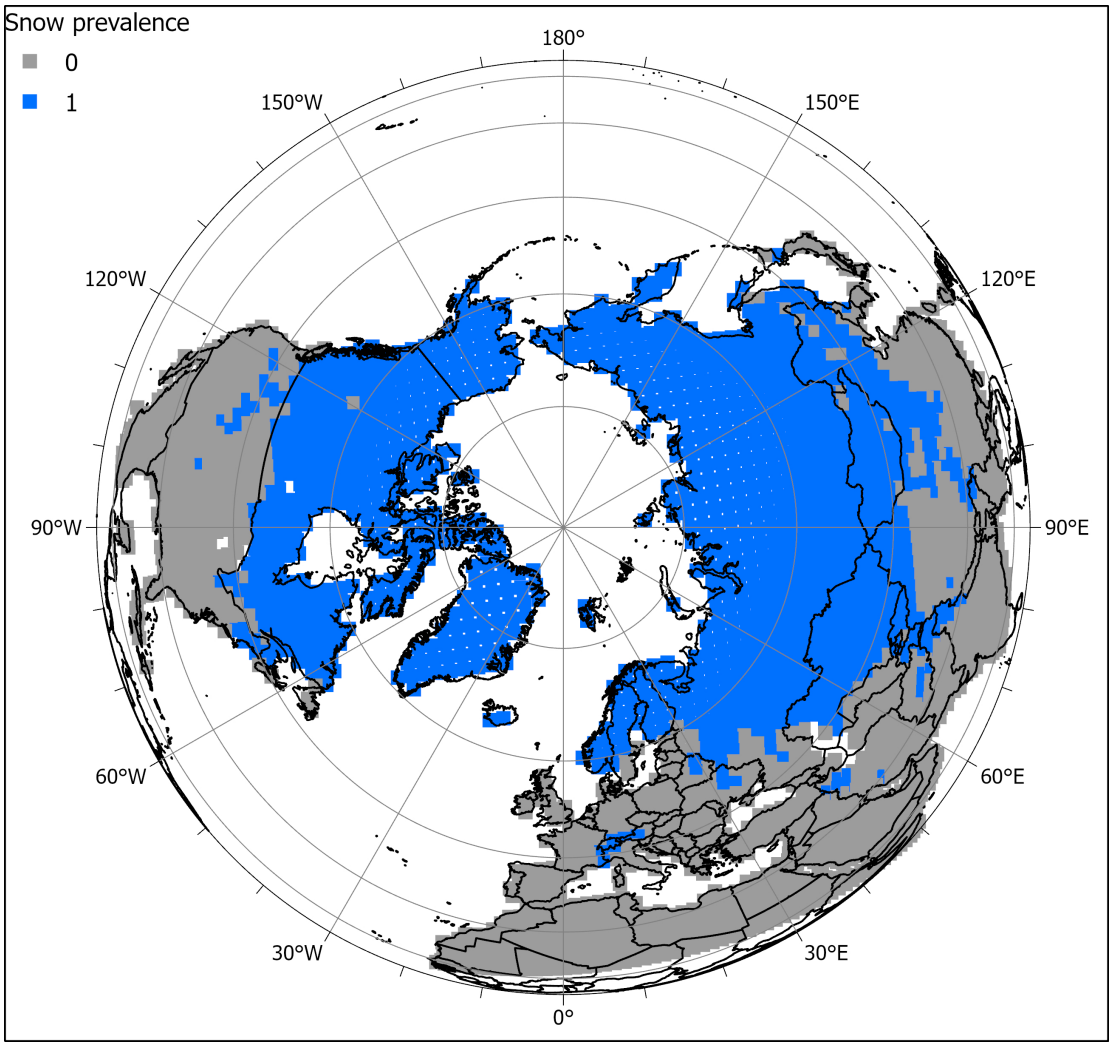

Figure 4.1: NH snow coverage reported by the NH Weekly Visible Satellite Charts (Robinson et al., 2012) for the week of December 1-7, 2020.

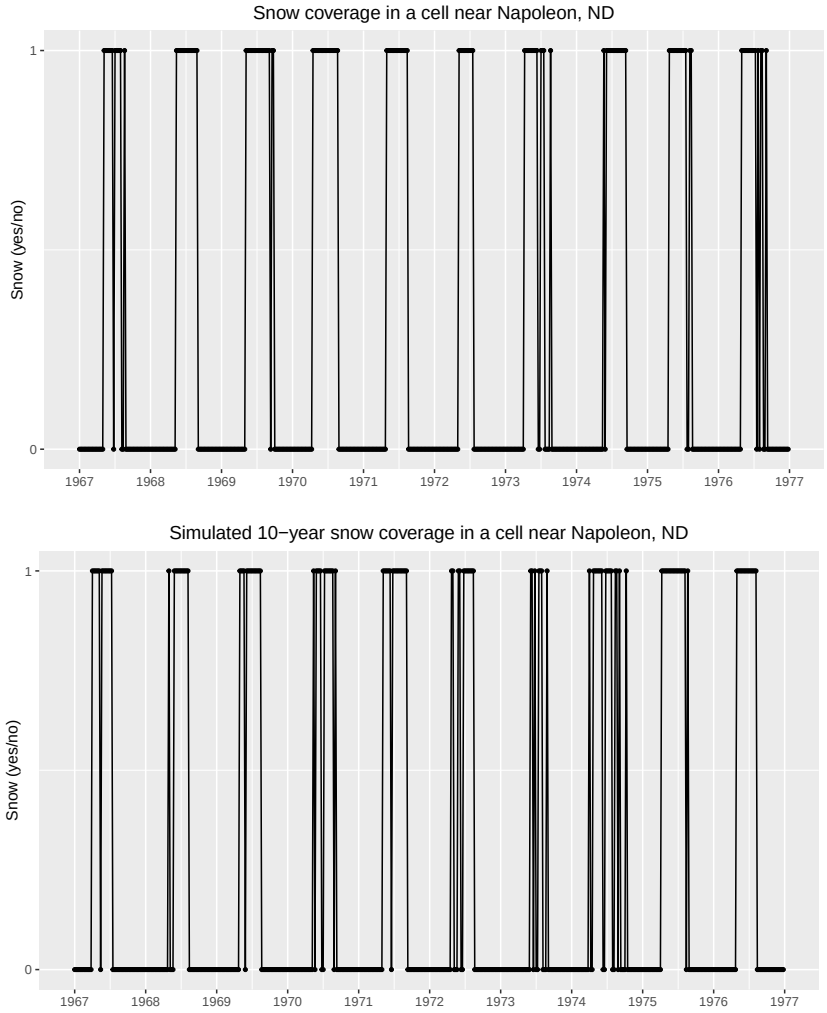

Figure 4.2: Top: ten years of snow presences/absences (August 1967 - July 1976) for a cell near Napoleon, ND (46.4309°N, 99.8852°W). Bottom: ten years (August 1967-July 1976) of simulated data. This simulation is discussed in Section 5 below. In both graphics, the yearly tickmarks refer to August 1^{st} of each calendar year, employing a winter centered year paradigm.

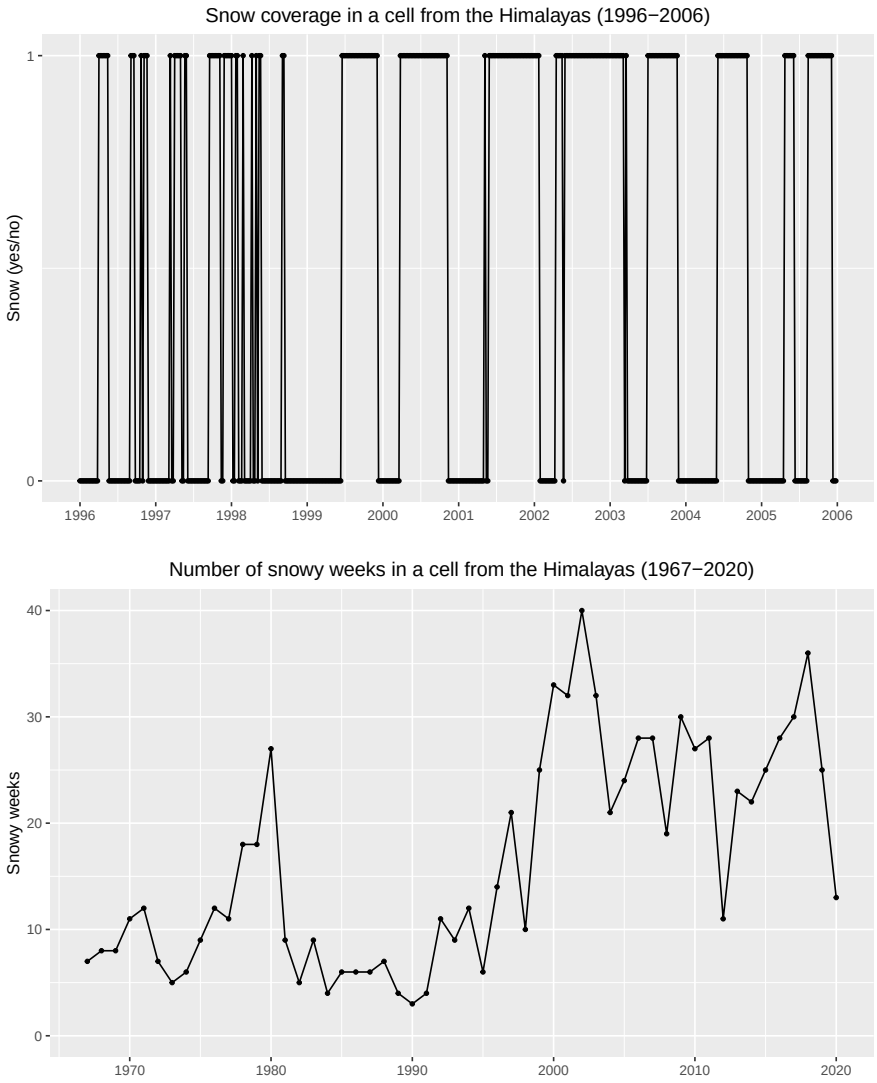

Figure 4.3: A cell from the Himalayas (27.9682°N, 97.7094°E) with untrustable data. Top: Ten years of snow presence/absence from August 1996 - July 2006. Tickmarks are placed at August 1 of each calendar year. Bottom: The number of snow covered weeks during the 1967-2020 period. Tickmarks are placed at August 1 of each calendar year.

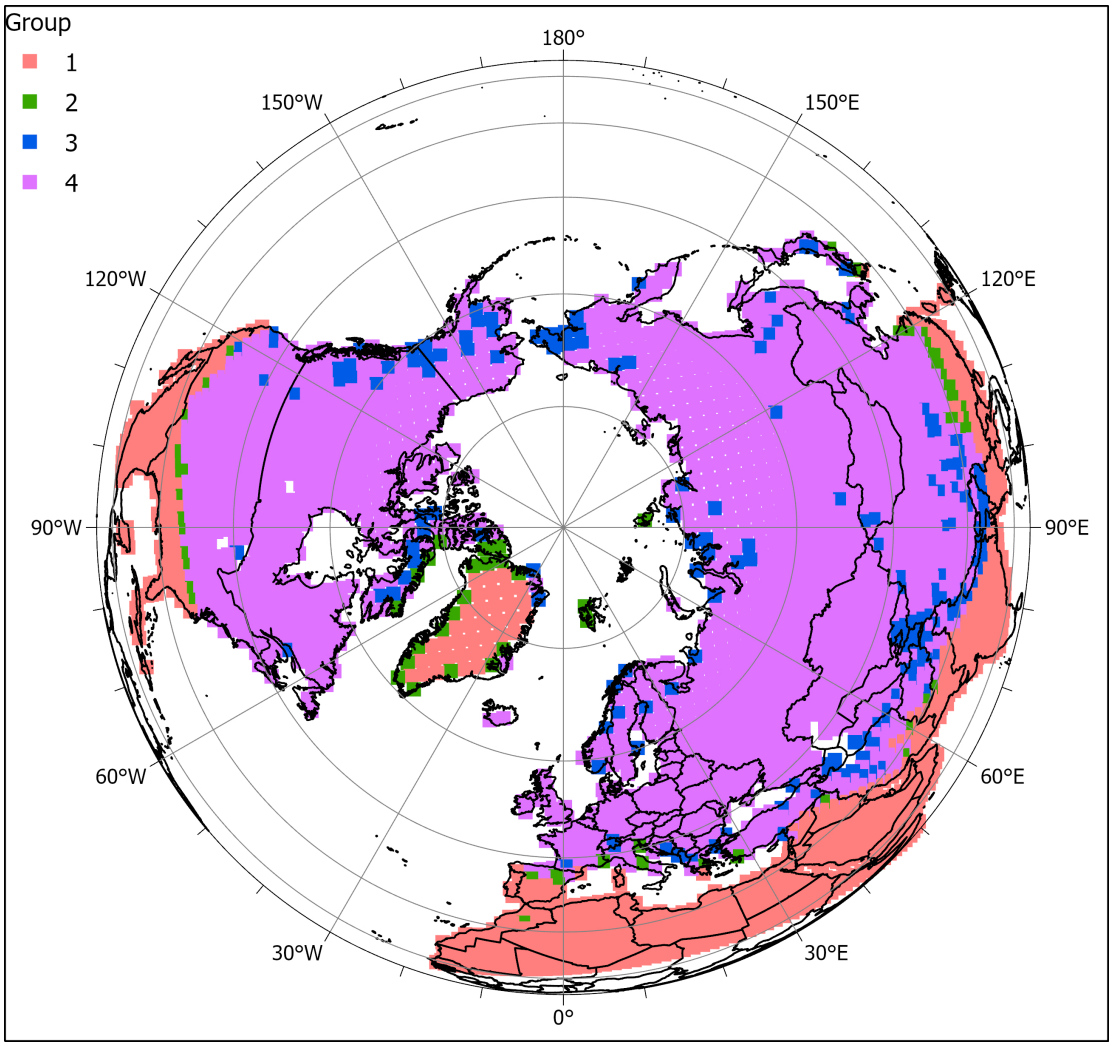

Figure 4.4: A graphical partition of this studies' cell groups. The violet colored cells (Group 4) were deemed analyzable. Group 1 cells are excluded because there are not enough changes from presence to absence (or visa versa) to fit our model. Group 3 cells were excluded as their data were deemed unreliable by our quality control methods, which agrees with the findings of other authors. Group 2 contains a small number cells whereby the standard errors of the trend estimates are so large as to make any trend estimates untrustable.

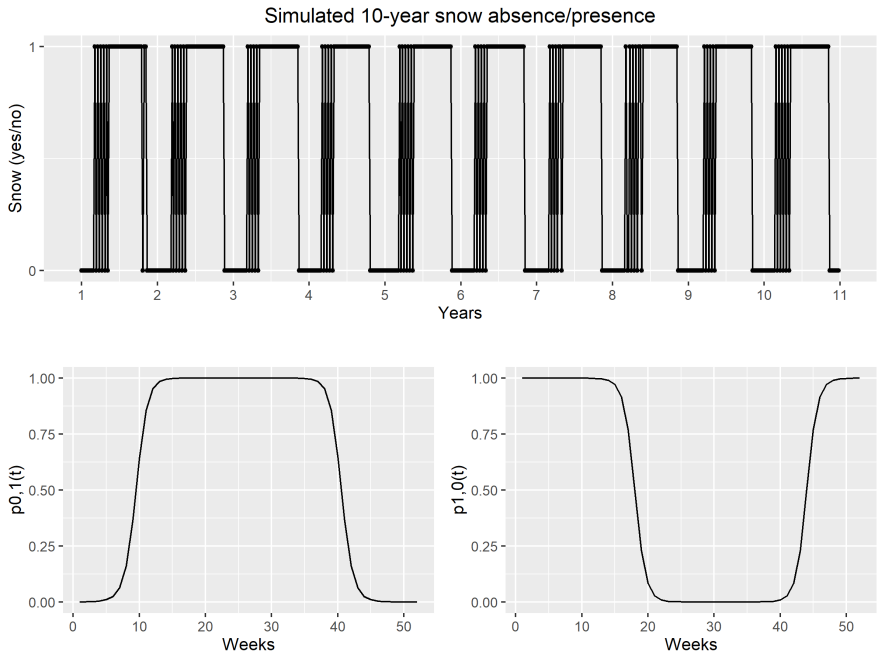

Figure 4.5: A simulated 10-year snow absence/presence series with plots of the transition probabilities $p_{0,1}(t)$ and $p_{1,0}(t)$. The parameters are $A_0 = 3, A_1 = 10, \tau = 25, \alpha = 0$, and $A_0^* = 0, A_1^* = 10, \tau^* = 5, \alpha^* = 0$ (no trend).

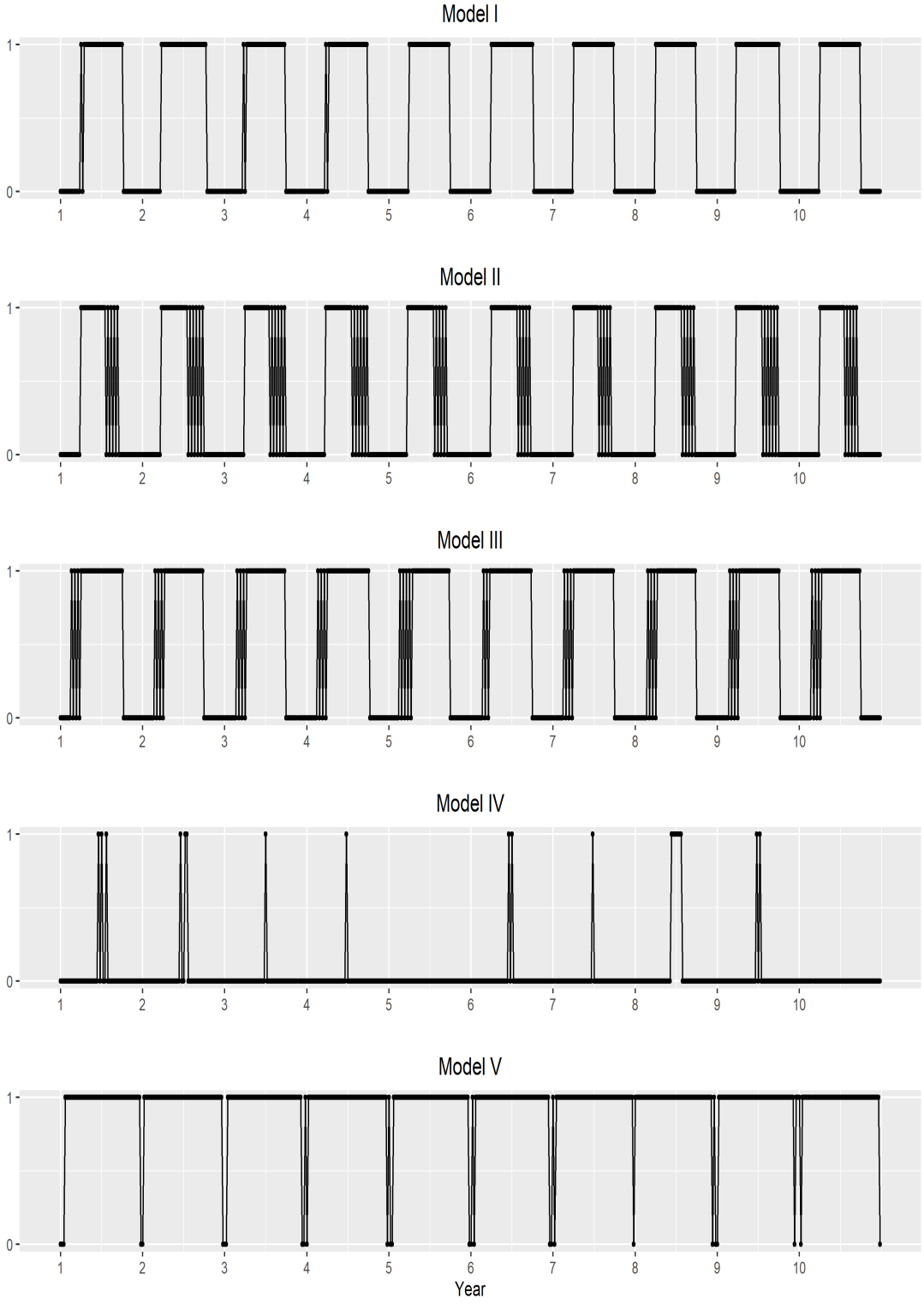

Figure 4.6: Ten year sample SCE series generated from Models I-V.

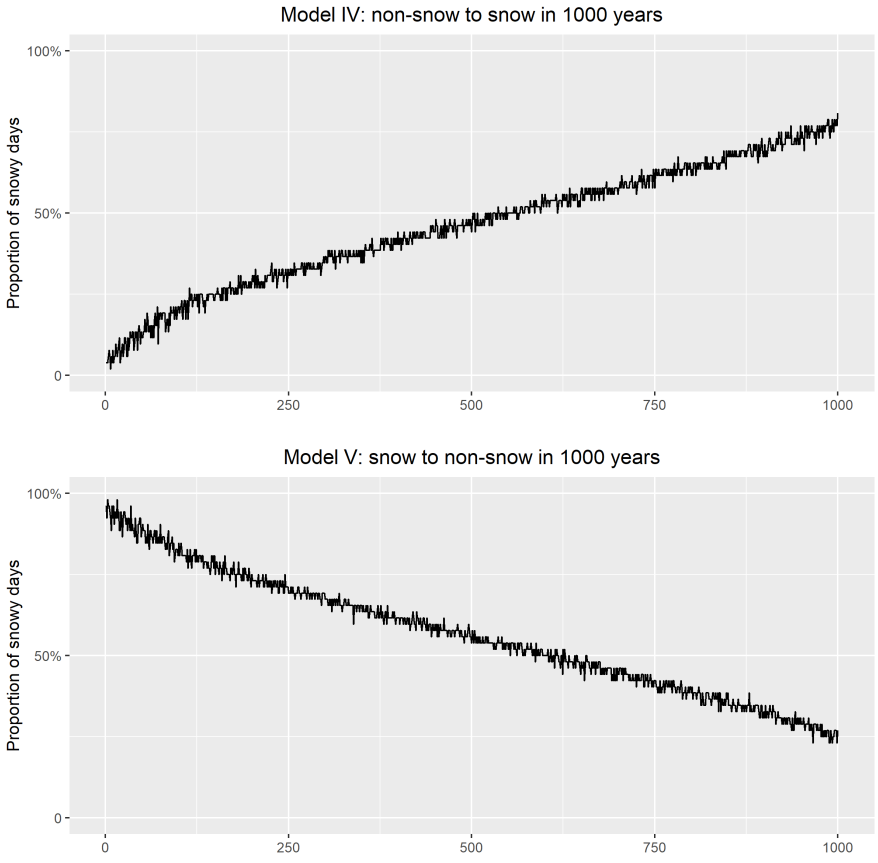

Figure 4.7: Annual proportions of snowy days from Models IV and V with non-zero trends.

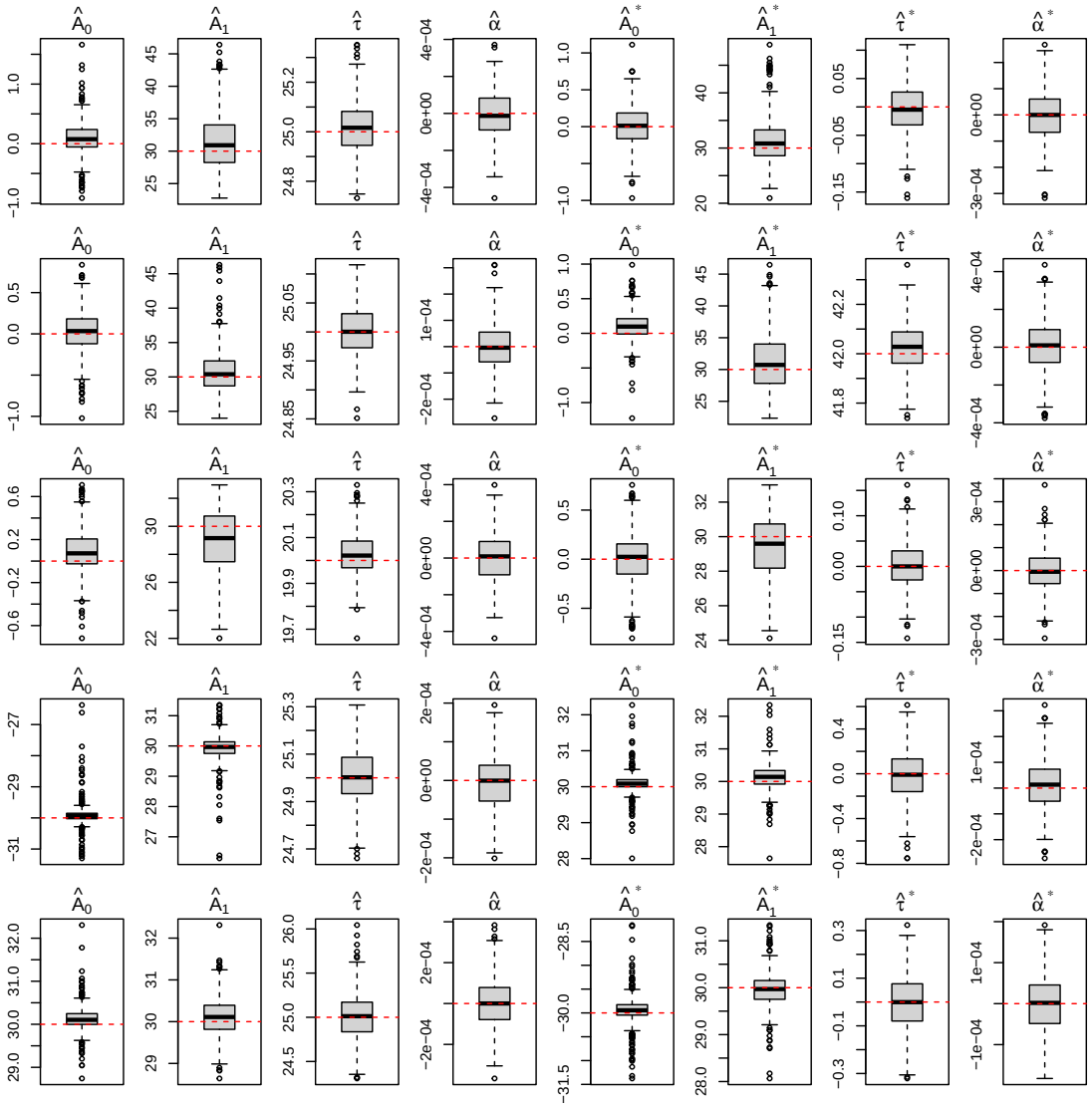

Figure 4.8: Boxplots of the parameter estimates from 1000 independent simulations. The red lines demarcate the true parameter values. Rows 1 to 5 denote estimators obtained from Model I - V, respectively.

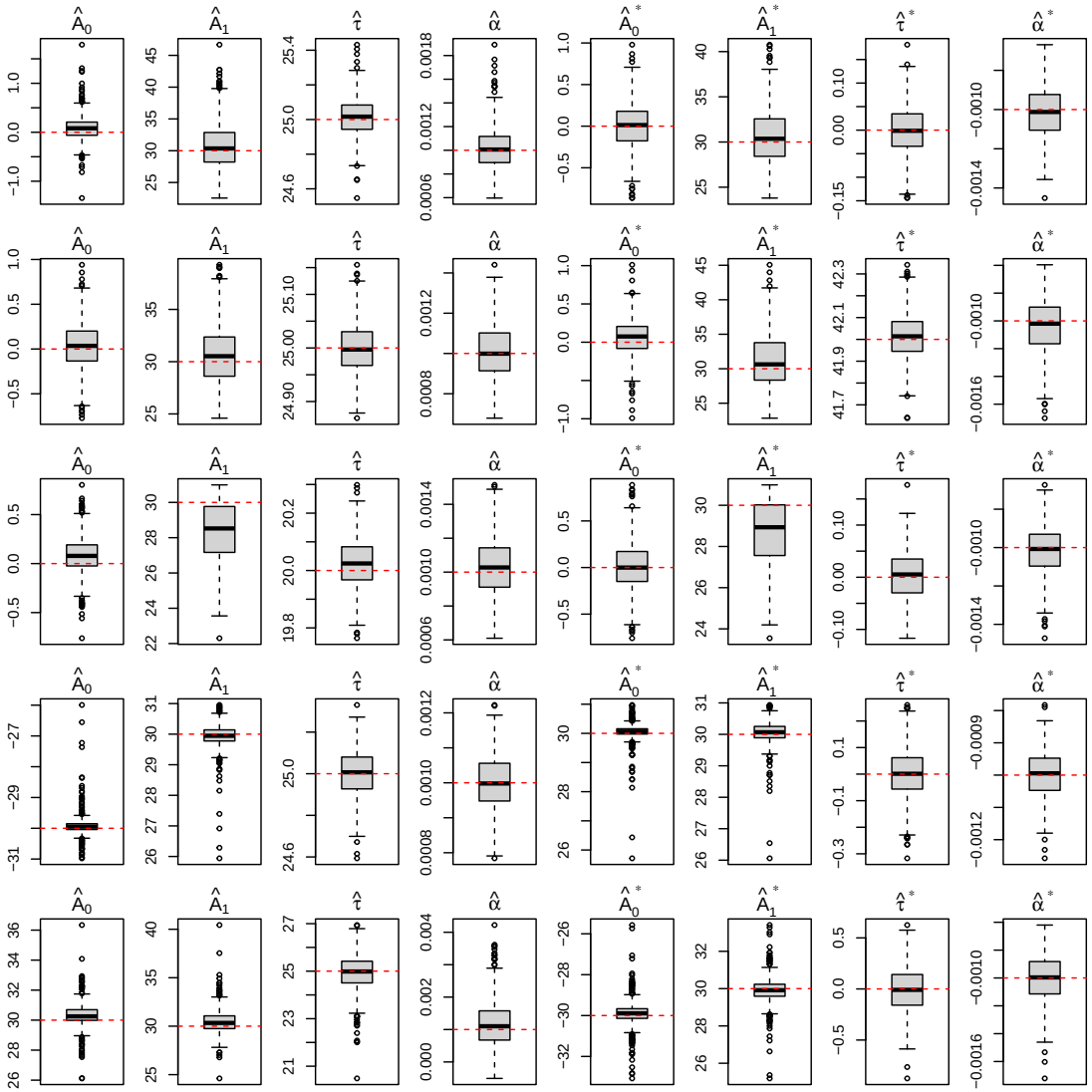

Figure 4.9: Boxplots of the parameter estimates aggregated from 1000 independent simulations. The red lines demarcate the true parameter values. Rows 1 to 5 denote estimators obtained from Model I - V with trend parameters $\alpha = 0.001$ and $\alpha^* = -0.001$.

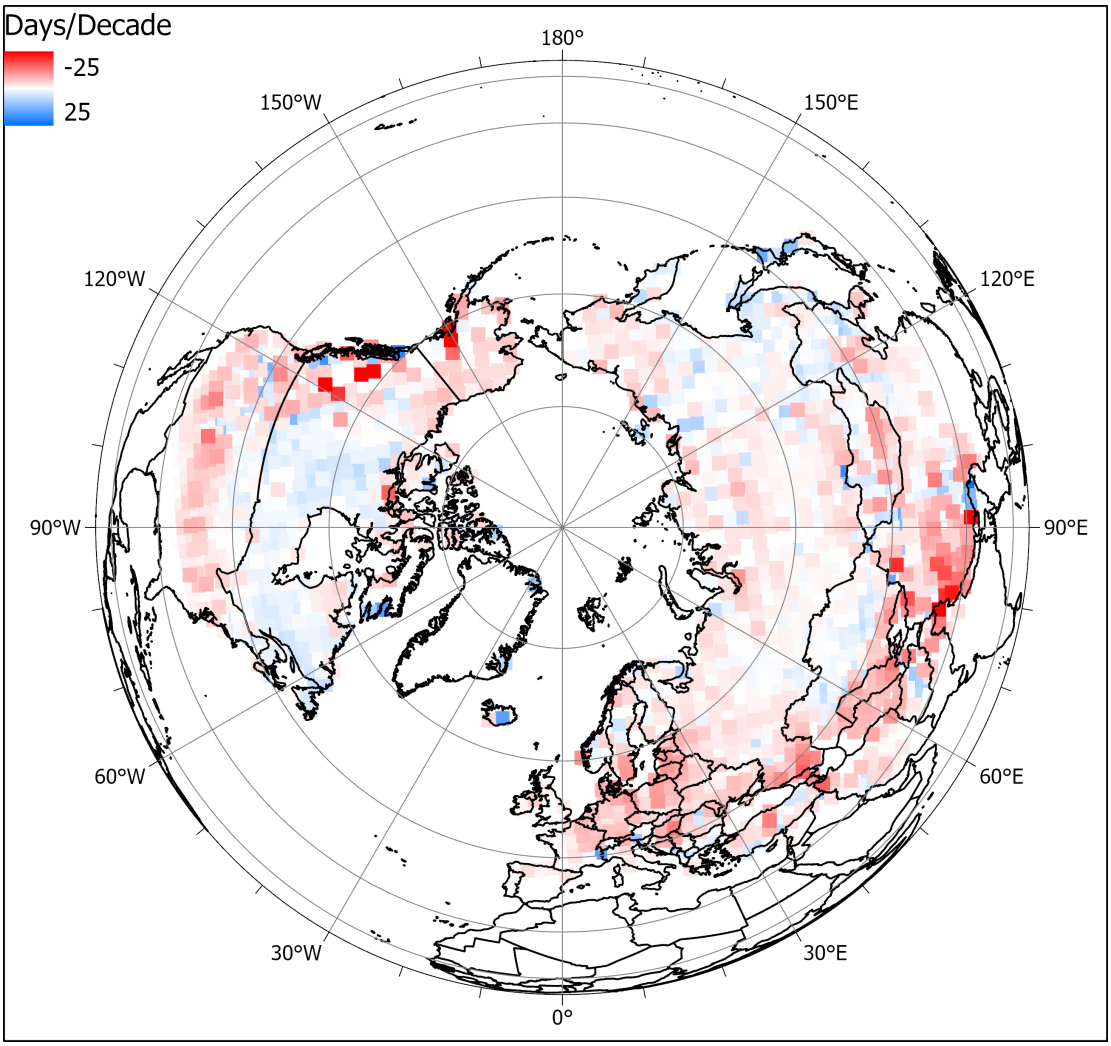

Figure 4.10: Raw trends in the SCE data converted to days gained/lost per decade. Red and blue depict SCE losses and increases, respectively. Declining SCE cells outnumber advancing SCE cells by roughly a two to one ratio.

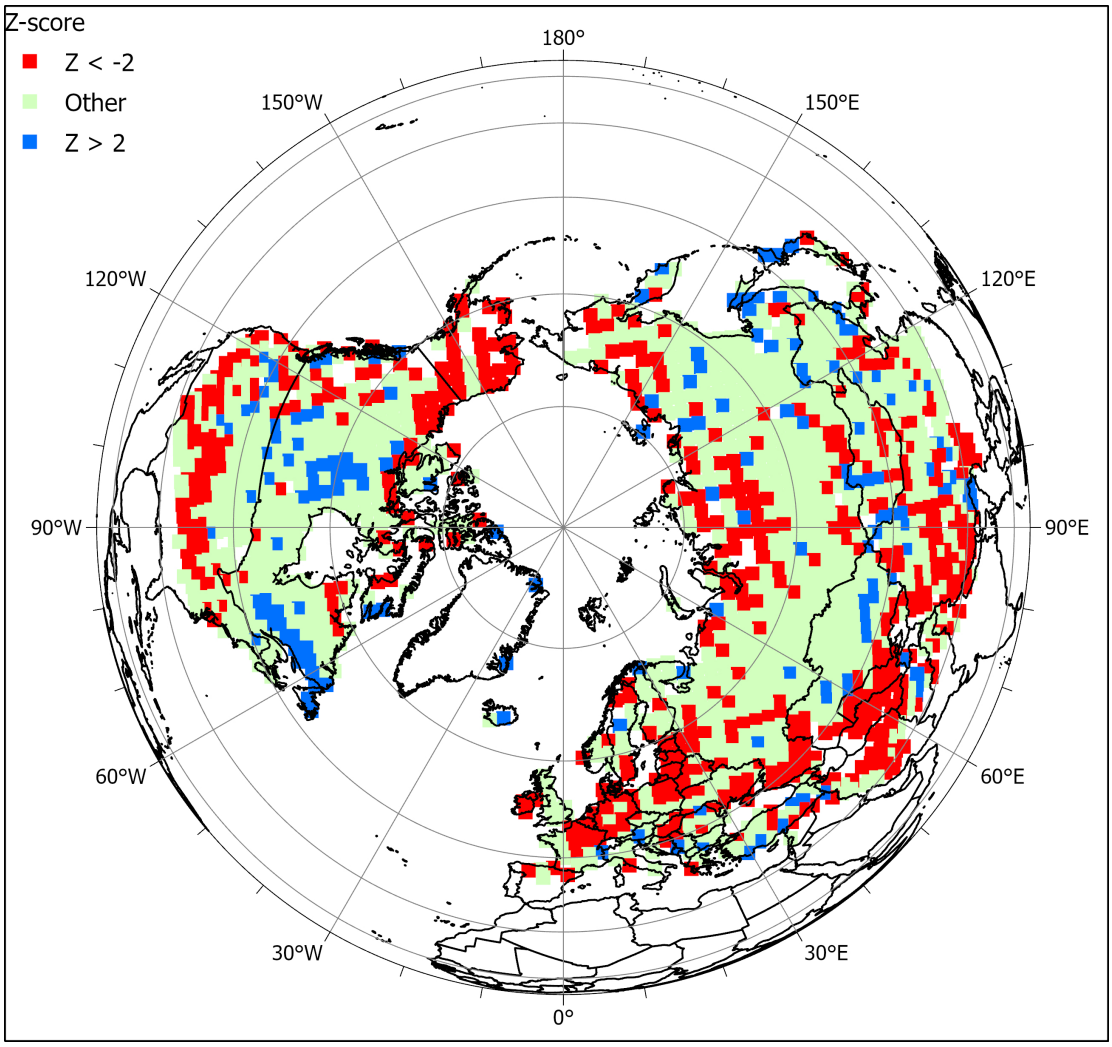

Figure 4.11: Z scores of the SCE trends. Trends in around half of the cells are not significantly changing (non-zero). Red indicates declining SCE and blue increasing SCE, with one-sided confidence of at least 97.5%.

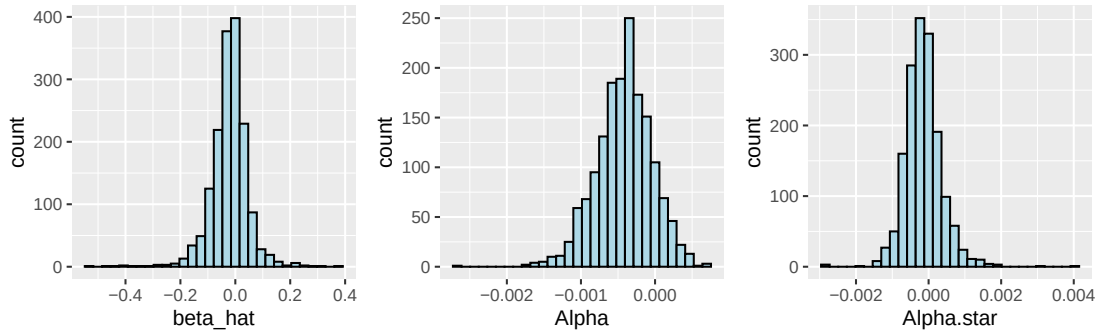

Figure 4.12: Histograms over all 1,618 analyzed cells of (left) the estimated SCE trends $\hat{\beta}$, (center) the $\hat{\alpha}$ estimates, and (right) the $\hat{\alpha}^*$ estimates. All histograms appear roughly unimodal (normally distributed). The mean of the left histogram is slightly negative.

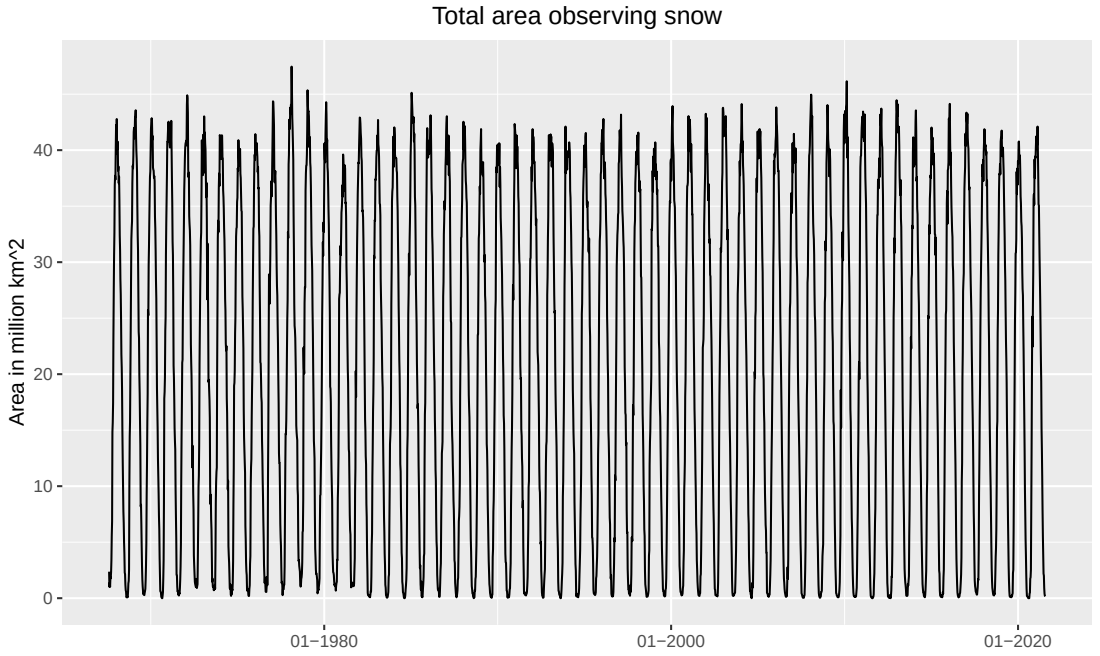

Figure 4.13: Total SCE area by week over the period of record. Trends are not visually obvious.

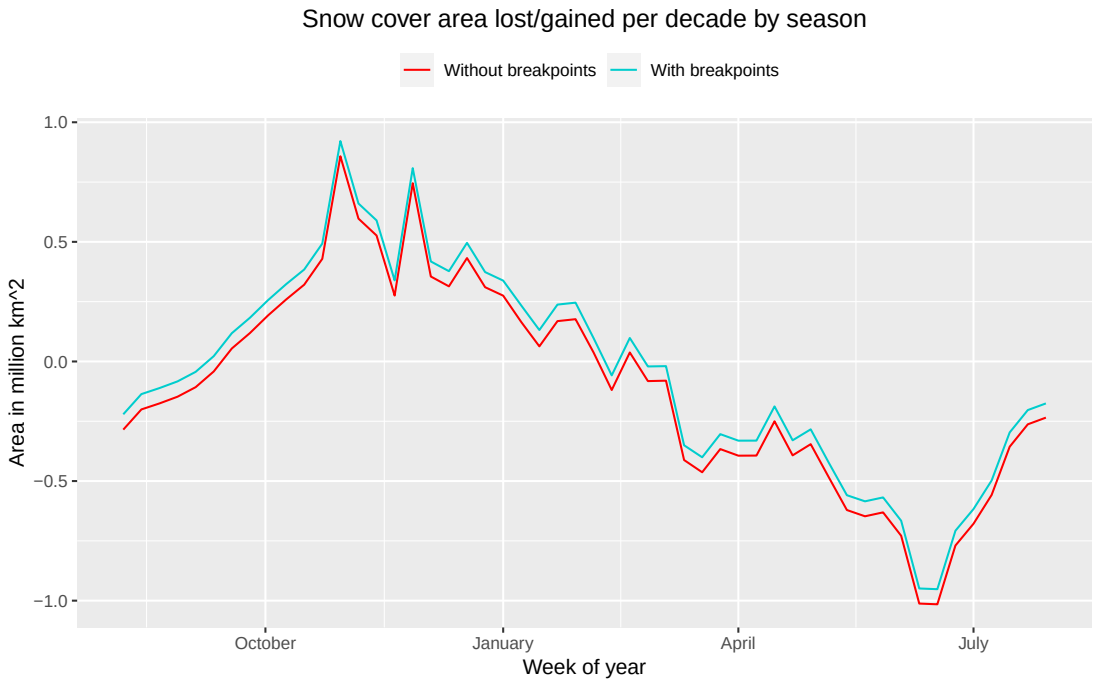

Figure 4.14: Seasonal trend estimates of SCE changes for each week of the year, scaled to area/gained lost per decade. Trends are larger when the Feb 1988, Jan 1989, and May 1999 breakpoints for method changes to extract the SCE data are considered.

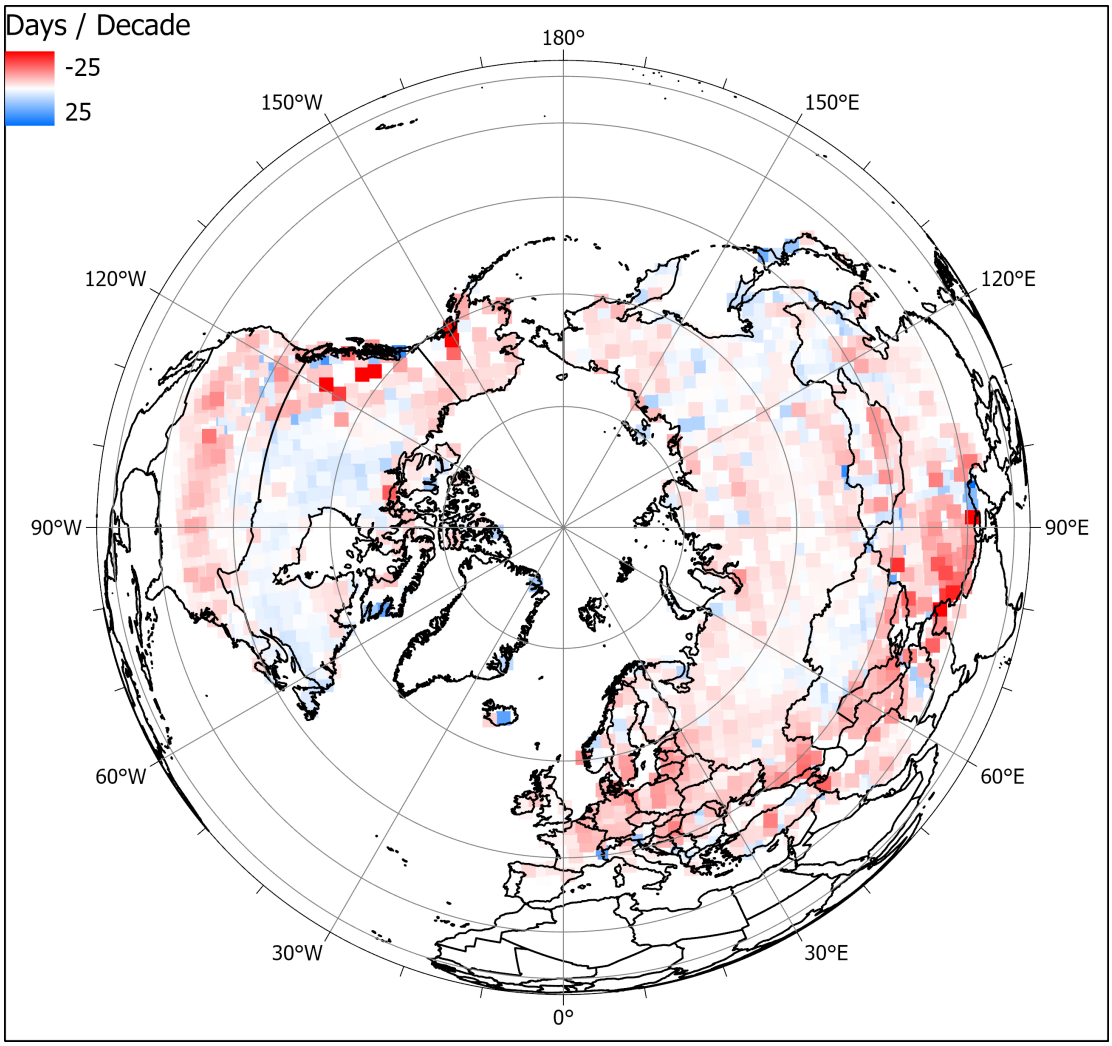

Figure 4.15: Model-based trends estimated via (4.7) and converted to days gained/lost per decade. Red and blue depict SCE losses and increases, respectively. Declining SCE cells outnumber advancing SCE cells by roughly a two to one ratio. The graphic is similar to Figure 4.10.

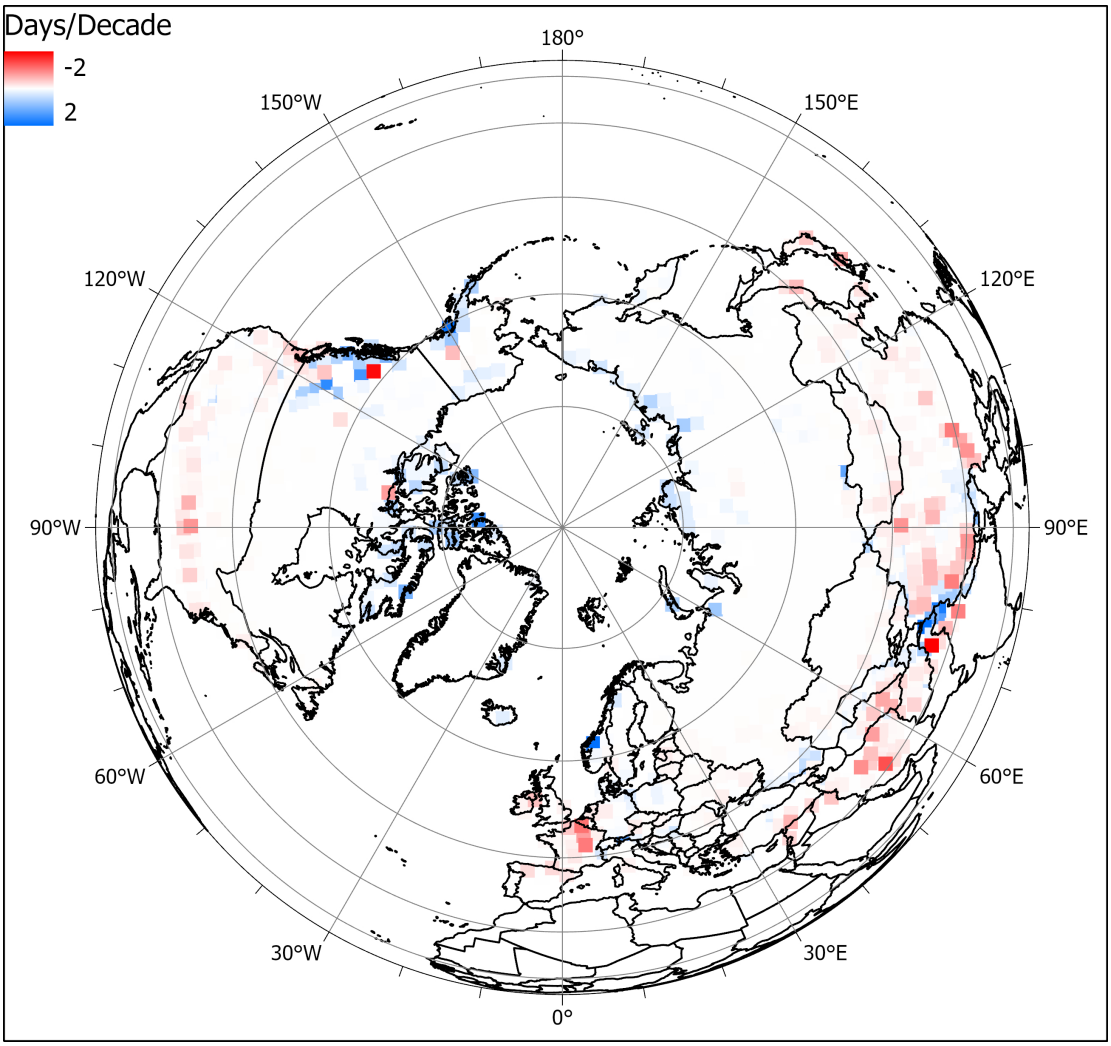

Figure 4.16: Raw trend (Figure 4.3) and the model-based trend (Figure 4.15) differences. The differences are small in most of the cells.

Appendix A

Appendix

Calculations in Example 3.2 The matrix on the left hand side of (2.7) can be rewritten as

$$\mathbf{I}_{T+1} - \phi \mathbf{P}_{T+1}, \tag{A.1}$$

where $\mathbf{P}_{T+1}$ is the $(T+1) \times (T+1)$ permutation matrix

$$\mathbf{P}_{T+1} = \mathbf{P}_{T+1}^{-1} = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 1 & \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \end{pmatrix} \tag{A.2}$$

and $\mathbf{I}_{T+1}$ is the $(T+1) \times (T+1)$ identity matrix. The inverse of the matrix in (A.1)

can be verified to be

$$[\mathbf{I}_{T+1} - \phi\mathbf{P}_{T+1}]^{-1} = \frac{1}{1-\phi^2}\left[\mathbf{I}_{T+1} + \phi\mathbf{P}_{T+1}\right].$$

Hence, the Yule-Walker equations (2.5) and (3.12) can be rewritten as

$$\begin{bmatrix} \rho_Z(0) \\ \vdots \\ \rho_Z(h) \\ \vdots \\ \rho_Z(T) \end{bmatrix} = \frac{1}{1-\phi^2}\left[\mathbf{I_{T+1}} + \phi\mathbf{P_{T+1}}\right] \begin{bmatrix} \frac{\alpha^0\sigma_\epsilon^2}{(1-\alpha^2)(1-\phi\alpha^T)} \\ \vdots \\ \frac{\alpha^h\sigma_\epsilon^2}{(1-\alpha^2)(1-\phi\alpha^T)} \\ \vdots \\ \frac{\alpha^T\sigma_\epsilon^2}{(1-\alpha^2)(1-\phi\alpha^T)} \end{bmatrix}. \tag{A.3}$$

Thus,

$$\begin{aligned} \rho_Z(h) &= \frac{1}{1-\phi^2}\left[\frac{\alpha^h\sigma_\epsilon^2}{(1-\alpha^2)(1-\phi\alpha^T)} + \phi\frac{\alpha^{T-h}\sigma_\epsilon^2}{(1-\alpha^2)(1-\phi\alpha^T)}\right] \\ &= \frac{\sigma_\epsilon^2}{(1-\phi^2)(1-\alpha^2)(1-\phi\alpha^T)}\left[\alpha^h + \phi\alpha^{T-h}\right], \quad \text{for } 0 \le h \le T. \end{aligned} \tag{A.4}$$

In our case, equation (2.6) shows that

$$\sigma_\epsilon^2 = \frac{(1-\phi^2)(1-\alpha^2)(1-\phi\alpha^T)}{1+\phi\alpha^T},$$

which gives $\mathrm{Var}(Z_t) \equiv 1$. Substituting this back into equation (A.4) gives

$$\rho_Z(h) = \frac{\alpha^h + \phi\alpha^{T-h}}{1+\phi\alpha^T}, \quad 0 \le h \le T. \tag{A.5}$$

For $h > T$, define $a = \lfloor h/T \rfloor$ and $b = h - aT$. Then Z_t can be rewritten as

$$Z_t = \phi^a Z_{t-aT} + \sum_{k=0}^{a-1} \phi^k \eta_{t-kT} \tag{A.6}$$

and hence

$$\begin{aligned} \text{Cov}(Z_t, Z_{t-h}) &= \text{Cov}(\phi^a Z_{t-aT} + \sum_{k=0}^{a-1} \phi^k \eta_{t-kT}, Z_{t-h}) \\ &= \phi^a Cov(Z_{t-aT}, Z_{t-h}) + \sum_{k=0}^{a-1} \phi^k \text{Cov}(\eta_{t-kT}, Z_{t-h}). \end{aligned}$$

Since the index difference between Z_{t-aT} and Z_{t-h} is less than T, (A.5) applies; using this and (2.5) and (2.6) produces

$$\rho_Z(h) = \phi^a \frac{\alpha^b + \phi\alpha^{T-b}}{1 + \phi\alpha^T} + \sum_{k=0}^{a-1} \phi^k \alpha^{h-Tk} \frac{1-\alpha^2}{1+\phi\alpha}, \quad h > T. \tag{A.7}$$

as claimed.

Proof of Proposition 3.1: We follow similar reasoning to Pipiras and Taqqu (2017) and Jia et al. (2023). We begin with a generalization of the Price Theorem (Theorem 5.8.5 in Pipiras and Taqqu (2017)), stated as follows and easily proven. Let G_{ν_1} and G_{ν_2} be two continuous differentiable functions. Then their link function has a derivative with form

$$L'(u) = \frac{1}{\sqrt{\text{Var}(X_1)\text{Var}(X_2)}} E[G'_{\nu_1}(Z_1) G'_{\nu_2}(Z_2)] \Big|_{\text{Corr}(Z_1, Z_2) = u}. \tag{A.8}$$

Here, Z_1 and Z_2 are a correlated Gaussian pair, each component standardized, and with $\mathrm{Corr}(Z_1, Z_2) = u$.

In our application, G_{ν_1} and G_{ν_2} are non-negative and non-decreasing since they are cumulative distribution functions. But because our data are counts, G_{ν_1} and G_{ν_2} are step functions and not necessarily differentiable on the integers. To remedy this, we approximate G_{ν_1} and G_{ν_2} by differentiable functions and take limits in the approximation.

To do this, let $U \stackrel{\mathcal{N}}{=} \mathcal{N}(0,1)$. For any $\epsilon > 0$ and $\ell \in \{1, 2\}$,

$$\begin{aligned} G_{\epsilon,\nu_\ell}(x) &:= E[G_{\nu_\ell}(x + \epsilon U)] = \int_{-\infty}^{\infty} G_{\nu_\ell}(z) \frac{e^{-\frac{(x-z)^2}{2\epsilon^2}}}{\sqrt{2\pi}\epsilon} dz \\ &= \sum_{j=0}^{\infty} j \int_{\Phi^{-1}(C_{j-1}(\nu_\ell))}^{\Phi^{-1}(C_j(\nu_\ell))} \frac{e^{-\frac{(x-z)^2}{2\epsilon^2}}}{\sqrt{2\pi}\epsilon} dz \\ &= \sum_{j=0}^{\infty} j \int_{\Phi^{-1}(C_{j-1}(\nu_\ell))-x}^{\Phi^{-1}(C_j(\nu_\ell))-x} \frac{e^{-\frac{w^2}{2\epsilon^2}}}{\sqrt{2\pi}\epsilon} dw. \end{aligned} \tag{A.9}$$

The "kernel"

$$\frac{e^{-\frac{(x-z)^2}{2\epsilon^2}}}{\sqrt{2\pi}\epsilon} \tag{A.10}$$

acts like Dirac's delta function $\delta_{\{x\}}(z)$ at $z = x$ as $\epsilon \downarrow 0$. Note that $G_{\epsilon,\nu_\ell}(x)$ is non-decreasing and differentiable with first derivative

$$G'_{\epsilon,\nu_\ell}(x) = \frac{1}{\sqrt{2\pi}\epsilon} \sum_{j=0}^{\infty} j \Big[e^{-\frac{(\Phi^{-1}(C_{j-1}(\nu_\ell))-x)^2}{2\epsilon^2}} - e^{-\frac{(\Phi^{-1}(C_j(\nu_\ell))-x)^2}{2\epsilon^2}} \Big] = \frac{1}{\sqrt{2\pi}\epsilon} \sum_{j=0}^{\infty} e^{-\frac{(\Phi^{-1}(C_j(\nu_\ell))-x)^2}{2\epsilon^2}}, \tag{A.11}$$

and define $X_\ell^{(\epsilon)} = G_{\epsilon,\nu_\ell}(Z_\ell)$ for $\ell \in \{1,2\}$. Equation (2.16) gives

$$\begin{aligned}
L'_\epsilon(u) &= \frac{1}{\sqrt{\mathrm{Var}(X_1^{(\epsilon)})\mathrm{Var}(X_2^{(\epsilon)})}} E[G'_{\epsilon,\nu_1}(Z_1)G'_{\epsilon,\nu_2}(Z_2)]\Big|_{\mathrm{Corr}(Z_1,Z_2)=u} \\
&= \frac{1}{\sqrt{\mathrm{Var}(X_1^{(\epsilon)})\mathrm{Var}(X_2^{(\epsilon)})}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G'_{\epsilon,\nu_1}(Z_1)G'_{\epsilon,\nu_2}(Z_2) \\
&\qquad \times \frac{1}{2\pi\sqrt{1-u^2}} e^{-\frac{1}{2(1-u^2)}\left(z_1^2+z_2^2-2uz_1z_2\right)} dz_1 dz_2 \\
&= \frac{1}{\sqrt{\mathrm{Var}(X_1^{(\epsilon)})\mathrm{Var}(X_2^{(\epsilon)})}} \sum_{j_1=0}^{\infty}\sum_{j_2=0}^{\infty} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\epsilon} e^{-\frac{(\Phi^{-1}(C_{j_1}(\nu_1))-z_1)^2}{2\epsilon^2}} \\
&\qquad \times \frac{1}{\sqrt{2\pi}\epsilon} e^{-\frac{(\Phi^{-1}(C_{j_2}(\nu_2))-z_2)^2}{2\epsilon^2}} \frac{1}{2\pi\sqrt{1-u^2}} e^{-\frac{1}{2(1-u^2)}\left(z_1^2+z_2^2-2uz_1z_2\right)} dz_1 dz_2.
\end{aligned}$$

Noting again that the quantity in (A.10) acts like a Dirac's delta function $\delta_{\{x\}}(z)$, the limit as $\epsilon \downarrow 0$ should be

$$\begin{aligned}
L'(u) = \frac{1}{\sqrt{\mathrm{Var}(X_1)\mathrm{Var}(X_2)}} \sum_{j_1=0}^{\infty}\sum_{j_2=0}^{\infty} \frac{1}{2\pi\sqrt{1-u^2}} \times \\
e^{-\frac{1}{2(1-u^2)}\left(\Phi^{-1}(C_{j_1}(\nu_1))^2+\Phi^{-1}(C_{j_2}(\nu_2))^2-2u\Phi^{-1}(C_{j_1}(\nu_1))\Phi^{-1}(C_{j_2}(\nu_2))\right)}, \quad \text{(A.12)}
\end{aligned}$$

which is always non-negative. The existence and form of $L'(u)$ stems from the fact that we are differentiating a power series with absolutely convergent coefficients inside its radius of convergence. That $\sum_{k=0}^{\infty} |\ell_k| < \infty$ follows from (2.15), the Cauchy-Schwarz inequality, and the finiteness of $\sum_{k=0}^{\infty} k! g_k(\nu_1)^2$ and $\sum_{k=0}^{\infty} k! g_k(\nu_2)^2$.

We now show that $L'_\epsilon(u)$ converges to $L'(u)$. For this, we first need an expression for the Hermite coefficients of $G_{\epsilon,\nu_\ell}(\cdot)$, denoted by $g_{\epsilon,k}(\nu_\ell)$ for $\ell \in \{1,2\}$. These will be compared to the Hermite coefficients $g_k(\nu_\ell)$ of G_{ν_ℓ}.

Taylor expanding the Hermite polynomial $H_k(x+y) = \sum_{d=0}^{k} \binom{k}{d} y^{k-d} H_d(x)$ implies

$$\begin{aligned} G_{\epsilon,\nu_\ell}(x) &= E[G_{\nu_\ell}(x+\epsilon U)] = E\left[\sum_{k=0}^{\infty} g_k(\nu_\ell) H_k(x+\epsilon U)\right] \\ &= E\left[\sum_{k=0}^{\infty} g_k(\nu_\ell) \sum_{d=0}^{k} \binom{k}{d} (\epsilon U)^{k-d} H_d(x)\right] \\ &= \sum_{d=0}^{\infty} H_d(x) \sum_{k=d}^{\infty} g_k(\nu_\ell) \epsilon^{k-d} \binom{k}{d} E[U^{k-d}]. \end{aligned}$$

After changing summation indices and using that $E[U^p] = 0$ if p is odd, and equal to $(p-1)!!$ if p is even, where $k!! = 1 \times 3 \times \cdots \times k$ when k is odd, we get

$$g_{\epsilon,k}(\nu_\ell) = g_k(\nu_\ell) + \sum_{q=1}^{\infty} g_{k+2q}(\nu_\ell) \epsilon^{2q} \binom{k+2q}{k} (2q-1)!! = g_k(\nu_\ell) + \sum_{q=1}^{\infty} g_{k+2q}(\nu_\ell) \epsilon^{2q} \frac{(k+2q)!}{k! 2^q q!}. \tag{A.13}$$

Then

$$\begin{aligned} &|g_k(\nu_1) g_k(\nu_2) - g_{\epsilon,k}(\nu_1) g_{\epsilon,k}(\nu_2)| \leq \\ &\qquad |g_k(\nu_1)| \sum_{q=1}^{\infty} |g_{k+2q}(\nu_1)| \epsilon^{2q} \frac{(k+2q)!}{k! 2^q q!} + |g_k(\nu_2)| \sum_{q=1}^{\infty} |g_{k+2q}(\nu_2)| \epsilon^{2q} \frac{(k+2q)!}{k! 2^q q!} \\ &\qquad + \Big(\sum_{q=1}^{\infty} |g_{k+2q}(\nu_1)| \epsilon^{2q} \frac{(k+2q)!}{k! 2^q q!} \Big) \Big(\sum_{q=1}^{\infty} |g_{k+2q}(\nu_2)| \epsilon^{2q} \frac{(k+2q)!}{k! 2^q q!} \Big). \end{aligned} \tag{A.14}$$

Use the Cauchy-Schwarz inequality to obtain the bounds

$$\sum_{q=1}^{\infty} |g_{k+2q}(\nu_\ell)| \epsilon^{2q} \frac{(k+2q)!}{k! 2^q q!} \leq \left(\sum_{q=1}^{\infty} g_{k+2q}^2(\nu_\ell) (k+2q)! \right)^{1/2} \left(\sum_{q=1}^{\infty} \epsilon^{4q} \frac{(k+2q)!}{(k!)^2 (2^q q!)^2} \right)^{1/2}$$

$$\leq \frac{M_{k,\ell}}{(k!)^{1/2}} \left(\sum_{q=1}^{\infty} \epsilon^{4q} \frac{(k+2q)!}{k!(2q)!} \right)^{1/2} \qquad \forall \ell \in \{1,2\},$$

where $M_{k,\ell}$ is some finite constant that converges to zero as $k \to \infty$. Since $\text{Var}(X_\ell) = \sum_{k=1}^{\infty} k! g_k^2(\nu_\ell)$ is finite and $(2^q q!)^2$ is of the same order as $(2q)!$, $\sum_{q=1}^{\infty} g_{k+2q}^2(\nu_\ell)(k+2q)! \to 0$ as $k \to \infty$. We use the fact that $\sum_{p=0}^{\infty} x^p \binom{k+p}{p} = (1-x)^{-k-1}$ for $|x| < 1$ to obtain a bound for $\sum_{p=1}^{\infty} \epsilon^{2p} \binom{k+p}{p}$. Then (A.14) gives

$$|g_k(\nu_1)g_k(\nu_1) - g_{\epsilon,k}(\nu_1)g_{\epsilon,k}(\nu_2)| \leq \sum_{\ell=1}^{2} \frac{M_{k,\ell}|g_k(\nu_\ell)|}{(k!)^{1/2}} \left[(1-\epsilon^2)^{-k-1} - 1 \right]^{1/2} + \frac{M_{k,1}M_{k,2}}{k!}[(1-\epsilon^2)^{-k-1} - 1]. \tag{A.15}$$

Now take the first derivative of the link function in (2.14) to obtain

$$L'(u) = \frac{1}{\sqrt{\text{Var}(X_1)\text{Var}(X_2)}} \sum_{k=1}^{\infty} g_k(\nu_1)g_k(\nu_2)k!ku^{k-1},$$

where the series converges absolutely for $u \in (-1,1)$ since the "extra" k gets dominated by u^{k-1}. Similarly,

$$L'_\epsilon(u) = \frac{1}{\sqrt{\text{Var}(X_1)\text{Var}(X_2)}} \sum_{k=1}^{\infty} g_{\epsilon,k}(\nu_1)g_{\epsilon,k}(\nu_2)k!ku^{k-1}.$$

The above expression agrees with Theorem 5.1.10 in Pipiras and Taqqu (2017). To show

that the difference between $L'_\epsilon(u)$ and $L'(u)$ converges to zero as $\epsilon \downarrow 0$, use

$$|L'(u) - L'_\epsilon(u)| \leq \Big| \frac{1}{\sqrt{\mathrm{Var}(X_1)\mathrm{Var}(X_2)}} - \frac{1}{\sqrt{\mathrm{Var}(X_1^{(\epsilon)})\mathrm{Var}(X_2^{(\epsilon)})}} \Big| \sum_{k=1}^{\infty} g_k(\nu_1) g_k(\nu_2) k! k |u|^{k-1}$$
$$+ \frac{1}{\sqrt{\mathrm{Var}(X_1^{(\epsilon)})\mathrm{Var}(X_2^{(\epsilon)})}} \sum_{k=1}^{\infty} |g_k(\nu_1) g_k(\nu_2) - g_{\epsilon,k}(\nu_1) g_{\epsilon,k}(\nu_2)| k! k |u|^{k-1}. \tag{A.16}$$

From (A.15), we see that $|g_k(\nu_1)g_k(\nu_2) - g_{\epsilon,k}(\nu_1)g_{\epsilon,k}(\nu_2)| \to 0$ as $\epsilon \downarrow 0$. Hence, $\sum_{k=1}^{\infty} |g_k(\nu_1)g_k(\nu_2) - g_{\epsilon,k}(\nu_1)g_{\epsilon,k}(\nu_2)| k! k |u|^{k-1}$ converges to zero by the dominated convergence theorem as $\epsilon \downarrow 0$. Using (2.13), we concluded that $\mathrm{Var}(X_1^{(\epsilon)}) \to \mathrm{Var}(X_1)$ and $\mathrm{Var}(X_2^{(\epsilon)}) \to \mathrm{Var}(X_2)$ as $\epsilon \downarrow 0$. Therefore,

$$\left| \frac{1}{\sqrt{\mathrm{Var}(X_1)\mathrm{Var}(X_2)}} - \frac{1}{\sqrt{\mathrm{Var}(X_1^{(\epsilon)})\mathrm{Var}(X_2^{(\epsilon)})}} \right| \to 0 \quad \text{as} \quad \epsilon \downarrow 0$$

follows by continuity of the function $x^{-1/2}$ away from $x = 0$ (the limiting variances are tacitly assumed positive to avoid degeneracy).

Bibliography

Aknouche, A., W. Bentarzi, and N. Demouche, 2018: On periodic ergodicity of a general periodic mixed Poisson autoregression. *Statistics & Probability Letters*, **134**, 15–21.

Alexandersson, H., 1986: A homogeneity test applied to precipitation data. *International Journal of Climatology*, **6**, 661–675.

Alzaid, A. A. and M. Al-Osh, 1990: An integer-valued pth-order autoregressive structure (INAR(p)) process. *Journal of Applied Probability*, **27**, no. 2, 314–324.

Anderson, P. L., Y. G. Tesfaye, and M. M. Meerschaert, 2007: Fourier-PARMA models and their application to river flows. *Journal of Hydrologic Engineering*, **12**, no. 5, 462–472.

Asmussen, S. and S. Foss, 2014: On exceedance times for some processes with dependent increments. *Journal of Applied Probability*, **51**, no. 1, 136–151.

Bai, Y., J. Kang, and P. X.-K. Song, 2014: Efficient pairwise composite likelihood estimation for spatial-clustered data. *Biometrics*, **70**, no. 3, 661–670.

Banerjee, S., B. P. Carlin, and A. E. Gelfand, 2014: *Hierarchical Modeling and Analysis for Spatial Data*. CRC press.

Barnett, T. P., J. C. Adam, and D. P. Lettenmaier, 2005: Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature*, **438**, no. 7066, 303–309.

Barreto-Souza, W. and M. Bourguignon, 2015: A skew INAR (1) process on $\mathbb{Z}$. *AStA Advances in Statistical Analysis*, **99**, no. 2, 189–208.

Basawa, I. and R. Lund, 2001: Large sample properties of parameter estimates for periodic ARMA models. *Journal of Time Series Analysis*, **22**, no. 6, 651–663.

Basawa, I., R. Lund, and Q. Shao, 2004: First-order seasonal autoregressive processes with periodically varying parameters. *Statistics & Probability Letters*, **67**, no. 4, 299–306.

Benjamin, M. A., R. A. Rigby, and D. M. Stasinopoulos, 2003: Generalized autoregressive moving average models. *Journal of the American Statistical Association*, **98**, no. 461, 214–223.

Bentarzi, M. and N. Aries, 2020: On some periodic INARMA (p, q) models. *Communications in Statistics-Simulation and Computation*, 1–21.

Bentarzi, M. and W. Bentarzi, 2017: Periodic integer-valued bilinear time series model. *Communications in Statistics-Theory and Methods*, **46**, no. 3, 1184–1201.

Bentarzi, M. and M. Hallin, 1994: On the invertibility of periodic moving-average models. *Journal of time series analysis*, **15**, no. 3, 263–268.

Bentarzi, M. and R. Souakri, 2023: On Periodic Generalized Poisson INAR (1) Model. *Communications in Statistics-Simulation and Computation*, 1–26.

Biller, B., 2009: Copula-based multivariate input models for stochastic simulation. *Operations Research*, **57**, no. 4, 878–892.

Bissell, A. F., 1969: Cusum techniques for quality control. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **18**, no. 1, 1–25.

Blight, P., 1989: Time series formed from the superposition of discrete renewal processes. *Journal of Applied Probability*, **26**, no. 1, 189–195.

Bloomfield, P., H. L. Hurd, and R. B. Lund, 1994: Periodic correlation in stratospheric ozone data. *Journal of Time Series Analysis*, **15**, no. 2, 127–150.

Bormann, K. J., R. D. Brown, C. Derksen, and T. H. Painter, 2018: Estimating snow-cover trends from space. *Nature Climate Change*, **8**, no. 11, 924–928.

Boudreault, M. and A. Charpentier, 2011: Multivariate integer-valued autoregressive models applied to earthquake counts. *arXiv preprint arXiv:1112.0929*.

Box, G. E., G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, 2015: *Time Series Analysis: Forecasting and Control*. John Wiley & Sons.

Brockwell, P. J. and R. A. Davis, 1991a: *Time Series: Theory and Methods*. 2nd ed., Springer.

— 1991b: *Time Series: Theory and Methods*. Secondnd ed., Springer, New York City.

— 2009: *Time Series: Theory and Methods*. Springer Science & Business Media.

Brown, R., C. Derksen, and L. Wang, 2007: Assessment of spring snow cover duration variability over northern Canada from satellite datasets. *Remote Sensing of Environment*, **111**, no. 2-3, 367–381.

Brown, R. and P. W. Mote, 2009: The response of Northern Hemisphere snow cover to a changing climate. *Journal of Climate*, **22**, no. 8, 2124–2145.

Brown, R. D., 2000: Northern Hemisphere snow cover variability and change, 1915–97. *Journal of Climate*, **13**, 2339–2355.

— 2010: Analysis of snow cover variability and change in Québec, 1948–2005. *Hydrological Processes*, **24**, 1929–1954.

Brown, R. D. and R. O. Braaten, 1998: Spatial and temporal variability of Canadian monthly snow depths, 1946–1995. *Atmosphere - Ocean*, **36**, 37–54.

Brown, R. D., B. Brasnett, and D. Robinson, 2003: Gridded North American monthly snow depth and snow water equivalent for GCM evaluation. *Atmosphere - Ocean*, **41**, 1–14.

Brown, R. D. and D. A. Robinson, 2011a: Northern Hemisphere spring snow cover

variability and change over 1922–2010 including an assessment of uncertainty. *The Cryosphere*, **5**, no. 1, 219–229.

— 2011b: Northern Hemisphere spring snow cover variability and change over 1922–2010 including an assessment of uncertainty. *The Cryosphere*, **5**, 219–229.

Callaghan, T. V., M. Johansson, R. D. Brown, P. Y. Groisman, N. Labba, V. Radionov, R. G. Barry, O. N. Bulygina, R. L. Essery, D. Frolov, et al., 2011: The changing face of Arctic snow cover: A synthesis of observed and projected changes. *Ambio*, **40**, no. 1, 17–31.

Cario, M. C. and B. L. Nelson, 1997: Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix. Citeseer.

Changnon, D., T. B. McKee, and N. J. Doesken, 1993: Annual snowpack patterns across the Rockies: long-term trends and associated 500-mb synoptic patterns. *Monthly Weather Review*, **121**, 633–647.

Chatfield, C. and D. Prothero, 1973: Box-Jenkins seasonal forecasting: problems in a case-study. *Journal of the Royal Statistical Society: Series A*, **136**, no. 3, 295–315.

Chen, H., 2001: Initialization for NORTA: Generation of random vectors with specified marginals and correlations. *INFORMS Journal on Computing*, **13**, no. 4, 312–331.

Chib, S. and E. Greenberg, 1998: Analysis of multivariate probit models. *Biometrika*, **85**, no. 2, 347–361.

Chipman, J. S., 1979: Efficiency of least-squares estimation of linear trend when residuals are autocorrelated. *Econometrica: Journal of the Econometric Society*, 115–128.

Cochrane, D. and G. H. Orcutt, 1949: Application of least squares regression to relationships containing auto-correlated error terms. *Journal of the American statistical association*, **44**, no. 245, 32–61.

Cressie, N., 2015: *Statistics for Spatial Data*. John Wiley & Sons.

Cressie, N. and C. K. Wikle, 2015: *Statistics for Spatio-Temporal data*. John Wiley & Sons.

Cui, Y., Q. Li, and F. Zhu, 2021: Modeling Z-valued time series based on new versions of the Skellam INGARCH model. *Brazilian Journal of Probability and Statistics*, **35**, no. 2, 293–314.

Cui, Y. and R. Lund, 2009: A new look at time series of counts. *Biometrika*, **96**, no. 4, 781–792.

Czado, C., T. Gneiting, and L. Held, 2009: Predictive model assessment for count data. *Biometrics*, **65**, no. 4, 1254–1261.

Davis, R. A., W. T. Dunsmuir, and S. B. Streett, 2005: Maximum likelihood estimation for an observation driven model for Poisson counts. *Methodology and Computing in Applied Probability*, **7**, 149–159.

Davis, R. A., K. Fokianos, S. H. Holan, H. Joe, J. Livsey, R. Lund, V. Pipiras, and

N. Ravishanker, 2021: Count time series: A methodological review. *Journal of the American Statistical Association*, **116**, 1533–1547.

Davis, R. A., S. H. Holan, R. Lund, and N. Ravishanker, 2016: *Handbook of Discrete-Valued Time Series*. CRC Press, New York City, NY.

Davis, R. A., T. C. M. Lee, and G. A. Rodriguez-Yam, 2006: Structural break estimation for nonstationary time series models. *Journal of the American Statistical Association*, **101**, 223–239.

Dawid, A. P., 1984: Present position and potential developments: Some personal views statistical theory the prequential approach. *Journal of the Royal Statistical Society: Series A*, **147**, no. 2, 278–290.

De Oliveira, V., 2013: Hierarchical Poisson models for spatial count data. *Journal of Multivariate Analysis*, **122**, 393–408.

Derksen, C. and R. Brown, 2012: Spring snow cover extent reductions in the 2008–2012 period exceeding climate model projections. *Geophysical Research Letters*, **39**, no. 19.

Déry, S. J. and R. D. Brown, 2007: Recent Northern Hemisphere snow cover extent trends and implications for the snow-albedo feedback. *Geophysical Research Letters*, **34**, no. 22.

Djurić, P. M. and M. F. Bugallo, 2013: Particle filtering for high-dimensional systems. *2013 5th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, IEEE, 352–355.

Doornik, J. and H. Hansen, 1994: A practical test for univariate and multivariate normality. Nuffield College, Oxford, UK, Discussion paper.

Douc, R., E. Moulines, and D. Stoffer, 2014: *Nonlinear Time Series: Theory, Methods and Applications with R Examples*. CRC press.

Du, J.-G. and Y. Li, 1991: The integer-valued autoregressive (INAR(p)) model. *Journal of Time Series Analysis*, **12**, no. 2, 129–142.

Dunsmuir, W. T., D. J. Scott, et al., 2015: The glarma package for observation-driven time series regression of counts. *Journal of Statistical Software*, **67**, no. 7, 1–36.

Dye, D. G., 2002: Variability and trends in the annual snow-cover cycle in Northern Hemisphere land areas, 1972–2000. *Hydrological Processes*, **16**, no. 15, 3065–3077.

Dyer, J. L. and T. L. Mote, 2006: Spatial variability and trends in observed snow depth over North America. *Geophysical Research Letters*, **33**, L16503–.

Escarela, G., R. H. Mena, and A. Castillo-Morales, 2006: A flexible class of parametric transition regression models based on copulas: application to poliomyelitis incidence. *Statistical Methods in Medical Research*, **15**, no. 6, 593–609.

Estilow, T. W., A. H. Young, and D. A. Robinson, 2015: A long-term Northern Hemisphere snow cover extent data record for climate studies and monitoring. *Earth System Science Data*, **7**, no. 1, 137–142.

Falarz, M., 2004: Variability and trends in the duration and depth of snow cover in

Poland in the 20th century. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, **24**, no. 13, 1713–1727.

Fernando, D., M. Alqawba, M. Samad, and N. Diawara, 2022: Review of Copula for Bivariate Distributions of Zero-Inflated Count Time Series Data. *International Journal of Statistics and Probability*, **11**, no. 6.

Fisher, T. J., R. B. Lund, and M. W. Robbins, 2020: A Statistical Analysis of North Atlantic Tropical Cyclone Changes. *Evaluating Climate Change Impacts*, Chapman and Hall/CRC, 25–43.

Flanner, M. G., C. S. Zender, P. G. Hess, N. M. Mahowald, T. H. Painter, V. Ramanathan, and P. Rasch, 2009: Springtime warming and reduced snow cover from carbonaceous particles. *Atmospheric Chemistry and Physics*, **9**, no. 7, 2481–2497.

Fokianos, K., 2012: Count time series models. *Handbook of Statistics*, C. R. Rao, C. Rao, and V. Govindaraju, Eds., Elsevier, volume 30, 315–347.

Fokianos, K., A. Rahbek, and D. Tjøstheim, 2009a: Poisson autoregression. *Journal of the American Statistical Association*, **104**, no. 488, 1430–1439.

— 2009b: Poisson autoregression. *Journal of the American Statistical Association*, **104**, no. 488, 1430–1439.

Fralix, B., J. Livsey, and R. Lund, 2012: Renewal sequences with periodic dynamics. *Probability in the Engineering and Informational Sciences*, **26**, no. 1, 1–15.

Franke, J. and T. Rao Subba, 1993: Multivariate first-order integer-valued autoregressions. Department of Mathematics, UMIST.

Franses, P. H., 1994: A multivariate approach to modeling univariate seasonal time series. *Journal of Econometrics*, **63**, no. 1, 133–151.

Franses, P. H. and R. Paap, 2004: *Periodic Time Series Models*. Oxford University Press.

Fryzlewicz, P., 2014: Wild binary segmentation for multiple changepoint detection. *Annals of Statistics*, **42**, 2243–2281.

Gallagher, C., R. Lund, and M. Robbins, 2013: Changepoint detection in climate time series with long-term trends. *Journal of Climate*, **26**, 4994–5006.

Gardner, W. and L. Franks, 1975: Characterization of cyclostationary random signal processes. *IEEE Transactions on Information Theory*, **21**, no. 1, 4–14.

Gardner, W. A., A. Napolitano, and L. Paura, 2006: Cyclostationarity: Half a century of research. *Signal Processing*, **86**, no. 4, 639–697.

Genz, A. and F. Bretz, 2002: Comparison of methods for the computation of multivariate t probabilities. *Journal of Computational and Graphical Statistics*, **11**, no. 4, 950–971.

Geweke, J., 1991: Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints and the evaluation of constraint probabilities.

Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface, Citeseer, volume 571, 578.

Gladyshev, E., 1963: Periodically and almost-periodically correlated random processes with a continuous time parameter. *Theory of Probability & Its Applications*, **8**, no. 2, 173–177.

Glasserman, P. and D. D. Yao, 1992: Some guidelines and guarantees for common random numbers. *Management Science*, **38**, no. 6, 884–908.

Goldenberg, S. B. and L. J. Shapiro, 1996: Physical mechanisms for the association of El Niño and West African rainfall with Atlantic major hurricane activity. *Journal of Climate*, **9**, no. 6, 1169–1187.

Goudie, A. S., 2018: *Human Impact on the Natural Environment*. John Wiley & Sons, Hoboken, NJ, USA.

Gray, W. M., 1984: Atlantic seasonal hurricane frequency. Part I: El Niño and 30 mb quasi-biennial oscillation influences. *Monthly Weather Review*, **112**, no. 9, 1649–1668.

Grundstein, A. and T. Mote, 2010: Trends in average snow depth across the western United States. *Physical Geography*, **31**, 172–185.

Hajivassiliou, V., D. McFadden, and P. Ruud, 1996: Simulation of multivariate normal rectangle probabilities and their derivatives theoretical and computational results. *Journal of Econometrics*, **72**, no. 1-2, 85–134.

Han, Z. and V. De Oliveira, 2016: On the correlation structure of Gaussian copula models for geostatistical count data. *Australian & New Zealand Journal of Statistics*, **58**, no. 1, 47–69.

— 2018: gcKrig: An R package for the analysis of geostatistical count data using Gaussian copulas. *Journal of Statistical Software*, **87**, no. 1, 1–32.

— 2020: Maximum likelihood estimation of Gaussian copula models for geostatistical count data. *Communications in Statistics-Simulation and Computation*, **49**, no. 8, 1957–1981.

Hurd, H. L. and A. Miamee, 2007: *Periodically Correlated Random Sequences: Spectral Theory and Practice*, volume 355. John Wiley & Sons.

Inouye, D. I., E. Yang, G. I. Allen, and P. Ravikumar, 2017: A review of multivariate distributions for count data derived from the Poisson distribution. *Wiley Interdisciplinary Reviews: Computational Statistics*, **9**, no. 3, e1398.

Jacobs, P. A. and P. A. Lewis, 1978a: Discrete time series generated by mixtures. I: Correlational and runs properties. *Journal of the Royal Statistical Society: Series B (Methodological)*, **40**, no. 1, 94–105.

— 1978b: Discrete time series generated by mixtures II: asymptotic properties. *Journal of the Royal Statistical Society: Series B (Methodological)*, **40**, no. 2, 222–228.

— 1978c: Discrete Time Series Generated by Mixtures. III. Autoregressive Processes (DAR (p)). Naval Postgraduate School, Monterey California.

Jang, J. and R. Oh, 2021: A review on Poisson, Cox, Hawkes, shot-noise Poisson and dynamic contagion process and their compound processes. *Annals of Actuarial Science*, **15**, no. 3, 623–644.

Jia, Y., S. Kechagias, J. Livsey, R. Lund, and V. Pipiras, 2023: Latent Gaussian count time series. *Journal of the American Statistical Association*, **118**, no. 541, 596–606.

Jia, Y., R. Lund, and J. Livsey, 2021: Superpositioned stationary count time series. *Probability in the Engineering and Informational Sciences*, **35**, no. 3, 538–556.

Joe, H., 1996: Time series models with univariate margins in the convolution-closed infinitely divisible class. *Journal of Applied Probability*, **33**, no. 3, 664–677.

— 2014: *Dependence Modeling with Copulas*. CRC press.

— 2016: Markov models for count time series. *Handbook of Discrete-Valued Time Series*, R. A. Davis, S. H. Holan, R. Lund, and N. Ravishanker, Eds., CRC Press, New York City, NY, 49–70.

— 2019: Likelihood inference for generalized integer autoregressive time series models. *Econometrics*, **7**, no. 4.

Kachour, M. and J.-F. Yao, 2009: First-order rounded integer-valued autoregressive (RINAR (1)) process. *Journal of Time Series Analysis*, **30**, no. 4, 417–448.

Karl, T. R., P. Y. Groisman, R. W. Knight, and R. R. H. Jr., 1993: Recent variations of snow cover and snowfall in North America and their relation to precipitation and temperature variations. *Journal of Climate*, **6**, 1327–1344.

Karl, T. R., J. M. Melillo, T. C. Peterson, and S. J. Hassol, 2009: *Global Climate Change Impacts in the United States*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Kazianka, H., 2013: Approximate copula-based estimation and prediction of discrete spatial data. *Stochastic Environmental Research and Risk Assessment*, **27**, no. 8, 2015–2026.

Kazianka, H. and J. Pilz, 2010: Copula-based geostatistical modeling of continuous and discrete data including covariates. *Stochastic Environmental Research and Risk Assessment*, **24**, no. 5, 661–673.

Kedem, B., 1980: Estimation of the parameters in stationary autoregressive processes after hard limiting. *Journal of the American Statistical Association*, **75**, no. 369, 146–153.

Kleinman, N. L., J. C. Spall, and D. Q. Naiman, 1999: Simulation-based optimization with stochastic approximation using common random numbers. *Management Science*, **45**, no. 11, 1570–1578.

Kluver, D., 2007: *Characteristics and trends in North American snowfall from a comprehensive gridded data set*. Masters thesis, University of Delaware.

Kluver, D., T. Mote, D. Leathers, W. Chan, G. R. Henderson, and D. A. Robinson, 2016: Creation and validation of a comprehensive 1° by 1° daily gridded North American

dataset for 1900-2009: Snowfall. *Journal of Atmospheric and Oceanic Technology*, **33**, 857–871.

Knowles, N., M. D. Dettinger, and D. R. Cayan, 2006: Trends in snowfall versus rainfall in the western United States. *Journal of Climate*, **19**, 4545–4559.

Kocherlakota, S. and K. Kocherlakota, 2017: *Bivariate discrete distributions*. CRC Press.

Kong, J. and R. Lund, 2023: Seasonal count time series. *Journal of Time Series Analysis*, **44**, no. 1, 93–124.

Kunkel, K. E., M. Palecki, L. Ensor, K. G. Hubbard, D. Robinson, K. Redmond, and D. Easterling, 2009: Trends in twentieth-century U.S. snowfall using a quality-controlled dataset. *Journal of Atmospheric and Oceanic Technology*, **26**, 33–44.

Kunkel, K. E., M. A. Palecki, K. G. Hubbard, D. A. Robinson, K. T. Redmond, and D. R. Easterling, 2007: Trend identification in twentieth-century U.S. snowfall: The challenges. *Journal of Atmospheric and Oceanic Technology*, **24**, 64–73.

Lambert, D., 1992: Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, **34**, no. 1, 1–14.

Latour, A., 1997: The multivariate GINAR (p) process. *Advances in Applied Probability*, 228–248.

Lawrence, D. M. and A. G. Slater, 2010: The contribution of snow condition trends to future ground climate. *Climate Dynamics*, **34**, no. 7, 969–981.

Lee, J., S. Li, and R. Lund, 2014: Trends in extreme U.S. temperatures. *Journal of Climate*, **27**, 4209–4225.

Lee, J. and R. Lund, 2004: Revisiting simple linear regression with autocorrelated errors. *Biometrika*, **91**, no. 1, 240–245.

Lee, J., R. Lund, J. Woody, and Y. Xu, 2020: Trend assessment for daily snow depths with changepoint considerations. *Environmetrics*, **31**, no. 1, e2580.

Lemke, P. et al., 2007: Observations: Changes in Snow, Ice, and Frozen Ground. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovenmental Panel on Climate Change*, S. Solomon et al., Eds., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Liston, G. E. and C. A. Hiemstra, 2011: The changing cryosphere: Pan-Arctic snow trends (1979–2009). *Journal of Climate*, **24**, no. 21, 5691–5712.

Livsey, J., R. Lund, S. Kechagias, V. Pipiras, et al., 2018: Multivariate integer-valued time series with flexible autocovariances and their application to major hurricane counts. *The Annals of Applied Statistics*, **12**, no. 1, 408–431.

Lu, Q., R. B. Lund, and T. C. M. Lee, 2010: An MDL approach to the climate segmentation problem. *Annals of Applied Statistics*, **4**, 299–319.

Lund, R., 2006: A seasonal analysis of riverflow trends. *Journal of Statistical Computation and Simulation*, **76**, 397–405.

— 2011: Choosing seasonal autocovariance structures: PARMA or SARMA. *Economic Time Series: Modelling and Seasonality*, W. R. Bell, S. H. Holan, and T. S. McElroy, Eds., Chapman and Hall/CRC, 63–80.

Lund, R. and I. Basawa, 1999: Modeling and inference for periodically correlated time series. *Asymptotics, Nonparametrics, and Time Series*, S. Ghosh, ed., CRC Press, 37–52.

— 2000: Recursive prediction and likelihood evaluation for periodic ARMA models. *Journal of Time Series Analysis*, **21**, no. 1, 75–93.

Lund, R., S. H. Holan, and J. Livsey, 2016: Long memory discrete-valued time series. *Handbook of Discrete-Valued Time Series*, R. A. Davis, S. H. Holan, R. Lund, and N. Ravishanker, Eds., CRC Press, New York City, NY.

Lund, R., H. Hurd, P. Bloomfield, and R. Smith, 1995a: Climatological time series with periodic correlation. *Journal of Climate*, **8**, no. 11, 2787–2809.

— 1995b: Climatological time series with periodic correlation. *Journal of Climate*, **8**, no. 11, 2787–2809.

Lund, R. and J. Livsey, 2016: Renewal-based count time series. *Handbook of Discrete-Valued Time Series*, R. A. Davis, S. H. Holan, R. Lund, and N. Ravishanker, Eds., CRC Press, New York City, NY.

Lund, R., L. Seymour, and K. Kafadar, 2001: Temperature trends in the United States. *Environmetrics*, **12**, 673–690.

Lund, R., Q. Shao, and I. Basawa, 2006: Parsimonious periodic time series modeling. *Australian & New Zealand Journal of Statistics*, **48**, no. 1, 33–47.

Lund, R., X. Wang, J. Reeves, Q. Lu, C. Gallagher, and Y. Feng, 2007a: Changepoint detection in periodic and autocorrelated time series. *Journal of Climate*, **20**, 5178–5190.

Lund, R., X. L. Wang, Q. Q. Lu, J. Reeves, C. Gallagher, and Y. Feng, 2007b: Changepoint detection in periodic and autocorrelated time series. *Journal of Climate*, **20**, no. 20, 5178–5190.

Masarotto, G. and C. Varin, 2012: Gaussian copula marginal regression. *Electronic Journal of Statistics*, **6**, 1517–1549.

Masarotto, G., C. Varin, et al., 2017: Gaussian copula regression in R. *Journal of Statistical Software*, **77**, no. 8, 1–26.

McKenzie, E., 1985: Some simple models for discrete variate time series 1. *Journal of the American Water Resources Association*, **21**, no. 4, 645–650.

— 1986: Autoregressive moving-average processes with negative-binomial and geometric marginal distributions. *Advances in Applied Probability*, **18**, no. 3, 679–705.

— 1988: Some ARMA models for dependent sequences of Poisson counts. *Advances in Applied Probability*, **20**, no. 4, 822–835.

Minkova, L. D. and E. Omey, 2014: A new Markov binomial distribution. *Communications in Statistics-Theory and Methods*, **43**, no. 13, 2674–2688.

Mitchell, J. M., Jr., 1953: On the causes of instrumentally observed secular temperature trends. *Journal of Meteorology*, **10**, 244–261.

Möller, T. A. and C. H. Weiß, 2020: Generalized discrete autoregressive moving-average models. *Applied Stochastic Models in Business and Industry*, **36**, no. 4, 641–659.

Monitoring, A. et al., 2017: Snow, Water, Ice and Permafrost in the Arctic (SWIPA) 2017.

Monteiro, M., M. G. Scotto, and I. Pereira, 2010: Integer-valued autoregressive processes with periodic structure. *Journal of Statistical Planning and Inference*, **140**, no. 6, 1529–1541.

— 2015: A periodic bivariate integer-valued autoregressive model. *Dynamics, Games and Science: International Conference and Advanced School Planet Earth, DGS II, Portugal, August 28–September 6, 2013*, J.-P. Bourguignon, R. Jeltsch, A. A. Pinto, and M. Viana, Eds., Springer, 455–477.

Mooley, D., 1980: Severe cyclonic storms in the Bay of Bengal, 1877–1977. *Monthly Weather Review*, **108**, no. 10, 1647–1655.

Moriña, D., P. Puig, J. Ríos, A. Vilella, and A. Trilla, 2011: A statistical model for hospital admissions caused by seasonal diseases. *Statistics in Medicine*, **30**, no. 26, 3125–3136.

Mote, P. W., 2003: Trends in snow water equivalent in the Pacific Northwest and their climatic causes. *Geophysical Research Letters*, **30**, no. 12.

Mote, P. W., S. Li, D. P. Lettenmaier, M. Xiao, and R. Engel, 2018: Dramatic declines in snowpack in the western U.S. *Npj Climate and Atmospheric Science*, **1**, no. 1, 1–6.

Mudryk, L. E., A. Chereque, R. Brown, C. Derksen, K. Luojus, and B. Decharme, 2020: Arctic Report Card 2020: Terrestrial Snow Cover.

Naesseth, C., F. Lindsten, and T. Schon, 2015: Nested sequential monte carlo methods. *International Conference on Machine Learning*, PMLR, 1292–1301.

Notarnicola, C., 2022: Overall negative trends for snow cover extent and duration in global mountain regions over 1982–2020. *Scientific Reports*, **12**, no. 1, 1–16.

Ouzzani, F. and M. Bentarzi, 2019: On mixture periodic integer-valued ARCH models. *Communications in Statistics-Simulation and Computation*, 1–27.

Painter, T. H., J. Dozier, D. A. Roberts, R. E. Davis, and R. O. Green, 2003: Retrieval of subpixel snow-covered area and grain size from imaging spectrometer data. *Remote Sensing of Environment*, **85**, no. 1, 64–77.

Pedeli, X. and D. Karlis, 2011: A bivariate INAR (1) process with application. *Statistical modelling*, **11**, no. 4, 325–349.

— 2013a: On composite likelihood estimation of a multivariate INAR (1) model. *Journal of Time Series Analysis*, **34**, no. 2, 206–220.

— 2013b: Some properties of multivariate INAR (1) processes. *Computational Statistics & Data Analysis*, **67**, 213–225.

Perona, P., A. Porporato, and L. Ridolfi, 2007: A stochastic process for the interannual snow storage and melting dynamics. *Journal of Geophysical Research Atmospheres*, **112**.

Pickands III, J., 1975: Statistical inference using extreme order statistics. *the Annals of Statistics*, 119–131.

Pierce, D. W. and D. R. Cayan, 2013: The uneven response of different snow measures to human-induced climate warming. *Journal of Climate*, **26**, 4148–4167.

Pipiras, V. and M. S. Taqqu, 2017: *Long-range Dependence and Self-similarity*, volume 45. Cambridge University Press.

Pithan, F. and T. Mauritsen, 2014: Arctic amplification dominated by temperature feedbacks in contemporary climate models. *Nature Geoscience*, **7**, 181–184.

R Development Core Team, 2018: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rebeschini, P. and R. Van Handel, 2015: Can local particle filters beat the curse of dimensionality? *The Annals of Applied Probability*, **25**, no. 5, 2809–2866.

Reeves, J., J. Chen, X. L. Wang, R. Lund, and Q. Q. Lu, 2007: A review and comparison of changepoint detection techniques for climate data. *Journal of Applied Meteorology and Climatology*, **46**, 900–915.

Reinsel, G. C., 2003: *Elements of Multivariate Time Series Analysis*. Springer Science & Business Media.

Ristic, M. and B. Popovic, 2019: A new bivariate binomial time series model. *Markov Process Relat*, **25**, 301–328.

Robbins, M. W., R. B. Lund, C. M. Gallagher, and Q. Lu, 2011: Changepoints in the North Atlantic tropical cyclone record. *Journal of the American Statistical Association*, **106**, no. 493, 89–99.

Robinson, D. A., K. F. Dewey, and R. R. Heim, 1993: Global snow cover monitoring: An update. *Bulletin of the American Meteorological Society*, **74**, no. 9, 1689–1696.

Robinson, D. A., T. W. Estilow, and NOAA CDR Program, 2012: *NOAA Climate Data Record (CDR) of Northern Hemisphere (NH) Snow Cover Extent (SCE), Version 1*.

Santos, C., I. Pereira, and M. Scotto, 2019: Periodic INAR(1) models with Skellam-distributed innovations. *International Conference on Computational Science and Its Applications*, Springer, 64–78.

Santos, C., I. Pereira, and M. G. Scotto, 2021: On the theory of periodic multivariate INAR processes. *Statistical Papers*, **62**, 1291–1348.

Scotto, M. G., C. H. Weiß, and S. Gouveia, 2015: Thinning-based models in the analysis of integer-valued time series: a review. *Statistical Modelling*, **15**, no. 6, 590–618.

Scotto, M. G., C. H. Weiß, M. E. Silva, and I. Pereira, 2014: Bivariate binomial autoregressive models. *Journal of Multivariate Analysis*, **125**, 233–251.

Serreze, M., J. Walsh, F. Chapin, T. Osterkamp, M. Dyurgerov, V. Romanovsky, W. Oechel, J. Morison, T. Zhang, and R. Barry, 2000: Observational evidence of

recent change in the northern high-latitude environment. *Climatic Change*, **46**, no. 1, 159–207.

Serreze, M. C., M. P. Clark, D. L. Mcginnis, and D. A. Robinson, 1998: Characteristics of snowfall over the eastern half of the United States and relationships with principal modes of low-frequency atmospheric variability. *Journal of Climate*, **11**, 234–250.

Shao, Q., 2006: Mixture periodic autoregressive time series models. *Statistics & Probability Letters*, **76**, no. 6, 609–618.

Shao, Q. and R. B. Lund, 2004: Computation and characterization of autocorrelations and partial autocorrelations in periodic ARMA models. *Journal of Time Series Analysis*, **25**, 359–372.

Shao, Q. and P. Ni, 2004: Least-squares estimation and ANOVA for periodic autoregressive time series. *Statistics & Probability Letters*, **69**, no. 3, 287–297.

Shumway, R. H. and D. S. Stoffer, 2017: ARIMA models. *Time Series Analysis and Its Applications*, Springer, 75–163.

Shumway, R. H., D. S. Stoffer, and D. S. Stoffer, 2000: *Time Series Analysis and its Applications*, volume 3. Springer.

Smith, M. S. and M. A. Khaled, 2012: Estimation of copula models with discrete margins via Bayesian data augmentation. *Journal of the American Statistical Association*, **107**, no. 497, 290–303.

Smith, W. L., 1958: Renewal theory and its ramifications. *Journal of the Royal Statistical Society: Series B (Methodological)*, **20**, no. 2, 243–284.

Solow, A. R. and A. R. Beet, 2008: On the incompleteness of the historical record of North Atlantic tropical cyclones. *Geophysical research letters*, **35**, no. 11.

Steenbergen, M. R., 2006: Maximum likelihood programming in R. University of North Carolina, Chapel Hill. Department of Political Science.

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

Teicher, H., 1954: On the multivariate Poisson distribution. *Scandinavian Actuarial Journal*, **1954**, no. 1, 1–9.

Tesfaye, Y. G., M. M. Meerschaert, and P. L. Anderson, 2004: Identification of PARMA models and their application to the modeling of river flows. *AGU Spring Meeting Abstracts*, volume 2004, H33A–10.

Tong, Y. L., 2014: *Probability Inequalities in Multivariate Distributions*. Academic Press.

Tong, Y. L. and Y. Tong, 1990: *Fundamental properties and sampling distributions of the multivariate normal distribution*. Springer.

Van Mantgem, P. J., N. L. Stephenson, J. C. Byrne, L. D. Daniels, J. F. Franklin, P. Z. Fulé, M. E. Harmon, A. J. Larson, J. M. Smith, A. H. Taylor, et al., 2009: Widespread

increase of tree mortality rates in the western United States. *Science*, **323**, no. 5913, 521–524.

Vecchia, A., 1985: Periodic autogressive-moving average (PARMA) modeling with applications to water resources. *Journal of the American Water Resources Association*, **21**, no. 5, 721–730.

Ver Hoef, J. M. and P. L. Boveng, 2007: Quasi-Poisson vs. Negative Binomial Regression: How Should We Model Overdispersed Count Data? *Ecology*, **88**, no. 11, 2766–2772.

Weiß, C. H., 2008: The combined INAR(p) models for time series of counts. *Statistics & probability letters*, **78**, no. 13, 1817–1822.

— 2018: *An Introduction to Discrete-Valued Time Series*. John Wiley & Sons.

Whitt, W., 1976: Bivariate distributions with given marginals. *The Annals of Statistics*, **4**, no. 6, 1280–1289.

Wiesnet, D., C. Ropelewski, G. Kukla, and D. Robinson, 1987: A discussion of the accuracy of NOAA satellite-derived global seasonal snow cover measurements. *Large Scale Effects of Seasonal Snow Cover*, **166**, 291–304.

Woody, J., R. Lund, A. Grundstein, and T. Mote, 2009a: A storage model approach to the assessment of snow depth trends. *Water Resources Research*, **45**, W10426.

Woody, J., R. Lund, A. J. Grundstein, and T. L. Mote, 2009b: A storage model approach to the assessment of snow depth trends. *Water Resources Research*, **45**, no. 10.

Yao, Q. and P. J. Brockwell, 2006: Gaussian maximum likelihood estimation for ARMA models II: spatial processes. *Bernoulli*, **12**, no. 3, 403–429.

Yue, S., P. Pilon, B. Phinney, and G. Cavadias, 2002: The influence of autocorrelation on the ability to detect trend in hydrological series. *Hydrological Processes*, **16**, no. 9, 1807–1829.

Zhang, X., L. A. Vincent, W. Hogg, and A. Niitsoo, 2000a: Temperature and precipitation trends in Canada during the 20th century. *Atmosphere-ocean*, **38**, no. 3, 395–429.

Zhang, X., L. A. Vincent, W. D. Hogg, and A. Niitsoo, 2000b: Temperature and precipitation trends in Canada during the 20th century. *Atmosphere – Ocean (Canadian Meteorological & Oceanographic Society)*, **38**, 395–429.

Zheng, X., A. Kottas, and B. Sansó, 2022: On construction and estimation of stationary mixture transition distribution models. *Journal of Computational and Graphical Statistics*, **31**, no. 1, 283–293.

Zhu, R. and H. Joe, 2003: A new type of discrete self-decomposability and its application to continuous-time Markov processes for modeling count data time series. *Stochastic Models*, **19**, no. 2, 235–254.

Zona, D., B. Gioli, R. Commane, J. Lindaas, S. C. Wofsy, C. E. Miller, S. J. Dinardo, S. Dengel, C. Sweeney, A. Karion, et al., 2016: Cold season emissions dominate the

Arctic tundra methane budget. *Proceedings of the National Academy of Sciences*, **113**, no. 1, 40–45.