Summer 2021

Count data time series models and their applications

Yi Zhang

Department: Mathematics and Statistics

COUNT DATA TIME SERIES MODELS AND THEIR APPLICATIONS

by

YI ZHANG

A DISSERTATION

Presented to the Graduate Faculty of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

in

MATHEMATICS WITH STATISTICS EMPHASIS

2021

Approved by:

V.A. Samaranayake, Advisor
Gayla R. Olbricht
Robert Paige
Gregory Gelles
Xuerong Wen
Matthew Thimgan

PUBLICATION DISSERTATION OPTION

This dissertation consists of the following three articles, formatted in the style used by the Missouri University of Science and Technology.

Paper I: Pages 12-42, an abbreviated version have been published in the proceedings of JSM Conference in Boulder, CO, in August 2019.

Paper II: Pages 43-70, an abbreviated have been published in the proceedings of JSM virtual Conference in August 2020.

Paper III: Pages 71-93, an abbreviated have been published in Journal: Stat, 10(1), e345.

ABSTRACT

Due to fast developments of advanced sensors, count data sets have become ubiquitous in many fields. Modeling and forecasting such time series have generated great interest. Modeling can shed light on the behavior of the count series and to see how they are related to other factors such as the environmental conditions under which the data are generated. In this research, three approaches to modeling such count data are proposed.

First, a periodic autoregressive conditional Poisson (PACP) model is proposed as a natural generalization of the autoregressive conditional Poisson (ACP) model. By allowing for cyclical variations in the parameters of the model, it provides a way to explain the periodicity inherent in many count data series. For example, in epidemiology the prevalence of a disease may depend on the season. The autoregressive conditional Poisson hidden Markov model (ACP-HMM) is developed to deal with count data time series whose mean, conditional on the past, is a function of previous observations, with this relationship possibly determined by an unobserved process that switches its state or regime as time progresses. This model, in a sense, is the combination of the discrete version of the autoregressive conditional heteroscedastic (ARCH) formulation and the Poisson hidden Markov model. Both the above models address the frequently present serial correlation and the clustering of high or low counts observed in time series of count data, while at the same time allowing the underlying data generating mechanism to change cyclically or according to a hidden Markov process. Applications to empirical data sets show that these models provide a better fit than the standard ACP models. In addition to the above models, a modification of a zero-inflated Poisson model is used to analyze activity counts of the fruit fly. The model captures the dynamic structure of activity patterns and the fly's propensity to sleep. The obtained results when fed to a convolutional neural network provides the possibility of building a predictive model to identify fruit flies with short and long lifespans.

ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere gratitude to my advisor, Prof. V.A. Samaranayake, for his support, encouragement, and insightful guidance during my study at Missouri University of Science and Technology. He guided me to sharpen my thinking and brought my academic work to a higher level. Without him, I could not reach this point. I also would like to thank Dr. Matthew Thimgan for funding part of my dissertation through a grant funded by the National Institute of General Medical Sciences of the National Institute of Health under award number R15GM117507. I am grateful to my dissertation committee members for their comments and suggestions. I also would like to thank all the faculty members of the department of Mathematics and Statistics, for their help in many ways with great patience. Finally, I would like to extend my special and sincere thanks to my grandparents, parents, husband, and all other family members, for their love and unwavering support.

TABLE OF CONTENTS

Page

LIST OF ILLUSTRATIONS

Figure Page

PAPER I

PAPER II

PAPER III

LIST OF TABLES

SECTION

1. INTRODUCTION

Nowadays, a wealth of time series data are generated via advanced technology, such as activity data recorded by wearable electronics. Further analysis of those time series plays an important role in helping people understanding the underlying data generating mechanism and providing guidance for future investigations. Among different types of time series, count data is being increasingly collected and important to model. One example is the high frequency activity data that are captured by high-tech sensors. Proper modeling and interpretation of such data would help us understand the relationships between behavior and biological characteristics.

Most count data have two main features that could not be ignored when modeling the underlying data generating process: serial correlation and seasonal or regime-based changes in the underlying mechanism. A high percentage of count data time series exhibit correlations among observations, resulting in clusters of high or low counts. In other words, a high count is likely followed by another high value, while a low count probably will be observed subsequent to a low number. For the second feature, the magnitude of the counts may change across time. For example, seasons have a huge impact on the case numbers of some epidemics; therefore, it is reasonable to assume the latent process changes its structure across different periods. Another example is the number of transactions of stocks observed intra-day, the interrelationship among which depends on the time of day, such as the beginning of the market, the middle of the day, and the closing of the market. These two characteristics lead to the necessity of developing models that could deal with such phenomena simultaneously. Thus, the models we proposed here are an attempt to provide formulations that can account for behaviors described above.

2. LITERATURE REVIEW

A time series is a collection of observations $\{X_t : t \in \mathbb{N}\}$ that are observed sequentially according to time, and such series can be found in many areas ranging from biology to economics. With the development of advanced technology, numerous time series data sets have become available for analysis, enabling us to understand the underlying data generating mechanisms. The count data series consist of the numbers of discrete events recorded over a period of time, and most often, the measurements are taken at evenly spaced time intervals. A large number of models have been developed to explain the underlying mechanisms that generate real-world series, especially continuous valued time series, but there is a need to develop more versatile models for count data processes.

Markov chains are one way to deal with count data [1]. This method requires the definition of the (stationary) transition probabilities between all possible outcomes that the random variable could take. However, a sequence $\{X_t : t \in \mathbb{N}\}$ of random variables taking values in some set is a Markov chain if and only if the conditional distribution of X_{n+1}, given $X_1, X_2, \ldots, X_n$, depends on X_n only, which would limit the use of this model because it ignores dependence on values prior to time n. Also, when the number of possible outcomes grows very large, this model is no longer easily tractable and its parameter estimation becomes cumbersome.

A related but more complex approach is to use a hidden Markov model. These models are an alternative application of Markov chains as it is assumed that an underlying unobserved state of the system, determined by a Markov process, governs the parameters of the counting process. The system's present state should determine the distribution of observations at the current time [2]. Sebastian et al. [3] developed the Markov ordinal logistic regression model with the transition probability defined as $P(Y_i|Y_{i-1}, \mathbf{Z}) = e^{\alpha_i - \mathbf{Z}'\beta}[1 - e^{\alpha_i - \mathbf{Z}'\beta}]^{-1}$, where Y_i represents the states of the Markov chain,

while $\mathbf{Z}$ denotes a vector of some known covariates. Cooper et al. [4] proposed a so-called 'structured hidden Markov model' for the epidemic process that intuitively follows a hidden Markov chain process since patients communicate with each other and the epidemic process usually gives out routine surveillance data that could often be partially observed. 'Structured' implies that a simple transition model is driving the underlying Markov chain. However, the need to determine the order of the Markov chain before applying the model accounts for one obvious drawback of this type of model. Another inevitable problem of such models is that the variability of the outcomes may be small.

Another category of methods developed from the application of the Markov chain is the binomial thinning process proposed by McKenzie [5] as a simple model to deal with discrete variate time series problems. The thinning operator takes the sum of $X_i, i \in \mathbb{N}$, identically independent Bernoulli random variables, each of which takes value 1 with probability α and 0 with probability $1 - \alpha$. The data generating mechanism is modeled similarly to an autoregressive (AR) process in the sense that the current count is dependent on the number of Bernoulli random variables given by the previous count. For example, Poisson AR(1) process is constructed as $X_n = \alpha * X_{n-1} + W_n$, where X_n and W_n are both Poisson process with means θ and $\theta(1 - \alpha)$ respectively. The thinning operator $*$ is defined as follows: $\alpha * X_{n-1}$ denotes the number of successes observed from X_{n-1} Bernoulli trails with success probability α. Geometric AR(1), negative binomial AR(1), binomial AR(1), and compound correlated bi-variate Poisson distribution were also proposed. They also investigated the seasonality problem in count data, and the seasonal mean was set as $\mu_n = a\cos(\omega n) + b\sin(\omega n)$. Similarly, the innovation mean was set as $\omega_n = A\cos(\omega n) + B\sin(\omega n)$, where $A = a - \alpha(a\cos\omega - b\sin\omega)$ and $B = b - \alpha(a\sin\omega + b\cos\omega)$. Zhu and Joe [6] further modified the model, incorporating covariates to the mean of the stationary Markov time series allowing time varying components. Also, they extended the model with a structure to mimic an AR(2) model to account for higher order dependence present in many time series of count data. The problem with this type of model is that the seasonal

pattern embedded in this fashion is not flexible enough to model data demonstrating complex periodic components, which may require the inclusion of a large number of trigonometric functions to model the cyclical behavior with reasonable accuracy.

Many count data models are based on the use of the Poisson distribution. The Poisson regression model, as the basic count model, is well described in the book by Cameron and Trivedi [7]. In this approach it is assumed that y_i, an independent observation from a Poisson distribution, given the vector of regressor $\mathbf{x}_i$, has a density function $f(y_i|\mathbf{x}_i) = e^{-\mu_i}\mu_i{}^{y_i}[y_i!]^{-1}, y_i = 0 \cup \mathbb{N}$. The relationship between the mean and regressors is shown by the link function $\mu_i = exp(\mathbf{x}_i'\beta)$. It is worth noting that empirical data usually shows more variation than can be accounted by such Poisson models.

In order to account for unobserved heterogeneity and the correlation of events in the observed data, Winkelmann [8] derived several compound Poisson models, taking additional unobserved heterogeneity into consideration by letting $\lambda_i = exp(x_i\beta + \epsilon_i) = exp(x_i\beta)\delta_i$. They applied the model to labor mobility data and their results illustrate the necessity to allow for the generalizations of the standard Poisson regression model. Coxe et al. [9] provided a clear review of some appropriate regression models applicable to count data. Starting with the standard Poisson regression model, two variants of Poisson regression, negative binomial regression, and over-dispersed Poisson regression, were formed to handle the over-dispersion phenomenon. A comparison of those models, using a simulated data set of drinks consumed by university students on Saturday night [10], demonstrated the strengths and weaknesses of each of these models.

In order to generalize the relationship between the mean and the variance imposed by the Poisson regression models, Linden and Mantyniemi [11] utilized a negative binomial formulation. In their approach, two parameters were introduced to accommodate different 'quadratic mean-variance relationships' [11]. They expressed the probability mass function of the random variable X as $P(X = x|r, p) = \Gamma(x + r)p^r(1 - p)^x[x!\Gamma(r)]^{-1}$, with the expectation (theoretical mean) $\mu = r(1 - p)p^{-1}$ and variance $\sigma^2 = r(1 - p)p^{-2}$. Based

on this setup, parameters r and p could be solved from their relationship with the mean and the variance, as $r = \mu^2[\sigma^2 - \mu]^{-1}$ and $p = \mu\sigma^{-2}$. The negative binomial distribution allows more flexible parameterization, which could be used to represent multiple types of over-dispersed Poisson type processes. By establishing a quadratic function of the mean to describe the variance, $\sigma^2 = \omega\mu + \theta\mu^2$, diverse relationships between the mean and variance can be obtained by varying the two over-dispersion parameters ω and θ as long as the condition $\sigma^2 > \mu$ is satisfied. Scenarios where over-dispersion might happen due to factors such as sampling, environmental dissimilarity, or flocking behavior were exemplified using bird migration data showing a high level of over-dispersion. In this study, the negative binomial distribution, using well-selected over-dispersion parameters, appropriately represented the mean-variance relationships in the considered scenarios. However, distinct assumptions about mean-variance relationships could lead to completely different coefficients, which need careful identification and interpretation.

Albert [12], taking advantage of the Poisson distribution and hidden Markov chain, developed the Poisson hidden Markov model (HMM) to handle count data with different status at different periods and applied to a time series of epileptic seizure counts [12]. It assumes the latent Poisson processes that generate the count data have significantly diverse means at different times. However, the Poisson processes are only determined by the status of the underlying Markov chain and conditioning on the past states. The current Markov chain state S_t only depends on the previous state S_{t-1}. As a result, the dependency amongst data points is not taken care of since a lot of time series demonstrate a characteristic that a high count would often be followed by several high counts.

The class of discrete-valued time series models analogous to Gaussian autoregressive moving average (ARMA) models were advocated by Jacobs and Lewis [2]. The data generating process they assumed was a probabilistic linear combination of independently and identically distributed discrete random variables [13]. Two simple stationary processes of discrete random variables, DARMA $(p, N+1)$ and NDARMA (p, N), whose first-order

marginal distributions are arbitrarily chosen, are listed in [14]. One major drawback of the DARMA model is that even if the distribution of independent identically distributed variables is continuous, the sequence has a high density around a single value.

In the article by Heinen [15], an autoregressive conditional Poisson (ACP) model with an autoregressive mean structure was developed, which is also independently proposed by Rene [16] as the integer-valued generalized autoregressive conditional heteroskedastic (INGARCH) (p, q) process to model integer-valued data with Poisson deviates. In the article, important conditions for the existence of the mentioned process were discussed. When it comes to the situation $p = 1$, $q = 1$, such an integer-valued GARCH process is, in essence, a standard ARMA (1, 1) model. The asymptotic properties of the maximum likelihood estimates of model parameters were studied. The numbers of people infected by Campylobacterosis (a bacteria caused disease) over a certain period was analyzed by the observation driven model, and a one-step ahead forecast was also provided.

After the above autoregressive Poisson model was proposed, a negative binomial version was built on this structure. Zhu [17] developed the negative binomial integer GARCH (NBINGARCH) model and discussed some properties of it. By letting $(1 - p_t)[p_t]^{-1} = \lambda_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i X_{t-i} + \sum_{j=1}^{p} \beta_j \lambda_{t-j}$, the negative binomial INGARCH could deal with problems that occur when fitting over-dispersed data by a Poisson based INGARCH model. The model could also handle potential extreme observations.

Another common feature in time series count data is the excessive zeros. The zero-inflated Poisson model [18] was introduced to deal with this type of data. The model assumes that the observed zeros arise from two processes: one from a binary process with probability π of getting a zero and another from an ordinary Poisson process. Later, Yang [19] linked the parameters of the rate of the Poisson process and the probability of the binary process with exogenous explanatory covariates. Möller [20] extended the binomial thinning AR(1) process by adding a bounded support to accommodate different zero-inflated data

types. They assumed the observation at time t is generated from two thinning processes $X_t = \alpha * X_{t-1} + \beta * (n - X_{t-1})$, where α and β are the probabilities of getting a one in a Bernoulli trial and n is the maximum value that X_t could take.

In the next section, ACP, Poisson HMM, and zero-inflated Poisson model are briefly introduced.

2.1. THE AUTOREGRESSIVE CONDITIONAL POISSON (ACP) MODEL

In the article by Heinen [15], an autoregressive conditional Poisson (ACP) model that handles count data from a Poisson distribution with an autoregressive mean was developed. At the same time, the model was also independently proposed by Ferland *et. al* [16] as the integer-valued generalized autoregressive conditional heteroskedasticity (INGARCH) process.

2.1.1. The ACP Model. Let $\mathcal{F}_t$ represent all information available before and including time t. Conditioning on the past information, the count present follows a Poisson distribution with a mean μ_t related to the past such that,

$$N_t|\mathcal{F}_{t-1} \sim Pois(\mu_t),$$

where the mean has an autoregressive conditional intensity structure inspired by the conditional variance formulation in GARCH [21] model of Bollerslev [22]. In the ACP model,

$$E[N_t|\mathcal{F}_{t-1}] = \mu_t = \omega + \sum_{i=1}^{q} \alpha_i N_{t-i} + \sum_{j=1}^{p} \beta_j \mu_{t-j},$$

under the condition that all of $\alpha_i's$, $\beta_j's$ and ω are positive.

2.1.2. Parameter Estimation of ACP Models. The Maximum Likelihood Estimation method is used to provide estimates of the parameters of the ACP models. Let $\theta \equiv (\omega, \alpha_i, \beta_j)$ for $i = 1, \ldots, q$, $j = 1, \ldots, p$, represent all parameters in the ACP model

and the likelihood function at time t is

$$l_t(\boldsymbol{\theta}) = N_t \ln \mu_t - \mu_t - \ln N_t!.$$

The corresponding score function and Hessian matrix take the form:

$$\frac{\partial l_t}{\partial \boldsymbol{\theta}} = \left(\frac{N_t - \mu_t}{\mu_t}\right) \frac{\partial \mu_t}{\partial \boldsymbol{\theta}},$$

$$\frac{\partial^2 l_t}{\partial \boldsymbol{\theta}^2} = \left(-\frac{N_t}{\mu_t^2}\right) \left(\frac{\partial \mu_t}{\partial \boldsymbol{\theta}}\right) \left(\frac{\partial \mu_t}{\partial \boldsymbol{\theta}}\right)',$$

where

$$\frac{\partial \mu_t}{\partial \boldsymbol{\theta}} = \boldsymbol{V}_t' + \sum_{j=1}^{p} \beta_j \frac{\partial \mu_{t-j}}{\partial \boldsymbol{\theta}},$$

and

$$\boldsymbol{V}_t = [1, N_{t-1}, N_{t-2}, \ldots, N_{t-q}, \mu_{t-1}, \mu_{t-2}, \ldots, \mu_{t-p}].$$

2.2. THE POISSON HIDDEN MARKOV MODEL

The Poisson hidden Markov model (HMM) is first discussed by Hopkins [23]. Later the model was used by Alber [12] to handle the epileptic seizure counts by using a Poisson process whose mean varies at times. The model assumes the underlying mechanism of the count data changes states determined by a Markov chain, and also accommodates over-dispersion characteristics, which is commonly observed in count data time series.

2.2.1. The Poisson-HMM Model. Let a sequence of discrete random variables $S_t : t \in \mathbb{N}$ to be a Markov Chain with m possible states and transition probability matrix $\boldsymbol{\Gamma}(t) = \{\gamma_{ij}^s(t)\}$, $i = 1, 2, \ldots, m, j = 1, 2, \ldots, m$, where $\gamma_{ij}^s(t) = P(S_{s+t} = j | S_s = i)$. In most cases, it is enough to use homogeneous Markov chains, which means γ_{ij}^s does not

depend on s. Unless there is an explicit indication, it is assumed that Markov chain under discussion is a homogeneous one with transition probabilities denoted as γ_{ij}. Given the underlying Markov state S_t, the observed count data N_t follows a Poisson distribution

$$P(N_t = k|S_t = i) = \frac{e^{-\lambda_i}\lambda_i^k}{k!},$$

where λ_i is the expected mean of the state dependent Poisson process when the corresponding latent Markov chain takes state i.

2.2.2. Parameter Estimation of Poisson HMM Models. The direct maximization of the likelihood could be accomplished by the following method.

Let $\boldsymbol{\theta} \equiv (\gamma_{ij}, \lambda_i)$ for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, m$, represents all parameters in the Poisson hidden Markov model. The log-likelihood function for the model is given by

$$l_T(\boldsymbol{\theta}) = log\Big(P(\boldsymbol{X}_T = \boldsymbol{x}_T)\Big) = log(\boldsymbol{\delta}\boldsymbol{P}(x_1)\boldsymbol{\Gamma}\boldsymbol{P}(x_2)\cdots\boldsymbol{\Gamma}\boldsymbol{P}(x_T)\mathbf{1}'),$$

where $\boldsymbol{\delta}$ is the initial distribution and

$$\boldsymbol{P}(x_t) = \begin{bmatrix} p_1(x_t) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \vdots & p_m(x_t) \end{bmatrix},$$

$$p_i(x_t) = P(X_t = x_t|S_t = i),$$

$$\boldsymbol{\Gamma} = \begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1m} \\ \vdots & \ddots & \vdots \\ \gamma_{m1} & \cdots & \gamma_{mm} \end{bmatrix}.$$

For the discrete case, elements in the likelihood function become progressively smaller as t increases, and thus scaling the forward probabilities is a common way to avoid underflow. The scaling yields:

$$\boldsymbol{\beta}_0 = \boldsymbol{\delta}\boldsymbol{P}(x_1),$$

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1}\boldsymbol{\Gamma}\boldsymbol{P}(x_t), \quad for\ t = 2, 3, \ldots, T.$$

$$\boldsymbol{\phi}_0 = \boldsymbol{\delta},$$

$$\boldsymbol{\phi}_t = \frac{\boldsymbol{\beta}_t}{\omega_t},$$

$$\omega_t \boldsymbol{\phi}_t = \omega_{t-1}\boldsymbol{\phi}_{t-1}\boldsymbol{B}_t,$$

where

$$\omega_t = \sum_i \beta_t(i) = \boldsymbol{\beta}_t \mathbf{1}',$$

$$\omega_0 = \boldsymbol{\delta}\mathbf{1}',$$

Thus the scaled log-likelihood function would be

$$log(L_T) = \sum_{t=1}^{T} log\frac{\omega_t}{\omega_{t-1}} = \sum_{t=1}^{T} log(\boldsymbol{\phi}_{t-1}\boldsymbol{B}_t\mathbf{1}').$$

2.3. ZERO INFLATED POISSON MODEL

Heinen [15] introduced an autoregressive conditional Poisson (ACP) model to deal with time series of count data with serial correlation. Building on Heinen's work, the theoretical properties of the general ACP model was derived by Ghahramani and Thavaneswaran [24]. In brief, the count X_t observed during the interval (t-1, t] is considered a realization from a Poisson distribution with a conditional mean that is dependent on past observations and past conditional means. The process is formally defined as follows. Let the count data series be denoted by $\{X_t : t \in \mathbb{N}\}$ and let $\mathcal{F}_t$ denote the sigma field generated by the set of random variables $\{X_i : i \leq t\}$. The counts are assumed to be realizations from a Poisson distribution with a conditional mean λ_t, whose dependence on past conditional means and

counts parallels the GARCH model of Bollerslev [21]:

$$X_t|\mathcal{F}_{t-1}\sim P(\lambda_t)$$

$$E\left(X_t \mid \mathcal{F}_{t-1}\right) = \lambda_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i X_{t-i} + \sum_{j=1}^{p} \beta_j \lambda_{t-j},$$

under the condition that all of $\alpha_i's, \beta_j's$ are positive.

The above formulation is suitable for modeling discrete time series with over-dispersion while taking serial correlation into consideration, with the latter property being of particular interest given that temporal dependence is not uncommon in times series of count data.

Zhu [25] proposed a generalized version of the ACP model by replacing the Poisson assumption with a zero-inflated Poisson (*ZIP*) distribution that has a zero-inflation parameter ω. A $ZIP\left(\lambda, \omega\right)$ distribution can be characterized by the probability mass function:

$$P\left(X=k\right) = \omega\delta_{k,0} + \left(1-\omega\right)\frac{e^{-\lambda}\lambda^k}{k!},\ k\in\mathbb{N}\cup\{0\},$$

where $0 < \omega < 1$ and $\delta_{k,0}$ is the Kronecker delta such that

$$\delta_{k,0} = \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{if } k \neq 0 \end{cases}.$$

Thus, under the formulation of Zhu [25], the time series $\{X_t : t \in \mathbb{N}\}$, conditional on $\mathcal{F}_{t-1}$, satisfy

$$X_t|\mathcal{F}_{t-1}\sim ZIP\left(\lambda_t, \omega\right),$$

$$\lambda_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i X_{t-i} + \sum_{j=1}^{p} \beta_j \lambda_{t-j},$$

where $0 < \omega < 1$, $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, for $i = 1, 2, \ldots, q$ and $j = 1, 2, \ldots, p$.

PAPER

I. MODELING TIME SERIES OF COUNT DATA USING A PERIODIC AUTOREGRESSIVE CONDITIONAL POISSON MODEL

Yi Zhang, V Samaranayake
Department of Mathematics, Missouri University of Science and Technology, Rolla, MO 65409, USA
Email: zybvf@umsystem.edu

ABSTRACT

A periodic version of the autoregressive conditional Poisson model (ACP), introduced by Heinen [1] in 2003, is proposed. In the ACP model, the conditional mean of the Poisson process at a given time is assumed to follow a formulation that links it to past counts and past means. The proposed periodic autoregressive conditional Poisson (PACP) model assumes that the data are generated by Poisson process whose conditional mean follows an ACP model with parameters that vary seasonally. Such models would be more appropriate when modeling count data series exhibiting conditional heteroskedastic behavior that varies from season to season. Properties of the model are investigated, and an alternative format of the model is presented to make it comparable to a vector autoregressive moving average (ARMA) process. A Monte Carlo simulation study that employs the maximum likelihood method to estimate the parameters shows an accurate estimation of the parameters with a relatively small Monte Carlo standard error. The simulation study also investigated the use of Akaike information criterion (AIC) and Bayesian information criterion (BIC) criteria to

differentiate between periodic and non-periodic cases with promising results. An analysis of simulated data is used to illustrate the importance of identifying the true structure of time series count data with periodic behavior and potential for the wide uses of such models.

Keywords: count data, discrete time series, seasonality, conditional heteroskedasticity, time varying parameters.

1. INTRODUCTION

Advanced data collection technologies are generating numerous time series of count data that exhibit periodic behavior. Examples of such time series range from the number of transactions per minute involving a given stock to the number of hourly clicks on a website. While traditional approaches such as Poisson regression can handle many of these time series, some count series exhibit clustering of high counts, similar to volatility clustering found in stock return series. For example, such clustering is seen in incident counts of common infectious diseases, where a high prevalence of the disease during the recent past gives rise to higher counts during the next data gathering period. Among count data models, the autoregressive conditional Poisson (ACP) model proposed by Heinen [1] allows for such behavior, specifically because the structure of the ACP formulation is very similar to the generalized autoregressive conditional heteroskedastic (GARCH) processes that are used to model economic data with volatility clustering. In addition, the ACP model also makes it possible to analyze discrete correlated data with over-dispersion. However, the ACP model is not structured to capture periodic behavior inherent in some count data series. Thus, we proposed a generalized version of the ACP model, namely the periodic autoregressive conditional Poisson (PACP) model, which could accommodate such characteristics.

2. REVIEW OF MODELS FOR TIME SERIES COUNT DATA

Markov chains are one way to deal with count data [2]. This method requires the definition of the (stationary) transition probabilities between all possible outcomes that the random variable could generate. However, a sequence $X_1, X_2, \ldots$ of random variables taking values in some set is a Markov chain if the conditional distribution of X_{n+1}, given $\{X_i : i \in \mathbb{N}\}$, depends on X_n only, which would limit the use of this model because it ignores dependence on values prior to time n. Also, when the number of possible outcomes grows very large, this model is no longer easily tractable, and its parameter estimation becomes cumbersome.

An alternative is to use a hidden Markov model. These models are a modified application of Markov chains as it is assumed that an underlying unobserved state of the system, determined by a Markov process, changes in time. The system's present state should determine the distribution of observations at the current time [3]. Sebastian et al. [4] developed the Markov ordinal logistic regression model with the transition probability defined as $P(Y_t|Y_{t-1}, \mathbf{Z}) = (e^{\alpha_i - \mathbf{Z}'\beta})[1 - e^{\alpha_i - \mathbf{Z}'\beta}]^{-1}$, where Y_i represents the states of the Markov chain, while $\mathbf{Z}$ denotes some known covariates. Cooper et al. [5] proposed a so-called 'structured hidden Markov model' for the epidemic process that intuitively follows a hidden Markov chain process since patients communicate with each other and the epidemic process usually gives out routine surveillance data that could often be partially observed. 'Structured' implies that a simple transition model is driving the underlying Markov chain. However, the need to determine the order of the Markov chain before applying the model accounts for one obvious drawback of this type of model. Another inevitable problem of such models is that the variability of the outcomes may be small.

Another branch of methods developed from the application of the Markov chain is the binomial thinning process proposed by McKenzie [6] as a simple model to deal with discrete variate time series problems. The thinning operator takes the sum of $\{X_i : i \in \mathbb{N}\}$ identically independent Bernoulli random variables, each of which takes value 1 with

probability α and 0 with probability $1-\alpha$. The data generating mechanism is modeled similar to an autoregressive (AR) process in the sense that the current count is dependent on the number of Bernoulli random variables given by the previous count. For example, Poisson AR(1) process is constructed as $X_n = \alpha * X_{n-1} + W_n$, where X_n and W_n are both Poisson process with means θ and $\theta(1-\alpha)$ respectively. The thinning operator $*$ is defined as follows: $\alpha * X_{n-1}$ denotes the number of successes observed from X_{n-1} Bernoulli trails with success probability α. Geometric AR(1), negative binomial AR(1), binomial AR(1), and compound correlated bi-variate Poisson distribution were also proposed. They also investigated the seasonality problem in count data, and the seasonal mean was set as $\mu_n = a\cos\omega n + b\sin\omega n$, similarly the innovation mean was set as $\omega_n = A\cos\omega n + B\sin\omega n$, where $A = a - \alpha(a\cos\omega - b\sin\omega)$ and $B = b - \alpha(a\sin\omega + b\cos\omega)$. Zhu and Joe [7] further modified the model, incorporating covariates to the mean of the stationary Markov time series allowing time varying components. Also, they extend the model and the structure to mimic an AR(2) model to solve higher order dependence structure contained in time series count data. The problem with this kind of model is that the seasonal pattern embedded in this fashion is not flexible enough to model data demonstrating complex periodic components that may require the inclusion of a large number of trigonometric functions to model the cyclical behavior with reasonable accuracy.

Many count data models are based on the use of the Poisson distribution. The Poisson regression model, as the basic count regression model, is well described in the book by Cameron and Trivedi [8]. In this approach it is assumed that y_i, an independent observation from a Poisson distribution, given the vector of regressor $\boldsymbol{x}_i$, has a density function $f(y_i|\boldsymbol{x}_i) = (e^{-\mu_i}{\mu_i}^{y_i})[y_i!]^{-1}, y_i \in 0 \cup \mathbb{N}$. The relationship between the mean and regressors is shown by the link function $\mu_i = exp(\boldsymbol{x}_i'\boldsymbol{\beta})$. It is worth noting that empirical data usually shows more variation than can be accounted by such Poisson models. In other words, empirical data shows over-dispersion.

In order to account for unobserved heterogeneity and the correlation of events in the observed data, Winkelmann [9] derived several compound Poisson models, taking additional unobserved heterogeneity into consideration by letting $\lambda_i = exp(\boldsymbol{x_i}'\boldsymbol{\beta} + \epsilon_i) = exp(\boldsymbol{x_i}'\boldsymbol{\beta})\mu_i$. They applied the model to labor mobility data and their results illustrate the necessity to allow for the generalizations of the standard Poisson regression model. Coxe et al. [10] provided a clear review of some appropriate regression models applicable to count data. Starting with the standard Poisson regression model, two variants of Poisson regression, negative binomial regression and over-dispersed Poisson regression are gradually formed to handle the over-dispersion phenomenon. A comparison among those models, using a simulated data set of drinks consumed by university students on Saturday night [11], demonstrates the strengths and weaknesses of these models.

In order to generalize the relationship between the mean and the variance imposed by the Poisson regression models, Linden and Mantyniemi [12] utilized a negative binomial formulation. In their approach, two parameters were introduced to accommodate different 'quadratic mean–variance relationships' [12]. They expressed the probability mass function of the random variable X as $P(X = x|r, p) = \Gamma(x + r)p^r(1 - p)^x[x!\Gamma(r)]^{-1}$, with the expectation (theoretical mean) $\mu = r(1 - p)p^{-1}$ and variance $\sigma^2 = r(1 - p)p^{-2}$. Based on this setup, parameters r and p could be solved from their relationship with the mean and the variance, as $r = \mu^2[\sigma^2 - \mu]^{-1}$ and $p = \mu\sigma^{-2}$. The negative binomial distribution allows more flexible parameterization, which could be used to represent multiple types of over-dispersed Poisson processes. By establishing a quadratic function of the mean to describe the variance, $\sigma^2 = \omega\mu + \theta\mu^2$, diverse relationships between the mean and variance can be obtained by varying the two over-dispersion parameters ω and θ, as long as the condition $\sigma^2 > \mu$ is satisfied. Scenarios where over-dispersion might happen due to factors such as sampling, environmental dissimilarity, or flocking behavior were exemplified using bird migration data showing a high level of over-dispersion. In this study, the negative binomial distribution, using well-selected over-dispersion parameters, appropriately represented the

mean-variance relationships in the considered scenarios. However, distinct assumptions about mean-variance relationships could lead to completely different coefficients which need careful identification and interpretation.

The class of discrete-valued time series models analogous to Gaussian ARMA models were advocated by Jacobs and Lewis [3]. The data generating process they assumed was a probabilistic linear combination of independently and identically distributed discrete random variables [13]. Two simple stationary processes of discrete random variables, DARMA (p, $N + 1$) and NDARMA (p, N), whose first-order marginal distributions are arbitrarily chosen, are listed in [14]. One major drawback of the DARMA model is that even if the distribution of independent identically distributed variables is continuous, the sequence has a high density around a single value.

The autoregressive conditional Poisson (ACP) model [1], or simultaneously independently proposed by Rene [15] as the integer-valued generalized autoregressive conditional heteroscedastic (INGARCH) (p,q) process, is developed to model integer-valued data with Poisson deviates. In the article, important conditions for the existence of the mentioned process were discussed. When it comes to the situation $p = 1, q = 1$, such an integer-valued GARCH process is, in essence, a standard ARMA (1, 1) model. The asymptotic properties of the maximum likelihood estimates of model parameters were studied. The numbers of people infected by Campylobacterosis (a bacteria caused disease) over a certain period was analyzed by the observation driven model, and a one-step ahead forecast was also provided.

After the above ACP model was proposed, a negative binomial version was built on this structure. Zhu [16] developed the negative binomial integer GARCH (NBINGARCH) model and discussed some properties of it. By letting $(1-p_t)p_t^{-1} = \lambda_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j}$, the negative binomial INGARCH could deal with problems that occur when fitting over-dispersed data by a Poisson based INGARCH model. The model could also handle potential extreme observations.

Another common feature in time series count data is the excessive zeros. The zero-inflated Poisson model [17] was introduced to deal with this type of data. The model assumes excessive zeros is the outcome from two processes: one as a binary process with probability π of getting a zero and another as an ordinary Poisson process. Later, Yang [18] linked the parameters of the rate of the Poisson process and the probability of the binary process with exogenous explanatory covariates. Möller [19] extended the binomial thinning AR(1) process by adding a bounded support to accommodate different zero-inflated data types. They assumed the observation at time t is generated from two thinning processes $X_t = \alpha * X_{t-1} + \beta * (n - X_{t-1})$, where α and β are the probabilities of getting a one in a Bernoulli trial and n is the maximum value that X_t could take.

3. PROPOSED PACP MODELS

In the article by Heinen [1], an autoregressive conditional Poisson (ACP) model that handles count data from a Poisson distribution with an autoregressive mean was developed. Let $\mathcal{F}_t$ represent all information available before and including time t. Conditioning on the past information, the count present follows a Poisson distribution with a mean μ_t related to the past,

$$N_t|\mathcal{F}_{t-1} \sim P(\mu_t),$$

where the mean has an autoregressive conditional intensity structure inspired by conditional variance in GARCH [20] model of Bollerslev [21]. In the ACP model,

$$E[N_t|\mathcal{F}_{t-1}] = \mu_t = \omega + \sum_{i=1}^{q} \alpha_i N_{t-i} + \sum_{j=1}^{p} \beta_j \mu_{t-j}, \tag{1}$$

under the condition that all of $\alpha_i's$, $\beta_j's$ and ω are positive.

However, there are empirical count data series that demonstrate periodic characteristics. Therefore, a periodically varying coefficient autoregressive conditional Poisson model would be more appropriate under such circumstances. Thus, we generalize the ACP Model to a periodic autoregressive conditional Poisson model (PACP), which provides more flexibility when modeling periodic count data.

To define the desired structure, let $\{N_t : t \in \mathbb{N}\}$ be the time series of interest, with N_t denoting the count at time t. We assume that t falls into one of s periods that recur in a periodic fashion, and let $s(t)$ denote the period to which t belongs. Denoting the $\sigma-$ algebra generated by $\{N_i : i \leq t\}$ as $\mathcal{F}_t$. Assume that

$$N_t|\mathcal{F}_{t-1} \sim P(\lambda_t),$$

with the mean having a time varying structure

$$\lambda_t = \omega_{s(t)} + \sum_{i=1}^{q} \alpha_{is(t)} N_{t-i} + \sum_{j=1}^{p} \beta_{js(t)} \lambda_{t-j}, \tag{2}$$

where $\omega'_{s(t)}s$, $\alpha'_{is(t)}s$, $\beta'_{js(t)}s$ and $i = 1, 2, , \ldots, q$, $j = 1, 2, \ldots, p$ are positive for all values of $s(t)$. Note the $s(t)$ represents the corresponding stage of the periodic cycle at time t. Note that the above formulation parallels that of a periodic GARCH(p,q) process [21].

4. SOME PROPERTIES OF THE MODEL

Following the derivations by Bollerslev [21], the equation (1) could be rewritten as

$$N_t = \epsilon_t + \omega_{s(t)} + \sum_{i=1}^{\max(p,q)} (\alpha_{is(t)} + \beta_{is(t)}) N_{t-i} - \sum_{j=1}^{p} \beta_{js(t)} \epsilon_{t-j},$$

where $\epsilon_t = N_t - \lambda_t$.

Now, model in equation (2) could be interpreted as an ARMA(max(p,q), p) process instead of a GARCH(p, q) process. Some properties of ϵ_t are discussed below. We have

$$E(\epsilon_t) = E(N_t - \lambda_t) = E(E(N_t - \lambda_t|\mathcal{F}_{t-1})) = E(\lambda_t - \lambda_t) = 0,$$

and

$$\begin{aligned} Var(\epsilon_t) &= Var\left(E(\epsilon_t|\mathcal{F}_{t-1})\right) + E\left(Var(\epsilon_t|\mathcal{F}_{t-1})\right) \\ &= Var\left(0\right) + E\left(Var(\epsilon_t|\mathcal{F}_{t-1})\right) \\ &= Var\left(0\right) + E\left(Var(N_t - \lambda_t|\mathcal{F}_{t-1})\right) \\ &= E\left(Var(N_t - \lambda_t|\mathcal{F}_{t-1})\right) \\ &= E\left(Var(N_t|\mathcal{F}_{t-1})\right) \\ &= E\left(\lambda_t\right). \end{aligned}$$

We considered two cases: (1) Periodic data with a single observation within each period and (2) Periodic data with multiple observations within each period. For the first case, a Vector ARMA form of the time series is derived analogous to [21], and this form is derived in the appendix.

5. LIKELIHOOD FUNCTION, SCORE, HESSIAN AND PARAMETER ESTIMATION

Let $\boldsymbol{\theta} \equiv (\omega_{s(t)}, \alpha_{is(t)}, \beta_{js(t)})$ for $i = 1, \ldots, q$, $j = 1, \ldots, p$, represent all parameters in the PACP model. The conditional log-likelihood function for the model could be written as the sum of log-likelihood for each observation. Thus we have

$$l_T(\boldsymbol{\theta}) = \sum_{t=1}^{T} -\lambda_t(\boldsymbol{\theta}) + N_t \log \lambda_t(\boldsymbol{\theta}) - \log(N_t!). \tag{3}$$

The corresponding score function and Hessian matrix are

$$\frac{\partial l_T}{\partial \boldsymbol{\theta}} = \sum_{t=1}^{T} -\frac{\partial \lambda_t}{\partial \boldsymbol{\theta}} + \frac{N_t}{\lambda_t}\left(\frac{\partial \lambda_t}{\partial \boldsymbol{\theta}}\right),$$

$$\frac{\partial^2 l_T}{\partial \boldsymbol{\theta}^2} = \sum_{t=1}^{T} \left(-\frac{N_t}{\lambda_t^2}\right)\left(\frac{\partial \lambda_t}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \lambda_t}{\partial \boldsymbol{\theta}}\right)',$$

where

$$\frac{\partial \lambda_t}{\partial \boldsymbol{\theta}} = \boldsymbol{V}_t' + \sum_{j=1}^{p} \beta_{js(t)} \frac{\partial \lambda_{t-j}}{\partial \boldsymbol{\theta}},$$

and

$$\boldsymbol{V}_t = [1, N_{t-1}, N_{t-2}, \ldots, N_{t-q}, \lambda_{t-1}, \lambda_{t-2}, \ldots, \lambda_{t-p}].$$

6. THE MONTE-CARLO SIMULATION STUDY

We conducted a Monte-Carlo simulation study to investigate how well the PACP model parameters are estimated by the MLE procedure. A simulation study was also performed to investigate the use of AIC and BIC criteria to differentiate between periodic and non-periodic cases. The maximum likelihood method is utilized to estimate the parameters of PACP model. The log-likelihood function is defined as (3).

The properties of estimates were studied across different combinations of parameters using 3,000 simulation runs for each combination. Biases and Monte Carlo standard errors (SE) were computed for each of the parameter combinations. In order to eliminate the artifacts arising out of initial conditions, the first 240 time series data points were discarded.

6.1. CASE OF A SINGLE OBSERVATION WITHIN A PERIOD.

6.1.1. Two Periods with Single Observation within a Period. Let $\theta = (\omega_1, \omega_2, \alpha_1, \alpha_2, \beta)$, represent all parameters in the PACP model. The simulated data were generated from a PACP process of 2 periods with $p, q = 1$ and only a single data point within each period. Two different parameter sets were used for analysis.

Set 1. $\theta = (3, 5, 0.1, 0.3, 0.1)$ (Table 1).

Set 2. $\theta = (10, 8, 0.25, 0.35, 0.2)$ (Table 2).

For each combination of parameter sets, sample size T=500 and T=1,000 were considered. Note that the time series lengths T=500 and 1,000 are comparable to the lengths of series of day and night counts of a given phenomenon over a few years. Maximum likelihood estimation results from 3,000 simulations based on the above sample sizes with a single observation within a period are reported in Tables 1 and 2.

Table 1. Maximum likelihood estimation results from 3,000 simulations based on different sample sizes (single observation within a period); parameter Set 1.

Parameters	True Values	T=500			T=1,000		
		Estimates	Bias	SE	Estimates	Bias	SE
ω_1	3	2.7783	-0.2217	0.0194	2.9019	-0.0981	0.0142
ω_2	5	4.8543	-0.1457	0.0130	4.9279	-0.0721	0.0097
α_1	0.1	0.0989	-0.0011	0.0009	0.0983	-0.0017	0.0007
α_2	0.3	0.2985	-0.0015	0.0015	0.2990	-0.0010	0.0011
β	0.1	0.1345	0.0345	0.0030	0.1166	0.0166	0.0021

Table 2. Maximum likelihood estimation results from 3,000 simulations based on different sample sizes (single observation within a period); parameter Set 2.

Parameters	True Values	T=500			T=1,000		
		Estimates	Bias	SE	Estimates	Bias	SE
ω_1	10	10.2290	0.2290	0.0394	10.0930	0.0930	0.0285
ω_2	8	8.2398	0.2398	0.0407	8.0681	0.0681	0.0298
α_1	0.25	0.2484	-0.0016	0.0012	0.2500	0.0000	0.0008
α_2	0.35	0.3477	-0.0023	0.0012	0.3511	0.0011	0.0008
β	0.2	0.1889	-0.0111	0.0023	0.1949	-0.0051	0.0017

6.1.2. Three Periods with Single Observation within a Period. Let $\theta = (\omega_1, \omega_2, \omega_3, \alpha_1, \alpha_2, \alpha_3, \beta)$, represent all parameters in the PACP model. The simulated data were generated from a PACP process of 3 periods with $p, q = 1$ and only a single data point within each period. Two different parameter sets were used for analysis.

Set 1. $\theta = (5, 8, 3, 0.4, 0.2, 0.2, 0.3)$ (Table 3).

Set 2. $\theta = (8, 9, 5, 0.1, 0.3, 0.2, 0.1)$ (Table 4).

For each combination of parameter sets, sample size T=1,040 and T=2,080 were considered. Note that the time series lengths T=1,040 is comparable to the lengths of series of morning, mid-day, and night counts of a given phenomenon over a year. Maximum likelihood estimation results from 3,000 simulations based on the above sample sizes with a single observation within a period are reported in Tables 3 and 4. Results show reasonably accurate parameters estimates with minimal bias and low standard error, with lower biases in the large sample case.

Table 3. Maximum likelihood estimation results from 3,000 simulations based on different sample sizes (single observation within a period); parameter Set 1.

Parameters	True Values	T=1,020			T=2,040		
		Estimates	Bias	SE	Estimates	Bias	SE
ω_1	5	5.1430	-0.1430	0.0288	5.0520	-0.0520	0.0113
ω_2	8	8.1236	-0.1236	0.0289	8.0307	-0.0307	0.0108
ω_3	3	3.1409	-0.1409	0.0323	3.0459	-0.0459	0.0189
α_1	0.4	0.1982	0.0018	0.0001	0.1980	0.0020	0.0001
α_2	0.2	0.4002	-0.0002	0.0001	0.4001	-0.0001	0.0001
α_3	0.2	0.2022	-0.0022	0.0001	0.2004	-0.0004	0.0001
β	0.3	0.2883	0.0117	0.0002	0.2969	0.0031	0.0001

Table 4. Maximum likelihood estimation results from 3,000 simulations based on different sample sizes (single observation within a period); parameter Set 2.

Parameters	True Values	T=1,020			T=2,040		
		Estimates	Bias	SE	Estimates	Bias	SE
ω_1	8	7.7426	0.2574	0.0292	7.8935	0.1065	0.0137
ω_2	9	8.7822	0.2178	0.0354	8.8883	0.1117	0.0172
ω_3	5	4.6857	0.3143	0.0365	4.8889	0.1111	0.0240
α_1	0.1	0.1017	-0.0017	0.0001	0.1005	-0.0005	0.0001
α_2	0.3	0.2955	0.0045	0.0001	0.2993	0.0007	0.0001
α_3	0.2	0.1982	0.0018	0.0001	0.1972	0.0028	0.0001
β	0.1	0.1261	-0.0261	0.0003	0.1114	-0.0114	0.0002

6.2. CASE OF MULTIPLE OBSERVATIONS WITHIN A PERIOD

6.2.1. Daily Data with Multiple Observation within a Period. Let $\theta = (\omega_1, \omega_2, \omega_3, \omega_4, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \beta)$, represent all parameters in the PACP model. The simulated data were generated from a PACP process of 4 periods with $p, q = 1$ and 90 data points within each period. Three different parameter sets were used for analysis.

Set 1. $\theta = (2, 5, 3, 4, 0.1, 0.05, 0.2, 0.2, 0.1)$ (Table 5).

Set 2. $\theta = (10, 7, 5, 12, 0.2, 0.3, 0.1, 0.3, 0.3)$ (Table 6).

Set 3. $\theta = (4, 6, 5, 4, 0.1, 0.2, 0.1, 0.2, 0.4)$ (Table 7).

For each combination of parameter sets, sample size T=540 and T=1,080 were considered. These are approximately equivalent to the numbers of observations for a 1.5-year and 3-year daily count data respectively. Maximum likelihood estimation results from 3,000 simulations based on different sample sizes are reported in Tables 5-7.

Table 5. Maximum likelihood estimation results from 3,000 simulations based on different sample sizes (multiple observations within a period); parameter Set 1.

Parameters	True Values	T=540			T=1,080		
		Estimates	Bias	SE	Estimates	Bias	SE
ω_1	2	1.9741	-0.0259	0.0074	1.9907	-0.0093	0.0057
ω_2	5	4.8703	-0.1297	0.0170	4.9391	-0.0609	0.0131
ω_3	3	3.0080	0.0080	0.0133	3.0010	0.0010	0.0095
ω_4	4	4.0179	0.0179	0.0174	3.9968	-0.0032	0.0126
α_1	0.1	0.0935	-0.0065	0.0012	0.0957	-0.0043	0.0010
α_2	0.05	0.0555	0.0055	0.0010	0.0525	0.0025	0.0009
α_3	0.2	0.1812	-0.0188	0.0018	0.1923	-0.0077	0.0011
α_4	0.2	0.1807	-0.0194	0.0019	0.1931	-0.0069	0.0011
β	0.1	0.1169	0.0169	0.0028	0.1074	0.0074	0.0021

Table 6. Maximum likelihood estimation results from 3,000 simulations based on different sample sizes (multiple observations within a period); parameter Set 2.

Parameters	True Values	T=540			T=1,080		
		Estimates	Bias	SE	Estimates	Bias	SE
ω_1	10	10.6510	0.6510	0.0433	10.3380	0.3380	0.0298
ω_2	7	7.5971	0.5971	0.0358	7.2991	0.2991	0.0248
ω_3	5	5.2499	0.2499	0.0197	5.1202	0.1202	0.0126
ω_4	12	13.1030	1.1030	0.0575	12.4000	0.4000	0.0328
α_1	0.2	0.1897	-0.0104	0.0012	0.1914	-0.0086	0.0010
α_2	0.3	0.2882	-0.0118	0.0012	0.2916	-0.0085	0.0010
α_3	0.1	0.0926	-0.0074	0.0015	0.0939	-0.0061	0.0010
α_4	0.3	0.2843	-0.0157	0.0017	0.2944	-0.0056	0.0010
β	0.3	0.2780	-0.0220	0.0020	0.2917	-0.0083	0.0013

Table 7. Maximum likelihood estimation results from 3,000 simulations based on different sample sizes (multiple observations within a period); parameter Set 3.

Parameters	True Values	T=540			T=1,080		
		Estimates	Bias	SE	Estimates	Bias	SE
ω_1	4	4.4755	0.4755	0.0252	4.2636	0.2636	0.0183
ω_2	6	6.8556	0.8556	0.0421	6.4561	0.4561	0.0293
ω_3	5	5.6538	0.6538	0.0329	5.3267	0.3267	0.0224
ω_4	4	4.7379	0.7379	0.0330	4.3430	0.3430	0.0217
α_1	0.1	0.0934	-0.0066	0.0012	0.0946	-0.0054	0.0010
α_2	0.2	0.1944	-0.0056	0.0013	0.1966	-0.0034	0.0010
α_3	0.1	0.0876	-0.0124	0.0014	0.0947	-0.0053	0.0010
α_4	0.2	0.1781	-0.0219	0.0018	0.1932	-0.0068	0.0010
β	0.4	0.3473	-0.0527	0.0031	0.3726	-0.0275	0.0022

6.2.2. Weekly Data with Multiple Observations within a Period. Let $\theta = (\omega_1, \omega_2, \omega_3, \omega_4, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \beta)$, represent all parameters in the PACP model. The simulated data were generated from a PACP process of 4 periods with $p, q = 1$ and 13 data points within each period. Three different parameter sets were used for analysis.

Set 1. $\theta = (2, 5, 3, 4, 0.1, 0.15, 0.2, 0.2, 0.1)$ (Table 8).

Set 2. $\theta = (2, 3, 1, 4, 0.2, 0.3, 0.1, 0.2, 0.2)$ (Table 9).

Set 3. $\theta = (7, 3, 8, 4, 0.2, 0.1, 0.1, 0.2, 0.3)$ (Table 10).

For each combination of parameter sets, sample size T=260 and T=520 were considered. These are approximately equivalent to the numbers of observations for a 5-year and 10-year daily count data respectively. Maximum likelihood estimation results from 3,000 simulations based on different sample sizes are reported in Tables 8-10.

Table 8. Maximum likelihood estimation results from 3,000 simulations based on different sample sizes (multiple observations within a period); parameter Set 1.

Parameters	True Values	T=280			T=560		
		Estimates	Bias	SE	Estimates	Bias	SE
ω_1	2	1.9865	-0.0135	0.0121	2.0085	0.0085	0.0094
ω_2	5	5.0565	0.0565	0.0273	5.0222	0.0222	0.0207
ω_3	3	3.0502	0.0502	0.0208	3.0319	0.0319	0.0160
ω_4	4	4.0444	0.0444	0.0270	4.0353	0.0353	0.0209
α_1	0.1	0.0931	-0.0069	0.0027	0.0908	-0.0092	0.0022
α_2	0.15	0.1300	-0.0200	0.0032	0.1403	-0.0097	0.0024
α_3	0.2	0.1747	-0.0253	0.0036	0.1863	-0.0137	0.0025
α_4	0.2	0.1808	-0.0192	0.0036	0.1863	-0.0137	0.0027
β	0.1	0.1100	0.0100	0.0037	0.1057	0.0057	0.0030

Table 9. Maximum likelihood estimation results from 3,000 simulations based on different sample sizes (multiple observations within a period); parameter Set 2.

Parameters	True Values	T=280			T=560		
		Estimates	Bias	SE	Estimates	Bias	SE
ω_1	2	2.0719	0.0719	0.0155	2.0468	0.0468	0.0112
ω_2	3	3.1726	0.1726	0.0238	3.1174	0.1174	0.0179
ω_3	1	1.0072	0.0072	0.0072	1.0148	0.0148	0.0050
ω_4	4	4.1770	0.1770	0.0240	4.0901	0.0901	0.0166
α_1	0.2	0.1801	-0.0199	0.0031	0.1915	-0.0085	0.0023
α_2	0.3	0.2712	-0.0289	0.0036	0.2829	-0.0171	0.0025
α_3	0.1	0.0953	-0.0047	0.0027	0.0932	-0.0068	0.0021
α_4	0.2	0.1759	-0.0241	0.0035	0.1883	-0.0117	0.0024
β	0.2	0.1978	-0.0022	0.0035	0.1956	-0.0044	0.0026

Table 10. Maximum likelihood estimation results from 3,000 simulations based on different sample sizes (multiple observations within a period); parameter Set 3.

Parameters	True Values	T=280			T=560		
		Estimates	Bias	SE	Estimates	Bias	SE
ω_1	7	7.3174	0.3174	0.0468	7.1385	0.1385	0.0351
ω_2	3	3.0347	0.0347	0.0172	3.0297	0.0297	0.0120
ω_3	8	8.1619	0.1619	0.0414	8.0981	0.0981	0.0299
ω_4	4	4.2597	0.2597	0.0340	4.0774	0.0774	0.0229
α_1	0.2	0.1781	-0.0219	0.0030	0.1871	-0.0129	0.0023
α_2	0.1	0.0935	-0.0065	0.0026	0.0891	-0.0109	0.0019
α_3	0.1	0.0882	-0.0118	0.0026	0.0883	-0.0117	0.0020
α_4	0.2	0.1719	-0.0282	0.0033	0.1863	-0.0137	0.0022
β	0.3	0.2983	-0.0017	0.0030	0.3035	0.0035	0.0021

From the simulation results, it is clear that the maximum likelihood estimation method gives a relatively small bias (10% or less in most cases) and low Monte Carlo standard errors. It is only in Table 7 we see a bias slightly larger than 10%. It may be because of the larger value for β. Note that the estimation bias is smaller and standard error is lower for the larger sample size, in both the single observation per period and multiple observations per period cases. This demonstrates that the MLE is a viable method for estimating the parameters of the suggested periodic autoregressive Poisson model and that larger sample sizes produce more accurate estimates.

7. MODEL SELECTION

To examine whether AIC and/or BIC are good criteria to distinguish the true structure of the count data, a small scale Monte Carlo simulation study was performed. All statistics reported here are calculated from N=3,000 replications, and each replication having sample size of T=1,080. In order to avoid artifacts created by initial conditions, the first 360 time series data points were discarded.

Mean AIC is averaged from AIC values for each of the replications and the percentage in the brackets indicates the percentage of times a simulation run yielded a smaller AIC value for the corresponding model. Thus, 90% indicate that the corresponding model had a smaller AIC value in 90% of the simulation runs.

Table 11 shows results for the case when the data were generated from an ACP process with true parameters $\omega = 2$, $\alpha = 0.1$, and $\beta = 0.15$. Both ACP and PACP model were used to fit the data. In 94 out of 100 times, the AIC values for ACP model are lower than those for the PACP models, which suggests that AIC performs well in identifying the true structure of the time series. BIC does not work in this situation since PACP model gives out the same estimates for each season as the ACP model. However, AIC put more penalty on large parameter sets than BIC. Thus it has better performance than BIC.

Table 11. ACP and PACP model selection by AIC criterion with simulated time series data with ACP as the data generating process.

Parameters	ACP Model	PACP Model
ω_1	1.8626	1.9312
ω_2	-	1.9271
ω_3	-	1.9307
ω_4	-	1.9337
α_1	0.0993	0.0959
α_2	-	0.0972
α_3	-	0.0962
α_4	-	0.0959
β	0.1558	0.1314
Mean AIC	3955.7701(94.233%)	3961.7950(5.767%)

Table 12. ACP and PACP model selection by AIC and BIC criteria with simulated time series data generated under a PACP model.

Parameters	ACP Model	PACP Model
ω_1	0.1594	1.9978
ω_2	-	4.9635
ω_3	-	3.0050
ω_4	-	3.9987
α_1	0.1523	0.0948
α_2	-	0.0515
α_3	-	0.1936
α_4	-	0.1951
β	0.8135	0.1048
Mean AIC	4726.2751 (0%)	4605.7300 (100%)
Mean BIC	4783.1362 (0%)	4650.5933 (100%)

Table 12 shows results when a PACP process with true parameters $\omega_1 = 2, \omega_2 = 5, \omega_3 = 3, \omega_4 = 4, \alpha_1 = 0.1, \alpha_2 = 0.05, \alpha_3 = 0.2, \alpha_4 = 0.2, \beta = 0.1$ is the underlying structure producing the count data. Both ACP model and PACP model were used to fit the data. In this case, AIC and BIC both show their strong ability to select the right structure. Notice that when there is periodicity in the count data, an ordinary ACP model gives out estimates of α and β with their sum close to one, suggesting near non-stationarity, raising questions about the appropriateness of the model.

8. VISUALIZATION OF SIMULATED DATA AND ESTIMATED INTENSITY PROCESS

The application of the proposed PACP model is demonstrated using a simulated data set. The data is generated from a PACP process with 4 seasons and each season has 90 data points. The parameter set is $\omega_1 = 2$, $\omega_2 = 5$, $\omega_3 = 3$, $\omega_4 = 4$, $\alpha_1 = 0.1$, $\alpha_2 = 0.05$, $\alpha_3 = 0.2$, $\alpha_4 = 0.2$, $\beta = 0.1$.

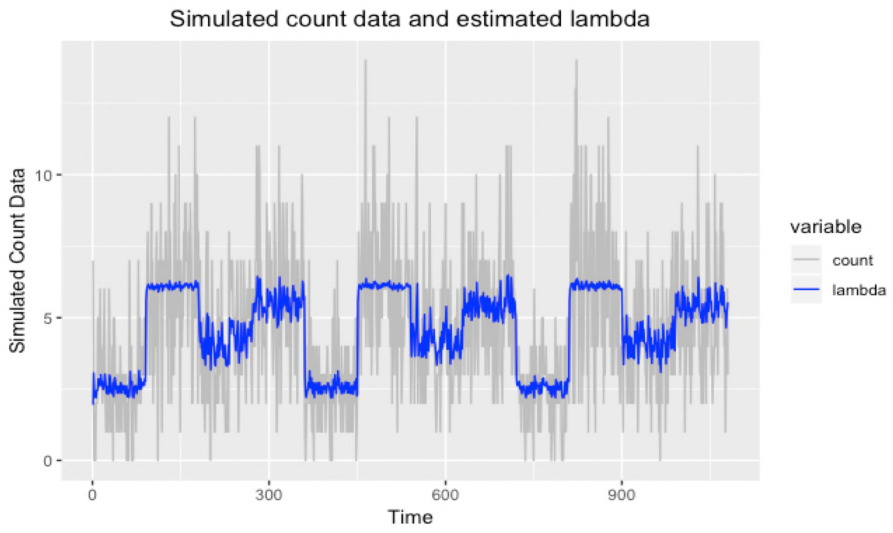


Figure 1. Simulated time series count data and the estimated intensity λ_t process.

As shown in Figure 1, the grey line represents the simulated data while the blue line indicates the estimated underlying periodic process. The results show the PACP model captures the cyclical movement of the process well.

9. REAL DATA ANALYSIS

In this section, the proposed PACP model is applied to fit a real dataset, and the performance is compared to a regular ACP model. From section 7, AIC is an appropriate standard to select the best model from all the competing models. The following example provides insight into the circumstances that a PACP model should be considered. The data set is weekly Astrovirus infection case numbers of males in Germany, which is available from Robert Koch Institute (website https://survstat.rki.de).

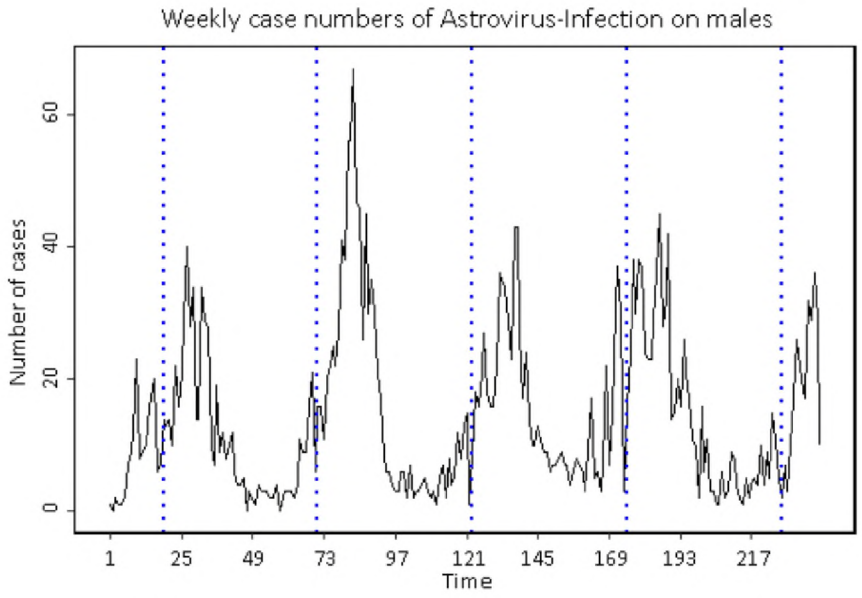


Figure 2. Weekly observations time series plot for Astrovirus infection cases of males in Germany. The time interval covering the count data is from week 36 of year 2015 to week 14 of year 2020, resulting in a total of 240 weekly observations of infection cases. The data set has a mean of 14.4417 and a variance of 159.7121.

The time interval covering the count data is from week 36 of year 2015 to week 14 of year 2020, resulting in a total of 240 weekly observations of infection cases. The data set has a mean of 14.4417 and variance of 159.7121. The periodicity is apparently shown in a time series plot of the data set in Figure 2, and it also suggested a varying structure of the volatility among the counts.

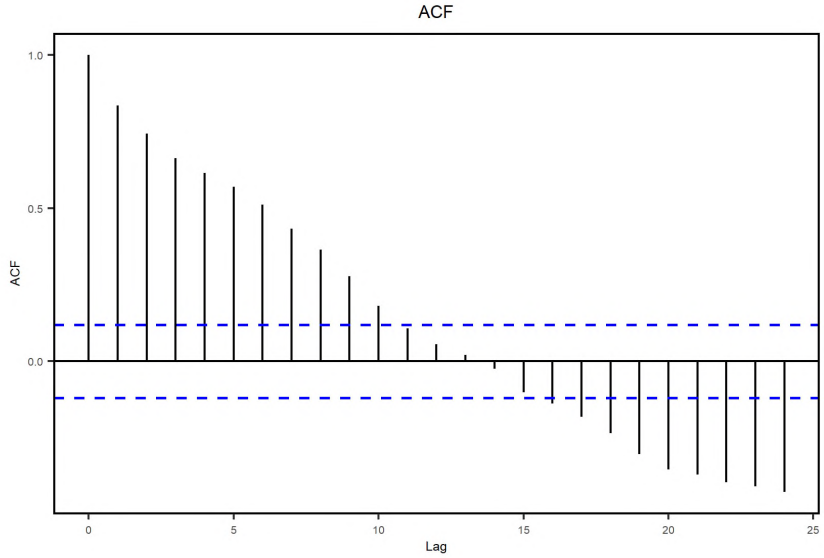
Figure 3 illustrates the time series are correlated. Combining the fact that an annual seasonality is observed in this count data, it is appropriate to consider fitting the PACP model to the time series. We need to consider several models to identify the best one that describes the periodicity and time varying auto-correlation structure shown in this data.

We fit three PACP model with 2, 3 and 4 periods separately, as shown in Table 14 with model names as PACP (S=2), PACP (S=3) and PACP (S=4) respectively. We also fit a regular ACP model to compare its performance with our proposed PACP model. MLE algorithm is used to estimate the model parameters and AIC values are listed as the criteria to select the best model.

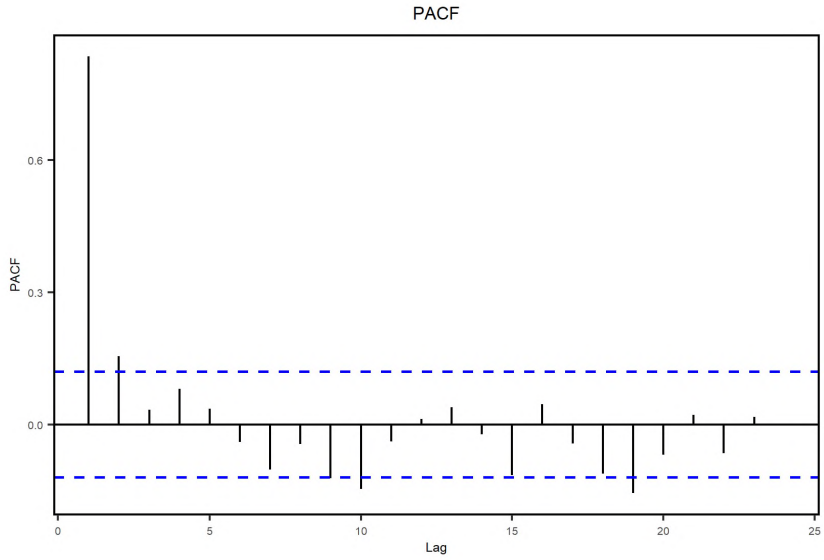
For PACP with 2 periods, we combined winter and spring since they have relatively large counts. Similarly we combined the summer and autumn since they have small counts. For PACP with 3 periods, we combined the summer and autumn since they all have low counts and less variation. Combining the spring and winter was also considered, resulting in a larger AIC value. For PACP with 4 periods, we naturally take the 4 seasons as it is to be the 4 periods.

Table 13. Periods defined for each PACP model.

Model	S1	S2	S3	S4
regular ACP	-	-	-	-
PACP(S=2)	Winter, Spring	Summer, Autumn	-	-
PACP(S=3)	Winter	Spring	Summer, Autumn	-
PACP(S=4)	Winter	Spring	Summer	Autumn



(a) ACF



(b) PACF

Figure 3. Auto-correlation function (ACF) and partial auto-correlation function (PACF) of Astrovirus infection count data.

According to the results shown in Table 14, the PACP model has an overall better performance than regular ACP model while PACP model with 3 seasons giving the best fit. The mean of the Poisson process estimated from regular ACP model is quite small (1.05), which in other words, is not quite reasonable for a time series count data with mean 14.4417. However, the means for each season from the PACP model describe the pattern

demonstrated in the time series plot well and the variations of the estimated alpha across the 3 seasons illustrate the periodic nature of the underlying serial correlations. Thus, the regular ACP model could not capture the time varying dependent structure like the proposed PACP model, which demonstrates the necessity of such a generalization.

Table 14. Estimated parameters, AIC for the Astrovirus infection cases of males in Germany.

Model	Regular ACP	PACP(S=2)	PACP(S=3)	PACP(S=4)
ω_1	1.0486	4.6719	5.0388	4.9886
ω_2	-	1.8572	6.5027	6.3540
ω_3	-	-	1.3623	1.8554
ω_4	-	-	-	1.6735
α_1	0.6081	0.5562	0.4013	0.5045
α_2	-	0.3969	0.5292	0.5224
α_3	-	-	0.4723	0.3559
α_4	-	-	-	0.4938
β	0.3238	0.2506	0.2329	0.2454
AIC	1725.7	1650.0	1644.1	1649.9

10. CONCLUSION

The model provided here is a natural generalization of the autoregressive conditional Poisson model, which allows periodicity to be taken into consideration when modeling count time series. The reported simulation results in Section 6 show that the MLE provides reasonable estimates of the model parameters of the PACP model. We also studied the utility of using AIC and BIC criteria in determining if the underlying data generating process is ACP or PACP. Results suggest that the use of AIC criterion is a trustworthy way

to differentiate between the underlying ACP or PACP structure. In addition, the simulated data and real data analysis illustrate there is indeed the necessity to generalize the original ACP model to accommodate periodic component in the count data.

APPENDIX

1. ACP EXTENSION

In this section, we derived the reformulation of the original set up equation for our model to an integer valued seasonal ARMA(max(p, q), p) process. The count data has a stochastic process satisfies

$$\lambda_t = \omega_{s(t)} + \sum_{i=1}^{q} \alpha_{is(t)} N_{t-i} + \sum_{j=1}^{p} \beta_{js(t)} \lambda_{t-j}.$$

Equivalently, it could be rewritten as an integer valued ARMA(max(p, q), p) model.

(i) Case1. $p \leq q$.

$$N_t + \lambda_t = N_t + \omega_{s(t)} + \sum_{i=1}^{q} \alpha_{is(t)} N_{t-i} + \sum_{j=1}^{p} \beta_{js(t)} \lambda_{t-j},$$

$$N_t = N_t - \lambda_t + \omega_{s(t)} + \sum_{i=1}^{q} \alpha_{is(t)} N_{t-i} + \sum_{j=1}^{p} \beta_{js(t)} \lambda_{t-j},$$

$$N_t = N_t - \lambda_t + \omega_{s(t)} + \sum_{i=1}^{q} \alpha_{is(t)} N_{t-i} + \sum_{j=1}^{p} \beta_{js(t)} N_{t-j} - \sum_{j=1}^{p} \beta_{js(t)} N_{t-j} + \sum_{j=1}^{p} \beta_{js(t)} \lambda_{t-j},$$

$$N_t = N_t - \lambda_t + \omega_{s(t)} + \sum_{i=1}^{q} \alpha_{is(t)} N_{t-i} + \sum_{j=1}^{p} \beta_{js(t)} N_{t-j} - \left(\sum_{j=1}^{p} \beta_{js(t)} N_{t-j} - \sum_{j=1}^{p} \beta_{js(t)} \lambda_{t-j} \right),$$

$$N_t = \underbrace{N_t - \lambda_t}_{\epsilon_t} + \omega_{s(t)} + \sum_{i=1}^{q} \alpha_{is(t)} N_{t-i} + \sum_{j=1}^{p} \beta_{js(t)} N_{t-j} - \underbrace{\left(\sum_{j=1}^{p} \beta_{js(t)} N_{t-j} - \sum_{j=1}^{p} \beta_{js(t)} \lambda_{t-j} \right)}_{\sum_{j=1}^{p} \beta_{js(t)} \epsilon_{t-j}},$$

$$N_t = \epsilon_t + \omega_{s(t)} + \sum_{i=1}^{q} \alpha_{is(t)} N_{t-i} + \sum_{j=1}^{p} \beta_{js(t)} N_{t-j} - \sum_{j=1}^{p} \beta_{js(t)} \epsilon_{t-j},$$

$$N_t = \epsilon_t + \omega_{s(t)} + \sum_{i=1}^{q} \alpha_{is(t)} N_{t-i} + \sum_{j=1}^{p} \beta_{js(t)} N_{t-j} - \sum_{j=1}^{p} \beta_{js(t)} \epsilon_{t-j}.$$

Choose $\beta_{js(t)} \equiv 0$ for $j > p$, then,

$$N_t = \epsilon_t + \omega_{s(t)} + \sum_{i=1}^{q} \alpha_{is(t)} N_{t-i} + \sum_{j=1}^{p} \beta_{js(t)} N_{t-j} - \sum_{j=1}^{p} \beta_{js(t)} \epsilon_{t-j},$$

$$N_t = \epsilon_t + \omega_{s(t)} + \sum_{i=1}^{q} \alpha_{is(t)} N_{t-i} + \sum_{j=1}^{q} \beta_{js(t)} N_{t-j} - \sum_{j=1}^{p} \beta_{js(t)} \epsilon_{t-j} \mid \beta_{js(t)} \equiv 0, \quad \forall p \leq q,$$

$$N_t = \epsilon_t + \omega_{s(t)} + \sum_{i=1}^{q} (\alpha_{is(t)} + \beta_{is(t)}) N_{t-i} - \sum_{j=1}^{p} \beta_{js(t)} \epsilon_{t-j},$$

$$N_t = \epsilon_t + \omega_{s(t)} + \sum_{i=1}^{max(p,q)} (\alpha_{is(t)} + \beta_{is(t)}) N_{t-i} - \sum_{j=1}^{p} \beta_{js(t)} \epsilon_{t-j}.$$

(ii) Case2. $p > q$.

$$N_t + \lambda_t = N_t + \omega_{s(t)} + \sum_{i=1}^{q} \alpha_{is(t)} N_{t-i} + \sum_{j=1}^{p} \beta_{js(t)} \lambda_{t-j},$$

$$N_t = N_t - \lambda_t + \omega_{s(t)} + \sum_{i=1}^{q} \alpha_{is(t)} N_{t-i} + \sum_{j=1}^{p} \beta_{js(t)} \lambda_{t-j},$$

$$N_t = N_t - \lambda_t + \omega_{s(t)} + \sum_{i=1}^{q} \alpha_{is(t)} N_{t-i} + \sum_{j=1}^{p} \beta_{js(t)} N_{t-j} - \sum_{j=1}^{p} \beta_{js(t)} N_{t-j} + \sum_{j=1}^{p} \beta_{js(t)} \lambda_{t-j},$$

$$N_t = N_t - \lambda_t + \omega_{s(t)} + \sum_{i=1}^{q} \alpha_{is(t)} N_{t-i} + \sum_{j=1}^{p} \beta_{js(t)} N_{t-j} - \left(\sum_{j=1}^{p} \beta_{js(t)} N_{t-j} - \sum_{j=1}^{p} \beta_{js(t)} \lambda_{t-j} \right),$$

$$N_t = \underbrace{N_t - \lambda_t}_{\epsilon_t} + \omega_{s(t)} + \sum_{i=1}^{q} \alpha_{is(t)} N_{t-i} + \sum_{j=1}^{p} \beta_{js(t)} N_{t-j} - \underbrace{\left(\sum_{j=1}^{p} \beta_{js(t)} N_{t-j} - \sum_{j=1}^{p} \beta_{js(t)} \lambda_{t-j} \right)}_{\sum_{j=1}^{p} \beta_{js(t)} \epsilon_{t-j}},$$

$$N_t = \epsilon_t + \omega_{s(t)} + \sum_{i=1}^{q} \alpha_{is(t)} N_{t-i} + \sum_{j=1}^{p} \beta_{js(t)} N_{t-j} - \sum_{j=1}^{p} \beta_{js(t)} \epsilon_{t-j},$$

$$N_t = \epsilon_t + \omega_{s(t)} + \sum_{i=1}^{q} \alpha_{is(t)} N_{t-i} + \sum_{j=1}^{p} \beta_{js(t)} N_{t-j} - \sum_{j=1}^{p} \beta_{js(t)} \epsilon_{t-j}.$$

Now choose $\alpha_{is(t)} \equiv 0$ for $i > q$. Then,

$$N_t = \epsilon_t + \omega_{s(t)} + \sum_{i=1}^{q} \alpha_{is(t)} N_{t-i} + \sum_{j=1}^{p} \beta_{js(t)} N_{t-j} - \sum_{j=1}^{p} \beta_{js(t)} \epsilon_{t-j},$$

$$N_t = \epsilon_t + \omega_{s(t)} + \sum_{i=1}^{p} \alpha_{is(t)} N_{t-i} + \sum_{j=1}^{p} \beta_{js(t)} N_{t-j} - \sum_{j=1}^{p} \beta_{js(t)} \epsilon_{t-j} \mid \alpha_{is(t)} \equiv 0, \quad \forall i \geq q,$$

$$N_t = \epsilon_t + \omega_{s(t)} + \sum_{i=1}^{p} (\alpha_{is(t)} + \beta_{is(t)}) N_{t-i} - \sum_{j=1}^{p} \beta_{js(t)} \epsilon_{t-j},$$

$$N_t = \epsilon_t + \omega_{s(t)} + \sum_{i=1}^{max(p,q)} (\alpha_{is(t)} + \beta_{is(t)}) N_{t-i} - \sum_{j=1}^{p} \beta_{js(t)} \epsilon_{t-j}.$$

For both cases, they all could be rewritten as an ARMA(max(p, q), p) process.

2. VARMA FORM (FOUR SEASONS WITHOUT REPEATED OBSERVATIONS)

In this section, we derived the Vector ARMA form of the periodic count time series from the integer valued seasonal ARMA(max(p,q),p) process. Note that we can write

$$\begin{bmatrix} 1 & -(\alpha_{14}+\beta_{14}) & 0 & 0 \\ 0 & 1 & -(\alpha_{13}+\beta_{13}) & 0 \\ 0 & 0 & 1 & -(\alpha_{12}+\beta_{12}) \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} N_{4\tau} \\ N_{4\tau-1} \\ N_{4\tau-2} \\ N_{4\tau-3} \end{bmatrix} = \begin{bmatrix} \omega_4 \\ \omega_3 \\ \omega_2 \\ \omega_1 \end{bmatrix} +$$
$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ (\alpha_{11}+\beta_{11}) & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} N_{4(\tau-1)} \\ N_{4(\tau-1)-1} \\ N_{4(\tau-1)-2} \\ N_{4(\tau-1)-3} \end{bmatrix} + \begin{bmatrix} 1 & -\beta_{14} & 0 & 0 \\ 0 & 1 & -\beta_{13} & 0 \\ 0 & 0 & 1 & -\beta_{12} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{4\tau} \\ \epsilon_{4\tau-1} \\ \epsilon_{4\tau-2} \\ \epsilon_{4\tau-3} \end{bmatrix}$$
$$+ \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -\beta_{11} & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \epsilon_{4(\tau-1)} \\ \epsilon_{4(\tau-1)-1} \\ \epsilon_{4(\tau-1)-2} \\ \epsilon_{4(\tau-1)-3} \end{bmatrix}.$$

Multiplying both sides by the inverse matrix given below,

$$\begin{bmatrix} 1 & \alpha_{14}+\beta_{14} & (\alpha_{13}+\beta_{13})(\alpha_{14}+\beta_{14}) & (\alpha_{12}+\beta_{12})(\alpha_{13}+\beta_{13})(\alpha_{14}+\beta_{14}) \\ 0 & 1 & \alpha_{13}+\beta_{13} & (\alpha_{12}+\beta_{12})(\alpha_{13}+\beta_{13}) \\ 0 & 0 & 1 & \alpha_{12}+\beta_{12} \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

we obtain the Vector ARMA form:

$$\begin{bmatrix} N_{4\tau} \\ N_{4\tau-1} \\ N_{4\tau-2} \\ N_{4\tau-3} \end{bmatrix} = \begin{bmatrix} \omega_1 (\alpha_{12}+\beta_{12}) (\alpha_{13}+\beta_{13}) (\alpha_{14}+\beta_{14}) + \omega_2 (\alpha_{13}+\beta_{13}) (\alpha_{14}+\beta_{14}) + \omega_3 (\alpha_{14}+\beta_{14}) + \omega_4 \\ \omega_1 (\alpha_{12}+\beta_{12}) (\alpha_{13}+\beta_{13}) + \omega_2 (\alpha_{13}+\beta_{13}) + \omega_3 \\ \omega_1 (\alpha_{12}+\beta_{12}) + \omega_2 \\ \omega_1 \end{bmatrix}$$

$$+ \begin{bmatrix} (\alpha_{11}+\beta_{11}) (\alpha_{12}+\beta_{12}) (\alpha_{13}+\beta_{13}) (\alpha_{14}+\beta_{14}) & 0 & 0 & 0 \\ (\alpha_{11}+\beta_{11}) (\alpha_{12}+\beta_{12}) (\alpha_{13}+\beta_{13}) & 0 & 0 & 0 \\ (\alpha_{11}+\beta_{11}) (\alpha_{12}+\beta_{12}) & 0 & 0 & 0 \\ \alpha_{11}+\beta_{11} & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} N_{4(\tau-1)} \\ N_{4(\tau-1)-1} \\ N_{4(\tau-1)-2} \\ N_{4(\tau-1)-3} \end{bmatrix}$$

$$+ \begin{bmatrix} 1 & \alpha_{14} & \alpha_{13} (\alpha_{14}+\beta_{14}) & \alpha_{12} (\alpha_{13}+\beta_{13}) (\alpha_{14}+\beta_{14}) \\ 0 & 1 & \alpha_{13} & \alpha_{12} (\alpha_{13}+\beta_{13}) \\ 0 & 0 & 1 & \alpha_{12} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{4\tau} \\ \epsilon_{4\tau-1} \\ \epsilon_{4\tau-2} \\ \epsilon_{4\tau-3} \end{bmatrix}$$

$$+ \begin{bmatrix} -\beta_{11} (\alpha_{12}+\beta_{12}) (\alpha_{13}+\beta_{13}) (\alpha_{14}+\beta_{14}) & 0 & 0 & 0 \\ -\beta_{11} (\alpha_{12}+\beta_{12}) (\alpha_{13}+\beta_{13}) & 0 & 0 & 0 \\ -\beta_{11} (\alpha_{12}+\beta_{12}) & 0 & 0 & 0 \\ -\beta_{11} & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \epsilon_{4(\tau-1)} \\ \epsilon_{4(\tau-1)-1} \\ \epsilon_{4(\tau-1)-2} \\ \epsilon_{4(\tau-1)-3} \end{bmatrix}.$$

Now take the expectation of both sides and assume that the observations from the same period have the same expected value, to get the expectation of each observation from each season:

$$\begin{cases} E(\lambda_1) = \dfrac{\omega 1 + \omega 2 (\alpha_{14}+\beta_{14}) + \omega 3 (\alpha_{13}+\beta_{13}) (\alpha_{14}+\beta_{14}) + \omega 4 (\alpha_{12}+\beta_{12}) (\alpha_{13}+\beta_{13}) (\alpha_{14}+\beta_{14})}{1-(\alpha_{11}+\beta_{11}) (\alpha_{12}+\beta_{12}) (\alpha_{13}+\beta_{13}) (\alpha_{14}+\beta_{14})} \\ E(\lambda_2) = \dfrac{\omega 1 (\alpha_{11}+\beta_{11}) (\alpha_{12}+\beta_{12}) (\alpha_{13}+\beta_{13}) + \omega 2 + \omega 3 (\alpha_{13}+\beta_{13}) + \omega 4 (\alpha_{12}+\beta_{12}) (\alpha_{13}+\beta_{13})}{1-(\alpha_{11}+\beta_{11}) (\alpha_{12}+\beta_{12}) (\alpha_{13}+\beta_{13}) (\alpha_{14}+\beta_{14})} \\ E(\lambda_3) = \dfrac{\omega 1 (\alpha_{11}+\beta_{11}) (\alpha_{12}+\beta_{12}) + \omega 2 (\alpha_{11}+\beta_{11}) (\alpha_{12}+\beta_{12}) (\alpha_{14}+\beta_{14}) + \omega 3 + \omega 4 (\alpha_{12}+\beta_{12})}{1-(\alpha_{11}+\beta_{11}) (\alpha_{12}+\beta_{12}) (\alpha_{13}+\beta_{13}) (\alpha_{14}+\beta_{14})} \\ E(\lambda_4) = \dfrac{\omega 1 (\alpha_{11}+\beta_{11}) + \omega 2 (\alpha_{11}+\beta_{11}) (\alpha_{14}+\beta_{14}) + \omega 3 (\alpha_{11}+\beta_{11}) (\alpha_{13}+\beta_{13}) (\alpha_{14}+\beta_{14}) + \omega 4}{1-(\alpha_{11}+\beta_{11}) (\alpha_{12}+\beta_{12}) (\alpha_{13}+\beta_{13}) (\alpha_{14}+\beta_{14})} \end{cases}.$$

Now let $a_{11} + b_{11} = A$, $a_{12} + b_{12} = B$, $a_{13} + b_{13} = C$ and $a_{14} + b_{14} = D$, the expectation of observation of each season could be rewritten as:

$$\begin{cases} E(\lambda_1) : \dfrac{\omega_1 + D\omega_2 + CD\omega_3 + BCD\omega_4}{1 - ABCD} \\ E(\lambda_2) : \dfrac{ABC\omega_1 + \omega_2 + C\omega_3 + BC\omega_4}{1 - ABCD} \\ E(\lambda_3) : \dfrac{AB\omega_1 + ABD\omega_2 + \omega_3 + B\omega_4}{1 - ABCD} \\ E(\lambda_4) : \dfrac{A\omega_1 + AD\omega_2 + ACD\omega_3 + \omega_4}{1 - ABCD} \end{cases} .$$

REFERENCES

[1] Andréas Heinen. Modelling time series count data: an autoregressive conditional poisson model. *Available at SSRN 1117187*, 2003.

[2] Siddhartha Chib and Rainer Winkelmann. Markov chain monte carlo analysis of correlated count data. *Journal of Business & Economic Statistics*, 19(4):428–435, 2001.

[3] Iain L MacDonald and Walter Zucchini. *Hidden Markov and other models for discrete-valued time series*, volume 110. CRC Press, 1997.

[4] Tunny Sebastian, Visalakshi Jeyaseelan, Lakshmanan Jeyaseelan, Shalini Anandan, Sebastian George, and Shrikant I Bangdiwala. Decoding and modelling of time series count data using poisson hidden markov model and markov ordinal logistic regression models. *Statistical methods in medical research*, 28(5):1552–1563, 2019.

[5] Ben Cooper and Marc Lipsitch. The analysis of hospital infection data using hidden markov models. *Biostatistics*, 5(2):223–237, 2004.

[6] Ed McKenzie. Some simple models for discrete variate time series 1. *JAWRA Journal of the American Water Resources Association*, 21(4):645–650, 1985.

[7] Rong Zhu and Harry Joe. Modelling count data time series with markov processes based on binomial thinning. *Journal of Time Series Analysis*, 27(5):725–738, 2006.

[8] A Colin Cameron and Pravin K Trivedi. *Regression analysis of count data*, volume 53. Cambridge university press, 1998.

[9] Rainer Winkelmann and Klaus F Zimmermann. Recent developments in count data modelling: theory and application. *Journal of economic surveys*, 9(1):1–24, 1995.

[10] Stefany Coxe, Stephen G West, and Leona S Aiken. The analysis of count data: A gentle introduction to poisson regression and its alternatives. *Journal of personality assessment*, 91(2):121–136, 2009.

[11] Tracy DeHart, Howard Tennen, Stephen Armeli, Michael Todd, and Glenn Affleck. Drinking to regulate negative romantic relationship interactions: The moderating role of self-esteem. *Journal of Experimental Social Psychology*, 44(3):527–538, 2008.

[12] Andreas Lindén and Samu Mäntyniemi. Using the negative binomial distribution to model overdispersion in ecological count data. *Ecology*, 92(7):1414–1421, 2011.

[13] Patricia A Jacobs and Peter AW Lewis. Discrete time series generated by mixtures. i: Correlational and runs properties. *Journal of the Royal Statistical Society: Series B (Methodological)*, 40(1):94–105, 1978.

[14] Patricia A Jacobs and Peter AW Lewis. Stationary discrete autoregressive-moving average time series generated by mixtures. *Journal of Time Series Analysis*, 4(1):19–36, 1983.

[15] René Ferland, Alain Latour, and Driss Oraichi. Integer-valued garch process. *Journal of Time Series Analysis*, 27(6):923–942, 2006.

[16] Fukang Zhu. A negative binomial integer-valued garch model. *Journal of Time Series Analysis*, 32(1):54–67, 2011.

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

[18] Ming Yang, Gideon KD Zamba, and Joseph E Cavanaugh. Markov regression models for count time series with excess zeros: A partial likelihood approach. *Statistical Methodology*, 14:26–38, 2013.

[19] Tobias A Möller, Christian H Weiß, Hee-Young Kim, and Andrei Sirchenko. Modeling zero inflation in count data time series with bounded support. *Methodology and Computing in Applied Probability*, 20(2):589–609, 2018.

[20] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.

[21] Tim Bollerslev and Eric Ghysels. Periodic autoregressive conditional heteroscedasticity. *Journal of Business & Economic Statistics*, 14(2):139–151, 1996.

II. AUTOREGRESSIVE CONDITIONAL HETEROSKEDASTIC HIDDEN MARKOV MODEL

Yi Zhang, V Samaranayake
Department of Mathematics, Missouri University of Science and Technology, Rolla, MO 65409, USA
Email: zybvf@umsystem.edu

ABSTRACT

Poisson hidden Markov models (P-HMM), which are widely used for modeling time series count data, were originally established and applied in the biometric area [1]. In the Poisson hidden Markov models, the mean of a Poisson process varies according to the states of the latent Markov chain. The proposed autoregressive conditional Poisson hidden Markov model (ACP-HMM) assumes the mean of a Poisson process not only depends on the underlying states of a Markov chain but also depends on past values. It is more appropriate to fit time series count data with ACP-HMM when data exhibits strong auto-correlation. A Monte Carlo simulation study is performed to evaluate the estimation of parameters of ACP-HMM model employing the maximum likelihood method. The relatively small mean standard errors show that the maximum likelihood estimation (MLE) method works reasonably well for the model with a small number of states ($n \leq 4$). A simulation study shows the importance of identifying the correct structure of an auto-correlated time series count data with possible hidden states. Finally, a real-life data is used to illustrates the potential for the wide uses of such models.

Keywords: count data, Markov chain, seasonality, conditional heteroskedasticity, time varying parameters.

1. INTRODUCTION

A large number of high frequency count data time series are available thanks to advanced technology. Many of them exhibit changes of status over time while auto-correlation between time points could not be ignored. Examples of such time series could be found everywhere from epidemic data to the field of economics. The Poisson hidden Markov model is one of the most common formulations used for modeling count data processes with different states. It was first developed and applied to a time series of epileptic seizure counts [1]. It assumes the latent Poisson processes that generate the count data have significantly diverse means at different times. However, the Poisson processes are only determined by the status of the underlying Markov chain and conditioning on the past states, the current Markov chain state S_t only depends on the previous state S_{t-1}. As a result, a longer-term dependency amongst data points is not taken care of. Many time series demonstrate a characteristic of a high count often followed by several high counts. Thus, the autoregressive conditional Poisson hidden Markov model (ACP-HMM) is proposed to deal with such types of data more appropriately. It could be seen as a combination of a hidden Markov model and an autoregressive structure, which admits the existence of various underlying mechanisms that shift back and forth while capturing the strong serial correlation among time series observations. It also provides more reasonable estimates of the mean values of Poisson processes in cases where a cluster of high or low counts is actually a result from strong correlation within observations.

2. REVIEW OF MODELS FOR TIME SERIES COUNT DATA

For modeling count data, one of the most common distributions utilized to analyze is the Poisson distribution. A lot of publications introducing the Poisson regression models are available [2, 3, 4]. The basic probability mass function for a Poisson distribution is given by [3]: $P(Y = y|\mu) = \mu^y e^{-\mu}[y!]^{-1}$, where y is the realization from a Poisson random variable

Y and μ is the arithmetic mean. The Poisson generalized regression model establishes a relationship between the expected mean and dependent variables $\boldsymbol{X} = [X_1, X_2, \ldots, X_p]$ by the link function $\log(E(Y|\boldsymbol{X})) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$. However, such Poisson models are not able to handle more complex count data in reality. Thus, several variants from this model are developed to accommodate more characteristics in time series data.

For some data, variance observed is considerably larger than the expected, such phenomenon is called over-dispersion. In order to deal with this problem, which is a very common feature in count time series, several compound Poisson models are developed in diverse ways. Hinde [5] assumed that the mean of the Poisson distribution follows some form of another distribution and various kinds of distributions could lead to a whole class of different compound Poisson models. For example, the observations are set to follow a Poisson distribution with mean μ, and the mean μ follows a gamma distribution $\mu \sim \Gamma(k, \Theta)$. The advantage of such a construction is that it provides a more analytically tractable model and the mean is preserved as the mean of gamma distribution under such assumption, while over-dispersion problem is also solved.

Although compound Poisson models could solve some problems, negative binomial distribution is introduced to allow a more flexible relationship between mean and variance. Negative binomial distributions are developed to well handle data with over-dispersion, which is a common phenomenon in empirical data sets. In to 1953, Bliss et al. [6] fitted the negative binomial distribution to biological data (e.g., plants and animals), where the variance is significantly larger than the mean. Linden and Mantyniemi [7] also adopted negative binomial distribution to model over-dispersion in ecological count data. They proposed a parameterization of the negative binomial distribution, where two over-dispersion parameters were introduced to allow for various quadratic mean–variance relationships, including the ones assumed in the most commonly used approaches. They wrote the probability mass function as $P(X = x|r, p) = \Gamma(x + r)p^r(1 - p)^x[x!\Gamma(r)]^{-1}$, where the expectation (theoretical mean) $\mu = r(1 - p)p^{-1}$ and variance $\sigma^2 = r(1 - p)p^{-2}$. Then, they parameterized r

and p in terms of the mean and the variance as $r = \mu^2[\sigma^2 - \mu]^{-1}$ and $p = \mu\sigma^{-2}$. Using bird migration data as an example, they presented hypothetical scenarios on how over-dispersion arose due to sampling, flocking behavior or aggregation, environmental variability, or combinations of these factors. For all considered scenarios, mean–variance relationships can be appropriately described by the negative binomial distribution with two over-dispersion parameters.

For unexplained heterogeneity, Zeger [8] discussed a model for regression analysis with a time series of counts. Correlation was assumed to arise from an unobservable process added as another component of the mean. Hinde [5] introduced a different class of compound Poisson models to describe the variation of the data into two components. The original setup of the mean of the Poisson process $\lambda_i = exp(x_i\beta)$ was added with an error term resulting in $\lambda_i^* = exp(x_i\beta + \epsilon_i)$, where the error term could reflect a specification error caused by unobserved variation from the system itself. The compound Poisson distributions are a natural generalization of the basic Poisson models, which provide more flexibility to accommodate the complexity and multi-source variations in count data. It should be noticed that for certain parametric forms, the compound Poisson might not have a closed form and can be computationally cumbersome.

In reality, count data can be time correlated. McKenzie [9] developed two special classes of observation-driven models, which are known as the integer-valued autoregressive (INAR) and moving average (INMA) processes. The binomial thinning operator was introduced to generate sum of identical independent Bernoulli random variables, where the number of those random variables were determined by the previous counts. Another class of models that are widely used to model correlated time series is autoregressive moving average (ARMA) models. It incorporates the influence from previous observations into the present one. Continuous autoregressive time series data was introduced by Box and Jenkins [10], which is very useful to deal with the trend and seasonality. Then the discrete version: discrete autoregressive moving average (DARMA) models [11] are proposed for

count data, which are analogous to Gaussian ARMA models and with arbitrarily chosen marginal distributions. These models were obtained by a probabilistic mixture of a sequence of independent identically distributed discrete random variables. As a result, the realization of the process requires many runs of a single value. Different orders of autoregressive form are available [12]. By coupling two simple stationary processes, DAR (p) and DMA(q), a mixed model of DARMA(p, q) can be generated. This model is simplified into a single equation as the NDARMA(p, q), $X'_n = V_n X'_{n-A_n} + (1 - V_n)Y_{n-D_n}$, see Jacobs and Lewis [11]. The problem of the DARMA model is that a single value might have a high density around it when the sequence is generated from such structures. Ferland et al. introduced an integer-valued generalized autoregressive conditional heteroskedastic (INGARCH) (p, q) model with Poisson deviates [13]. For the case $p = 1$, $q = 1$, this INGARCH becomes a standard autoregressive moving average (1,1) process. Zhu [14] developed a negative binomial integer-valued GARCH model, aiming to handle over-dispersion and extreme observations. After that, Zhu [15] introduced a class of generalized Poisson integer-valued GARCH models, which can account for both over-dispersion and under-dispersion, mostly over-dispersion. The author considered the maximum likelihood estimators for the parameters and established their consistency and asymptotic normality. Chen et al. [16] proposed an autoregressive conditional negative binomial model for time series of counts that has a time-varying conditional autoregressive mean functions and heteroskedasticity. Since Poisson distribution could take care of integer data set and GARCH could incorporate the influence of previous observations, Heinen [17] proposed an autoregressive conditional Poisson model, of which the mean of the Poisson process parallels a GARCH structure.

Traditional linear models or similar regression based approaches are not able to capture the strong time-dependent trends, which exhibit the repeated patterns of seasonality. Markov chains represent a general class of models that can handle such time series of counts [9, 18, 19]. The Markov chain provides different states which correspond to certain count numbers, and the possible outcomes are controlled by transition probabilities. When

conditioning on the past value, the distribution of X_t only depends on X_{t-1}. However, they tend to be overparameterized for most practical data as there are usually more than two possible outcomes. Also, due to the stationary correlation structure, Markov chains are not easy to generalize. Raftery [20] introduced a new structure for higher order Markov chains, which greatly reduced the required number of parameters and is more general. These models have autoregressive-like structures, and were later called mixture transition distribution (MTD) models. Jung and Tremayne reported a high order Markov model, with a focus on the second order case [21]. They provided means of obtaining estimated standard errors which were not accessible by analytical methods.

Based on Markov chains, the original concept and theories about hidden Markov models were advocated by Baum [22, 23]. It was first applied to speech recognition technology [24]. Hidden Markov models (HMMs) have become more and more popular thanks to their ability to accommodate different states that allow series' structures to change over time. It is assumed that an unobserved latent state of the system changes over time according to a Markov process. As the result of booming bio-technology, count data carrying periodicity are available and often could be seen as the outcomes from repeatedly occurring different hidden states. Poisson hidden Markov model was first developed by Hopkins [25] and applied to a time series of epileptic seizure counts by Albert [1]. This model allows for the mean of a Poisson distribution to change according to an underlying two-state Markov chain. In case of excessive zero counts, DeSantis et al. [26] introduced a heterogeneous zero-inflated Poisson HMM (ZIP-HMM) by adding Markov chain into a zero-inflated Poisson process. The number of cocaine abuses per week before and after participation in a stress- and cue-reactivity study was used to demonstrate the performance of such a model. It is shown that the ZIP-HMM formulation performs better when compared with other time series models. What is proposed here borrows ideas from the P-HMM and the ACP model.

3. PROPOSED ACP-HMM MODELS

Let a sequence of discrete random variables $\{S_t : t \in \mathbb{N}\}$ to be a Markov Chain with m possible states and transition probability matrix $\mathbf{\Gamma}(t) = \{\gamma_{ij}^s(t)\}$, $i = 1, 2, \ldots, m, j = 1, 2, \ldots, m$, where $\gamma_{ij}^s(t) = P(S_{s+t} = j | S_s = i)$. In most cases, it is enough to use homogeneous Markov chains, which means γ_{ij}^s does not depends on s. Unless there is an explicit indication, it is assumed that Markov chain under discussion is a homogeneous one with transition probabilities denoted as γ_{ij}. Given the underlying Markov state S_t, the observed count data X_t follows a Poisson distribution

$$P(X_t = k | S_t = i) = \frac{e^{-\lambda_i} \lambda_i^k}{k!},$$

where λ_i is the expected mean of the state dependent Poisson process when the corresponding latent Markov chain takes state i.

However, the traditional Poisson hidden Markov model would not take care of the auto-correlation among time series data. Conditioning on the states, all observations are assumed to be independent. Thus the autoregressive conditional Poisson hidden Markov model proposed here is more appropriate to fit correlated count data whose means change with different states.

To define the suggested structure, let $\{X_t : t \in \mathbb{N}\}$ denote observed time series count data, with X_t representing the count at time t. It is assumed that the means of the Poisson processes are generated from different Markov chain status and are also affected by the near previous counts. Let S_t denotes the state of the Markov chain to which t belongs and the $\sigma-$ algebra generated by $\{X_i, S_i : i \leq t\}$ as $\mathcal{F}_t$. Given the past information $\mathcal{F}_{t-1}$

$$X_t | \mathcal{F}_{t-1} \sim Pois(\lambda_t),$$

where λ_t is a time varying structure defined as

$$\lambda_{t,S_t} = \omega_{S_t} + \sum_{i=1}^{q} \alpha_{i,S_t} X_{t-i},$$

and ω_{S_t}, $\alpha'_{i,St}s$, $i = 1, 2, , \ldots, q$, are positive for all values of S_t. Note the S_t illustrates the corresponding state of the underlying Markov chain at time t.

Note that the above formulation parallels that of an ARCH(q) process, but with the parameters varying with the state of the Markov chain. A simpler model, where the α_{i,S_t} remain constant across all states S_t can be adopted and may be sufficient to model some empirical count data series.

4. SOME PROPERTIES OF THE MODEL

Denote the unconditional probabilities of a Markov chain at time t as $P(S_t = i)$, and probabilities of all possible outcomes at time t as column vector $\mathbf{u}_t = (P(S_t = 1), P(S_t = 2), \ldots, P(S_t = m)), t \in \mathbb{N}$, m representing the number of Markov states. Let $\mathbf{\Gamma}(t) = \{\gamma_{ij}(t)\}, i = 1, 2, \ldots, m, j = 1, 2, \ldots, m$, where $\gamma_{ij}(t) = P(S_{s+t} = j | S_s = i)$, and the mean of the Poisson process constructed as $\lambda_{t,S_t} = \omega_{S_t} + \sum_{i=1}^{q} \alpha_{i,S_t} X_{t-i}$. In order to express the expected mean and variance of observation X_t by vector and matrix calculation, define the row vector of mean of the Poisson process under different states as $\boldsymbol{\lambda}_t = (\lambda_t(S_t = 1), \lambda_t(S_t = 2), \ldots, \lambda_t(S_t = m))$. Let $\boldsymbol{\delta} = \mathbf{u}_1 = (P(S_1 = 1), P(S_1 = 2), \ldots, P(S_1 = m))$ as the initial distribution of the Markov chain, and the $\mathbf{u}_t$ could be deduced from relation $\mathbf{u}_{t+1} = \mathbf{u}_t \mathbf{\Gamma}(t)$. We restrict the scope of the study here within homogeneous Markov chain model, thus $\mathbf{\Gamma}(t)$ will be abbreviated as $\mathbf{\Gamma}$. So we have

$$\mathbf{u}_t = \boldsymbol{u}_{t-1} \mathbf{\Gamma} = \boldsymbol{\delta} \mathbf{\Gamma}^{t-1},$$

$$E(X_t) = \mathbf{u}_t \boldsymbol{\lambda}_t = \boldsymbol{\delta} \mathbf{\Gamma}^{t-1} \boldsymbol{\lambda}_t,$$

$$
\begin{aligned}
Var(X_t) &= E[Var(X_t|S_t)] + Var[E(X_t|S_t)] \\
&= E(X_t) + Var(\prod_{S_t=1}^{m} \lambda_{t,S_t} I_{\{S_t=i\}}) \\
&= E(X_t) + E[(\prod_{S_t=1}^{m} \lambda_{t,S_t} I_{\{S_t=i\}})^2] + [E(\prod_{S_t=1}^{m} \lambda_{t,S_t} I_{\{S_t=i\}})]^2 \\
&= \boldsymbol{\delta}\boldsymbol{\Gamma}^{t-1}\boldsymbol{\lambda}_t + \boldsymbol{\delta}\boldsymbol{\Gamma}^{t-1}\boldsymbol{\lambda}_t^2 + (\boldsymbol{\delta}\boldsymbol{\Gamma}^{t-1}\boldsymbol{\lambda}_t)^2.
\end{aligned}
$$

5. LIKELIHOOD FUNCTION AND PARAMETER ESTIMATION

Let $\boldsymbol{\theta} \equiv (\gamma_{kj}, \omega_{S_t}, \alpha_{i,S_t})$ for $i = 1, 2, \ldots, q$, $k, j = 1, 2, \ldots, m$ represents all parameters in autoregressive conditional Poisson hidden Markov model. The log-likelihood function for the model is given by

$$
l_T(\boldsymbol{\theta}) = log\Big(P(\boldsymbol{X}_T = \boldsymbol{x}_T)\Big) = log(\boldsymbol{\delta}\boldsymbol{P}(x_1)\boldsymbol{\Gamma}\boldsymbol{P}(x_2)\cdots\boldsymbol{\Gamma}\boldsymbol{P}(x_T)\mathbf{1}'),
$$

where $\boldsymbol{\delta}$ is the initial distribution and

$$
\boldsymbol{P}(x_t) = \begin{bmatrix} p_1(x_t) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & p_m(x_t) \end{bmatrix},
$$

with diagonal elements of the matrix defined as

$$
p_i(x_t) = P(X_t = x_t|S_t = i),
$$

and the matrix $\boldsymbol{\Gamma}$defined as

$$
\boldsymbol{\Gamma} = \begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1m} \\ \vdots & \ddots & \vdots \\ \gamma_{m1} & \cdots & \gamma_{mm} \end{bmatrix}.
$$

For the discrete case, elements in the likelihood function become progressively smaller as t increases, scaling the forward probabilities is a common way to avoid underflow. The steps to scale the forward probabilities are shown as below.

$$\boldsymbol{\beta}_0 = \boldsymbol{\delta}\boldsymbol{P}(x_1),$$

with $\boldsymbol{\beta}_t$ defined recursively as

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1}\boldsymbol{\Gamma}\boldsymbol{P}(x_t), \quad for\ t = 2, 3, \ldots, T.$$

with

$$\boldsymbol{\phi}_0 = \boldsymbol{\delta},$$

$$\boldsymbol{\phi}_t = \frac{\boldsymbol{\beta}_t}{\omega_t},$$

$$\omega_t\boldsymbol{\phi}_t = \omega_{t-1}\boldsymbol{\phi}_{t-1}\boldsymbol{B}_t,$$

where

$$\omega_t = \sum_i \beta_t(i) = \boldsymbol{\beta}_t\mathbf{1}',$$

$$\omega_0 = \boldsymbol{\delta}\mathbf{1}'.$$

Thus the scaled log likelihood function would be

$$log(L_T) = \sum_{t=1}^{T} log\frac{\omega_t}{\omega_{t-1}} = \sum_{t=1}^{T} log(\boldsymbol{\phi}_{t-1}\boldsymbol{B}_t\mathbf{1}'). \tag{1}$$

Note that the EM algorithm could also be derived and used. However, there are some parts required for the algorithm that do not have closed form of solutions, leading to complications in the computations. Thus maximum likelihood estimation method is more convenient and gives better estimates.

6. THE MONTE-CARLO SIMULATION STUDY

We conducted a Monte-Carlo simulation study to investigate the performance of maximum likelihood estimators of ACP-HMM model, the log likelihood function used is defined as (1). A simulation study was also used to investigate the use of Akaike information criterion (AICc) and Bayesian information criterion (BIC) criteria to differentiate between highly correlated count data and regular Poisson HMM process.

The properties of estimates were studied across different combinations of parameters using 1,000 simulation runs for each combination. Bias, mean squared error (MSE) and mean absolute deviation (MAD) were computed for each of the parameter combination sets. In order to eliminate the artifacts arising out of initial conditions, the first 240 time series data points were discarded.

We provide the parameter sets used in simulation study here before we move to details of each cases.

Case 1. A time series of count data with 2 states and 2 lags is generated. Each state has the same lag coefficients.

$$\mathbf{\Gamma} = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}, \omega_1 = 20, \omega_2 = 10, \alpha_1 = 0.1, \alpha_2 = 0.2.$$

Case 2. A time series of count data with 3 states and 2 lags is generated. Each state has the same lag coefficients.

$$\mathbf{\Gamma} = \begin{bmatrix} 0.8 & 0.15 & 0.05 \\ 0.15 & 0.75 & 0.1 \\ 0.05 & 0.15 & 0.8 \end{bmatrix}, \omega_1 = 20, \omega_2 = 13, \omega_3 = 8, \alpha_1 = 0.2, \alpha_2 = 0.1.$$

Case 3. A time series of count data with 4 states and 2 lags is generated. Each state has same lag coefficients.

$$\mathbf{\Gamma} = \begin{bmatrix} 0.6 & 0.2 & 0.15 & 0.05 \\ 0.15 & 0.6 & 0.2 & 0.05 \\ 0.05 & 0.2 & 0.6 & 0.15 \\ 0.05 & 0.15 & 0.2 & 0.6 \end{bmatrix}, \omega_1 = 10, \omega_2 = 5, \omega_3 = 15, \omega_4 = 20, \alpha_1 = 0.2, \alpha_2 = 0.3.$$

Case 4. A time series of count data with 2 states and 1 lag is generated. Each state has a different lag coefficient α.

$$\mathbf{\Gamma} = \begin{bmatrix} 0.75 & 0.25 \\ 0.2 & 0.8 \end{bmatrix}, \omega_1 = 10, \omega_2 = 20, \alpha_{1,1} = 0.3, \alpha_{1,2} = 0.2.$$

Case 5. A time series of count data with 3 states and 1 lag is generated. Each state has a different lag coefficient α.

$$\mathbf{\Gamma} = \begin{bmatrix} 0.7 & 0.25 & 0.05 \\ 0.15 & 0.7 & 0.15 \\ 0.05 & 0.2 & 0.75 \end{bmatrix}, \omega_1 = 3, \omega_2 = 8, \omega_3 = 4, \alpha_{1,1} = 0.1, \alpha_{1,2} = 0.5, \alpha_{1,3} = 0.3.$$

Case 6. A time series of count data with 4 states and 1 lag is generated. Each state has a different lag coefficient α.

$$\mathbf{\Gamma} = \begin{bmatrix} 0.6 & 0.2 & 0.15 & 0.05 \\ 0.15 & 0.6 & 0.2 & 0.05 \\ 0.05 & 0.2 & 0.6 & 0.15 \\ 0.05 & 0.15 & 0.2 & 0.6 \end{bmatrix}, \omega_1 = 10, \omega_2 = 5, \omega_3 = 8, \omega_4 = 7, \alpha_{1,1} = 0.2, \alpha_{1,2} = 0.05, \alpha_{1,3} = 0.7, \alpha_{1,4} = 0.3.$$

6.1. CASE 1. A TIME SERIES OF COUNT DATA WITH 2 STATES AND 2 LAGS IS GENERATED. EACH STATE HAS SAME LAG COEFFICIENTS

The simulated data were generated from a ACP-HMM process with 2 states and 2 lags. Each state has same lag coefficients. For Table 1, the true parameter set is

$$\mathbf{\Gamma} = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}, \omega_1 = 20, \omega_2 = 10, \alpha_1 = 0.1, \alpha_2 = 0.2.$$

Table 1. Maximum likelihood estimation results from 1,000 simulations based on different number of Markov states ($m = 2$, $q = 2$)

Parameter	True Coefficient	Estimates	MSE	MAD
p_{11}	0.7	0.70083	0.000973	0.031194
p_{12}	0.3	0.29917	0.000973	0.031194
p_{21}	0.4	0.39909	0.001284	0.035827
p_{22}	0.6	0.60091	0.001284	0.035827
ω_1	20	20.09900	0.796033	0.892207
ω_2	10	10.08003	0.641469	0.800917
α_1	0.1	0.09775	0.000640	0.025289
α_2	0.2	0.19806	0.001001	0.031631

For simulation of the parameter set, sample size T=1,440 was considered, which is comparable to the length of series of minute count data of a day or daily count data of approximately 4 years. Maximum likelihood estimation results from 1,000 simulations based on the above sample size with a time series of count data with 2 states and 2 lags were generated. Each state has same lag coefficients.

6.2. CASE 2. A TIME SERIES OF COUNT DATA WITH 3 STATES AND 2 LAGS IS GENERATED. EACH STATE HAS SAME LAG COEFFICIENTS

The simulated data were generated from a ACP-HMM process with 3 states and 2 lags. Each state has same lag coefficients.

Table 2. Maximum likelihood estimation results from 1,000 simulations based on different number of Markov states ($m = 3$, $q = 2$).

Parameter	True Coefficient	Estimates	MSE	MAD
p_{11}	0.8	0.75734	0.025900	0.069293
p_{12}	0.15	0.15960	0.017900	0.075879
p_{13}	0.05	0.08306	0.007890	0.053554
p_{21}	0.15	0.16917	0.015500	0.061574
p_{22}	0.75	0.71248	0.030000	0.090094
p_{23}	0.1	0.11835	0.016601	0.066250
p_{31}	0.05	0.08872	0.015400	0.063711
p_{32}	0.15	0.19528	0.029200	0.094614
p_{33}	0.8	0.71600	0.056100	0.113975
ω_1	20	18.88701	7.213412	1.907206
ω_2	13	12.01882	6.581093	1.725269
ω_3	8	7.33770	2.882121	1.068537
α_1	0.2	0.21170	0.001421	0.029035
α_2	0.1	0.13157	0.005563	0.050574

For Table 2, the true parameter set is

$$\mathbf{\Gamma} = \begin{bmatrix} 0.8 & 0.15 & 0.05 \\ 0.15 & 0.75 & 0.1 \\ 0.05 & 0.15 & 0.8 \end{bmatrix}, \omega_1 = 20, \omega_2 = 13, \omega_3 = 8, \alpha_1 = 0.2, \alpha_2 = 0.1.$$

For simulation of the parameter set, sample size T=1,440 was considered, which is comparable to the length of series of minute count data of a day or daily count data of approximately 4 years. Maximum likelihood estimation results from 1,000 simulations based on the above sample size with a time series of count data with 3 states and 2 lags were generated. Each state has same lag coefficients.

6.3. CASE 3. A TIME SERIES OF COUNT DATA WITH 4 STATES AND 2 LAGS IS GENERATED. EACH STATE HAS SAME LAG COEFFICIENTS

The simulated data were generated from a ACP-HMM process with 4 states and 2 lags. Each state has same lag coefficients. For Table 3, the true parameter set is

$$\mathbf{\Gamma} = \begin{bmatrix} 0.6 & 0.2 & 0.15 & 0.05 \\ 0.15 & 0.6 & 0.2 & 0.05 \\ 0.05 & 0.2 & 0.6 & 0.15 \\ 0.05 & 0.15 & 0.2 & 0.6 \end{bmatrix}, \omega_1 = 10, \omega_2 = 5, \omega_3 = 15, \omega_4 = 20, \alpha_1 = 0.2, \alpha_2 = 0.3.$$

For simulation of the parameter set, sample size T=1,440 was considered, which is comparable to the length of series of minute count data of a day or daily count data of approximately 4 years. Maximum likelihood estimation results from 1,000 simulations based on the above sample size with a time series of count data with 4 states and 2 lags were generated. Each state has same lag coefficients.

Table 3. Maximum likelihood estimation results from 1,000 simulations based on different number of Markov states ($m = 4$, $q = 2$).

Parameter	True Coefficient	Estimates	MSE	MAD
p_{11}	0.6	0.59072	0.021195	0.121462
p_{12}	0.2	0.14493	0.020742	0.124613
p_{13}	0.15	0.14698	0.021254	0.128451
p_{14}	0.05	0.11737	0.020622	0.106100
p_{21}	0.15	0.17713	0.020913	0.123070
p_{22}	0.6	0.56450	0.014809	0.092928
p_{23}	0.2	0.16797	0.020054	0.121475
p_{24}	0.05	0.09040	0.012942	0.079465
p_{31}	0.05	0.11546	0.021349	0.108176
p_{32}	0.2	0.13073	0.019580	0.119710
p_{33}	0.6	0.59138	0.024016	0.131772
p_{34}	0.15	0.16243	0.019272	0.116864
p_{41}	0.05	0.13782	0.025745	0.123332
p_{42}	0.15	0.09882	0.015453	0.109651
p_{43}	0.2	0.19901	0.186920	0.400990
p_{44}	0.6	0.56435	0.157511	0.364333
ω_1	10	9.18312	6.838601	2.186901
ω_2	5	4.97969	1.658411	0.994961
ω_3	15	14.55797	6.065909	1.926202
ω_4	20	19.64506	6.986931	2.021010
α_1	0.2	0.18946	0.001192	0.027813
α_2	0.3	0.27415	0.003118	0.044445

6.4. CASE 4. A TIME SERIES OF COUNT DATA WITH 2 STATES AND 1 LAG IS GENERATED. EACH STATE HAS A DIFFERENT LAG COEFFICIENT α

The simulated data were generated from a ACP-HMM process with 2 states and 1 lag. Each state has a different lag coefficient. For Table 4, the true parameter set is

$$\mathbf{\Gamma} = \begin{bmatrix} 0.75 & 0.25 \\ 0.2 & 0.8 \end{bmatrix}, \omega_1 = 10, \omega_2 = 20, \alpha_{1,1} = 0.3, \alpha_{1,2} = 0.2.$$

For simulation of the parameter set, sample size T=1,440 was considered, which is comparable to the length of series of minute count data of a day or daily count data of approximately 4 years. Maximum likelihood estimation results from 1,000 simulations based on the above sample size with a time series of count data with 2 states and 1 lag were generated. Each state has a different lag coefficient α.

Table 4. Maximum likelihood estimation results from 1,000 simulations based on different number of Markov states (m = 2, q = 1).

Parameter	True Coefficient	Estimates	MSE	MAD
p_{11}	0.75	0.74789	0.001008	0.024260
p_{12}	0.25	0.25211	0.001008	0.024260
p_{21}	0.2	0.20461	0.001951	0.025965
p_{22}	0.8	0.79539	0.001951	0.025965
ω_1	10	9.98615	0.470935	0.536199
ω_2	20	20.05603	1.107474	0.833189
$\alpha_{1,1}$	0.3	0.30087	0.001984	0.034581
$\alpha_{1,2}$	0.2	0.19915	0.001885	0.032861

6.5. CASE 5. A TIME SERIES OF COUNT DATA WITH 3 STATES AND 1 LAG IS GENERATED. EACH STATE HAS A DIFFERENT LAG COEFFICIENT α

The simulated data were generated from a ACP-HMM process with 3 states and 1 lag. Each state has a different lag coefficient. For Table 5, the true parameter set is

$$\mathbf{\Gamma} = \begin{bmatrix} 0.7 & 0.25 & 0.05 \\ 0.15 & 0.7 & 0.15 \\ 0.05 & 0.2 & 0.75 \end{bmatrix}, \omega_1 = 3, \omega_2 = 8, \omega_3 = 4, \alpha_{1,1} = 0.1, \alpha_{1,2} = 0.5, \alpha_{1,3} = 0.3.$$

Table 5. Maximum likelihood estimation results from 1,000 simulations based on different number of Markov states ($m = 3$, $q = 1$).

Parameter	True Coefficient	Estimates	MSE	MAD
p_{11}	0.7	0.68446	0.010054	0.067461
p_{12}	0.25	0.21080	0.011515	0.075060
p_{13}	0.05	0.10474	0.018946	0.090441
p_{21}	0.15	0.15906	0.006665	0.060033
p_{22}	0.7	0.68963	0.004557	0.041534
p_{23}	0.15	0.15131	0.007579	0.065768
p_{31}	0.05	0.09801	0.014336	0.080578
p_{32}	0.2	0.18475	0.009468	0.073481
p_{33}	0.75	0.71724	0.012719	0.075851
ω_1	3	2.86848	0.240990	0.314269
ω_2	8	8.22341	1.214701	0.592089
ω_3	4	4.26150	1.426600	0.830676
$\alpha_{1,1}$	0.1	0.23549	0.088734	0.246055
$\alpha_{1,2}$	0.5	0.42867	0.067414	0.204649
$\alpha_{1,3}$	0.3	0.31042	0.037574	0.173398

For simulation of the parameter set, sample size T=1,440 was considered, which is comparable to the length of series of minute count data of a day or daily count data of approximately 4 years. Maximum likelihood estimation results from 1,000 simulations based on the above sample size with a time series of count data with 3 states and 1 lag were generated. Each state has a different lag coefficient α.

6.6. CASE 6. A TIME SERIES OF COUNT DATA WITH 4 STATES AND 1 LAG IS GENERATED. EACH STATE HAS A DIFFERENT LAG COEFFICIENT α

The simulated data were generated from a ACP-HMM process with 4 states and 1 lag. Each state has a different lag coefficient. For Table 5, the true parameter set is

$$\mathbf{\Gamma} = \begin{bmatrix} 0.6 & 0.2 & 0.15 & 0.05 \\ 0.15 & 0.6 & 0.2 & 0.05 \\ 0.05 & 0.2 & 0.6 & 0.15 \\ 0.05 & 0.15 & 0.2 & 0.6 \end{bmatrix}, \omega_1 = 10, \omega_2 = 5, \omega_3 = 15, \omega_4 = 20, \alpha_1 = 0.2, \alpha_2 = 0.3.$$

For simulation of the parameter set, sample size T=1,440 was considered, which is comparable to the length of series of minute count data of a day or daily count data of approximately 4 years. Maximum likelihood estimation results from 1,000 simulations based on the above sample size with a time series of count data with 4 states and 1 lag were generated. Each state has a different lag coefficient α.

As was down in previous cases, we kept the probabilities of staying in the current state much higher than transitioning to a different state. This was intentionally down in order to make the state changes infrequent. The ACP-HMM model was built to allow the ACP component to model the high frequency changes in the count data and the HMM part to model the low frequency changes in the states.

Table 6. Maximum likelihood estimation results from 1,000 simulations based on different number of Markov states ($m = 4$, $q = 1$).

Parameter	True Coefficient	Estimates	MSE	MAD
p_{11}	0.6	0.52966	0.024245	0.129561
p_{12}	0.2	0.16533	0.017840	0.110500
p_{13}	0.15	0.15392	0.021644	0.128321
p_{14}	0.05	0.15109	0.030337	0.133185
p_{21}	0.15	0.10575	0.016595	0.113758
p_{22}	0.6	0.55517	0.012743	0.078093
p_{23}	0.2	0.16731	0.021481	0.125965
p_{24}	0.05	0.17177	0.033844	0.146897
p_{31}	0.05	0.12094	0.020470	0.107010
p_{32}	0.2	0.17373	0.011320	0.083275
p_{33}	0.6	0.53404	0.019775	0.108371
p_{34}	0.15	0.17129	0.017105	0.107620
p_{41}	0.05	0.14922	0.028730	0.129490
p_{42}	0.15	0.16034	0.013154	0.094199
p_{43}	0.2	0.17588	0.202360	0.424126
p_{44}	0.6	0.51456	0.113191	0.314562
ω_1	10	11.04803	4.878910	1.592500
ω_2	5	4.67898	0.664762	0.494271
ω_3	8	8.76332	2.276821	1.220924
ω_4	7	6.86989	1.973631	1.184100
$\alpha_{1,1}$	0.2	0.26203	0.053422	0.183901
$\alpha_{1,2}$	0.05	0.04577	0.000885	0.024387
$\alpha_{1,3}$	0.7	0.70129	0.008947	0.068687
$\alpha_{1,4}$	0.3	0.28628	0.036385	0.149790

Based on the simulation results, it can be seen that the maximum likelihood estimation method provides relatively good estimates for the parameters with low Monte Carlo standard errors. For cases that $m \leq 3$, the estimates are pretty accurate since the estimates are close to the true coefficients. When the number of states increases, fitting models with many omega and alpha values becomes more difficult. But the variance is still pretty reasonable given the fact that Poisson hidden Markov models give bigger variance [18]. This demonstrates the MLE is a promising method for estimating the parameters of the suggested autoregressive conditional Poisson hidden Markov model (ACP-HMM).

7. MODEL SELECTION

To illustrate the importance of selecting the correct structure of count data and also examine if AICc and/or BIC are good criteria to distinguish the true generation process, a small scale Monte Carlo simulation study was performed. All statistics reported here are calculated from N=1,000 replications and each replication having sample size T=1,440. In order to avoid artifacts created by initial conditions, the first 240 time series data points were discarded.

Table 7 shows results for the case when the data were generated from an ACP-HMM process with true parameters

$$\mathbf{\Gamma} = \begin{bmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{bmatrix}, \omega_1 = 5, \omega_2 = 20, \alpha_1 = 0.3, \alpha_2 = 0.1.$$

Both ACP-HMM and Poisson HMM Model were ultilized to fit the data. The AICc and BIC values for ACP-HMM are lower than those for the Poisson HMM model, which suggests AICc and BIC perform well in identifying the true structure of the time series.

Table 7. Poisson HMM and ACP-HMM selection by AICc and BIC criteria with simulated time series data for small α's that do not differ much.

Parameter	True Coefficient	ACP-HMM	Poisson HMM
p_{11}	0.7	0.70224	0.73894
p_{12}	0.3	0.29776	0.26106
p_{21}	0.2	0.20249	0.15565
p_{22}	0.8	0.79751	0.84435
ω_1	5	5.02370	12.21275
ω_2	20	20.01910	29.89603
α_1	0.3	0.29968	-
α_2	0.1	0.10057	-
AICc		10181	10792
BIC		10213	10827

Table 8 shows results when an ACP-HMM process is the underlying structure producing the count data with true parameters

$$\mathbf{\Gamma} = \begin{bmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{bmatrix}, \omega_1 = 2.5, \omega_2 = 9, \alpha_1 = 0.8, \alpha_2 = 0.1.$$

Both ACP-HMM model and Poisson HMM model were utilized to fit the data. In this case, AICc and BIC also showed their ability to select the right structure. Note if the data generating process of a count data time series has an autoregressive conditional heteroskedastic structure, and its parameters are governed by a hidden Markov process, then the regular Poisson HMM provides a poor fit, especially when one or more of the α's are high.

Table 8. Poisson HMM and ACP-HMM selection by AICc and BIC criteria with simulated time series data with one large α and the other small.

Parameter	True Coefficient	ACP-HMM	Poisson HMM
p_{11}	0.7	0.71857	0.92571
p_{12}	0.3	0.28143	0.07429
p_{21}	0.2	0.20091	0.07384
p_{22}	0.8	0.79909	0.92616
ω_1	2.5	2.56610	43.83731
ω_2	9	8.90310	68.34237
α_1	0.8	0.79107	-
α_2	0.1	0.09657	-
AICc		10461	12650
BIC		10493	12686

8. VISUALIZATION OF SIMULATED DATA

In order to provide a visual representation of how the ACP-HMM formulation behaves, data was generated from parameter set

$$\mathbf{\Gamma} = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}, \omega_1 = 20, \omega_2 = 10, \alpha_1 = 0.1, \alpha = 0.2.$$

In Figure 1, the grey line represents the simulated data while the blue line indicates the underlying mean of the Poisson process. The red bars at the bottom indicate time periods where the underlying process is at State 1 while the green bars indicate the periods in which the process is at State 2.

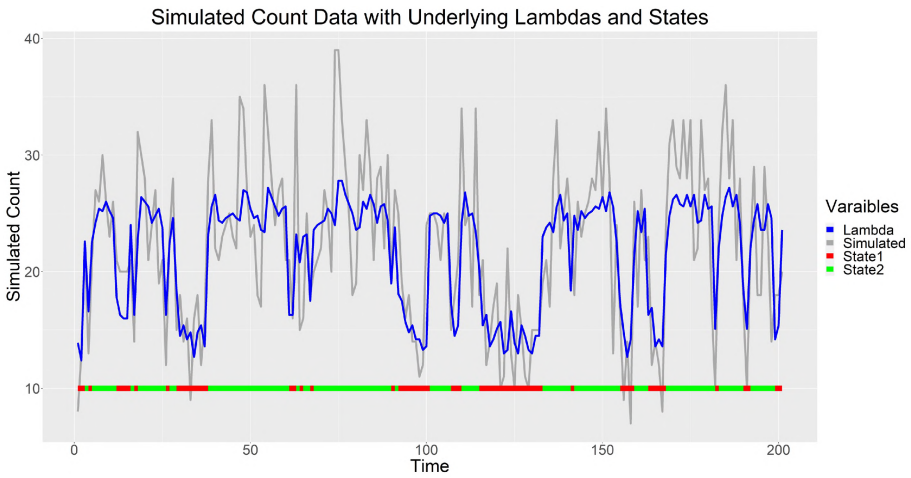


Figure 1. Simulated time series count data, and the underlying λ_{t,S_t} and states.

9. APPLICATION TO A REAL-LIFE DATA SET

Figure 2 demonstrates the daily number of deaths in Evora, Portugal from 01/01/1996 to 12/31/2007. The sample mean equals 6.119 and the variance is 7.483. There seems to be irregular periodicity present in this time series.

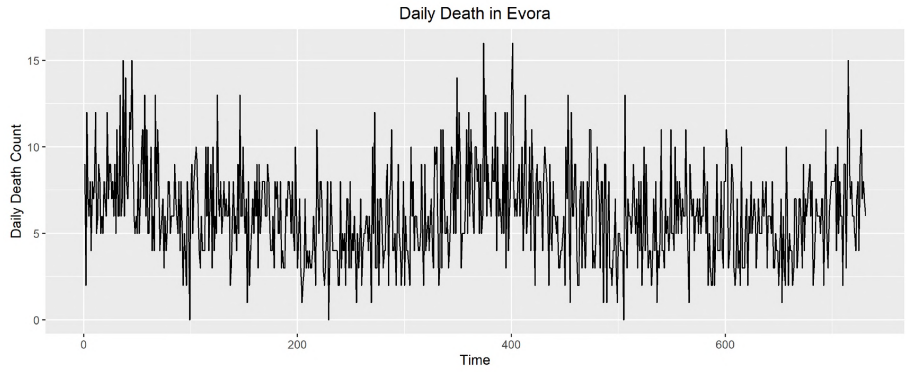


Figure 2. Daily Death in Evora from 01/01/1996 to 12/31/2007.

The auto-correlation function plot of the count data (Figure 3) suggests there is auto-correlation in the count data and hence ACP structure is better than regular Poisson process. Some irregular periodicity is also observed, hence the motivation for fitting an ACP-HMM could be seen.

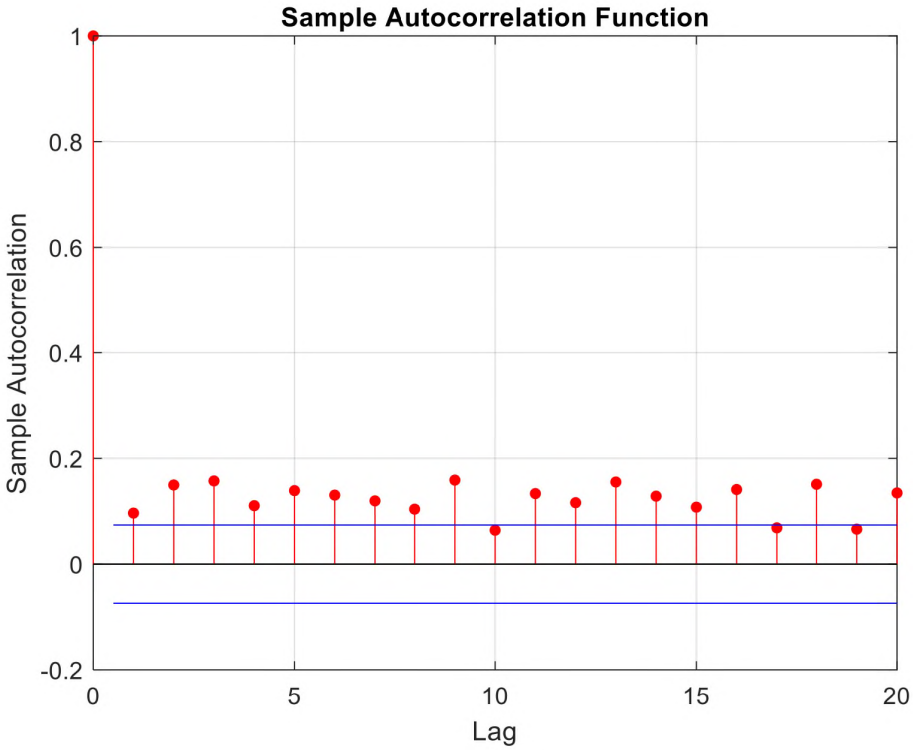


Figure 3. Autocorrelation of Daily Death Count Data.

Table 9. Daily death data fitted by Poisson HMM and ACP-HMM with AICc provided.

Parameter	ACP-HMM	Poisson HMM
p_{11}	0.9954	0.999999
p_{12}	0.0046	0.000001
p_{21}	0.0127	0.0422
p_{22}	0.9873	0.9578
ω_1	5.5312	6.1230
ω_2	6.8751	33.0782
$\alpha_{1,1}$	0.0001	-
$\alpha_{1,2}$	0.1086	-
AICc	3445.2	3533.7

Daily death data is fitted by Poisson HMM and ACP-HMM with 2 states and their corresponding AICc's are provided. Without considering the correlation inside the data, Poisson HMM gives a pretty large mean for the second Poisson process and the transition matrix has extreme estimates for its probabilities. Based on two AICc, the ACP-HMM provides a better fit and indicates a need for this generalization of the poisson HMM in real data.

10. CONCLUSION

The model provided here is a natural generalization of the Poisson hidden Markov model, which takes the influence of previous observations into consideration when modeling auto-correlated count time series. The reported simulation results in Section 6 show that the MLE method provides reasonable estimates of the model parameters of the ACP-HMM model. However, when the number of states gets larger, the estimations are not very accurate and this is especially seen in estimates of transition probabilities that are close to zero. The study of the utility of using AICc and BIC criteria in determining the true structure of the count data shows promising results. Finally, we use a real-life data set to illustrate the importance of developing such a model.

REFERENCES

[1] Paul S Albert. A two-state markov mixture model for a time series of epileptic seizure counts. *Biometrics*, pages 1371–1381, 1991.

[2] PoC Consul and Felix Famoye. Generalized poisson regression model. *Communications in Statistics-Theory and Methods*, 21(1):89–109, 1992.

[3] Stefany Coxe, Stephen G West, and Leona S Aiken. The analysis of count data: A gentle introduction to poisson regression and its alternatives. *Journal of personality assessment*, 91(2):121–136, 2009.

[4] Joseph M Hilbe. *Modeling count data*. Cambridge University Press, 2014.

[5] John Hinde. Compound poisson regression models. In *GLIM 82: Proceedings of the International Conference on Generalised Linear Models*, pages 109–121. Springer, 1982.

[6] Chester Ittner Bliss and Ronald A Fisher. Fitting the negative binomial distribution to biological data. *Biometrics*, 9(2):176–200, 1953.

[7] Andreas Lindén and Samu Mäntyniemi. Using the negative binomial distribution to model overdispersion in ecological count data. *Ecology*, 92(7):1414–1421, 2011.

[8] Scott L Zeger. A regression model for time series of counts. *Biometrika*, 75(4):621–629, 1988.

[9] Ed McKenzie. Some simple models for discrete variate time series 1. *JAWRA Journal of the American Water Resources Association*, 21(4):645–650, 1985.

[10] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[11] Patricia A Jacobs and Peter AW Lewis. Stationary discrete autoregressive-moving average time series generated by mixtures. *Journal of Time Series Analysis*, 4(1):19–36, 1983.

[12] Eddie McKenzie. Discrete variate time series. *Handbook of statistics*, 21(573606):1, 2003.

[13] René Ferland, Alain Latour, and Driss Oraichi. Integer-valued garch process. *Journal of Time Series Analysis*, 27(6):923–942, 2006.

[14] Fukang Zhu. A negative binomial integer-valued garch model. *Journal of Time Series Analysis*, 32(1):54–67, 2011.

[15] Fukang Zhu. Modeling overdispersed or underdispersed count data with generalized poisson integer-valued garch models. *Journal of Mathematical Analysis and Applications*, 389(1):58–71, 2012.

[16] Cathy WS Chen, Mike KP So, Jessica C Li, and Songsak Sriboonchitta. Autoregressive conditional negative binomial model applied to over-dispersed time series of counts. *Statistical Methodology*, 31:73–90, 2016.

[17] Andréas Heinen. Modelling time series count data: an autoregressive conditional poisson model. *Available at SSRN 1117187*, 2003.

[18] Iain L MacDonald and Walter Zucchini. *Hidden Markov and other models for discrete-valued time series*, volume 110. CRC Press, 1997.

[19] Siddhartha Chib and Rainer Winkelmann. Markov chain monte carlo analysis of correlated count data. *Journal of Business & Economic Statistics*, 19(4):428–435, 2001.

[20] Adrian E Raftery. A model for high-order markov chains. *Journal of the Royal Statistical Society: Series B (Methodological)*, 47(3):528–539, 1985.

[21] Robert C Jung and Andrew R Tremayne. Coherent forecasting in integer time series models. *International Journal of Forecasting*, 22(2):223–238, 2006.

[22] Leonard E Baum and Ted Petrie. Statistical inference for probabilistic functions of finite state markov chains. *The annals of mathematical statistics*, 37(6):1554–1563, 1966.

[23] Leonard E Baum and John Alonzo Eagon. An inequality with applications to statistical estimation for probabilistic functions of markov processes and to a model for ecology. *Bulletin of the American Mathematical Society*, 73(3):360–363, 1967.

[24] Biing Hwang Juang and Laurence R Rabiner. Hidden markov models for speech recognition. *Technometrics*, 33(3):251–272, 1991.

[25] Anthony Hopkins, Philip Davies, and Charles Dobson. Mathematical models of patterns of seizures: their use in the evaluation of drugs. *Archives of neurology*, 42(5):463–467, 1985.

[26] Stacia M DeSantis and Dipankar Bandyopadhyay. Hidden markov models for zero-inflated poisson counts with an application to substance use. *Statistics in medicine*, 30(14):1678–1694, 2011.

III. PREDICTING LIFESPAN OF DROSOPHILA MELANOGASTER: A NOVEL APPLICATION OF CONVOLUTIONAL NEURAL NETWORKS AND ZERO-INFLATED AUTOREGRESSIVE CONDITIONAL POISSON MODEL

Yi Zhang[a], V.A. Samaranayake[a], Gayla R. Olbricht[a], Matthew Thimgan[b]
[a]Department of Mathematics and Statistics
Missouri University of Science and Technology
Rolla, Missouri 65409–0050
[b]Department of Biological Science
Missouri University of Science and Technology
Rolla, Missouri 65409–0050

ABSTRACT

A model to classify the lifespan of the fruit fly is developed using a two-stage process. Stage one models the per minute activity counts of each fly using a zero-inflated autoregressive conditional Poisson model. These probabilities are allowed to vary hourly, reflecting the circadian and other cycles present in a fly's sleep architecture. A five-day moving window was used to model data from five days at a time, allowing the model parameters to vary over the course of the fly's life. The resulting probabilities capture information about changes in sleep patterns with age and are hypothesized to contain features that help categorize flies into short and long-lived groups. The resulting hourly zero-inflation probabilities over a 24 day period are utilized to create a "heat map" containing information on the 24-hour daily sleep cycle and its changes across the 24-day observation period. In stage two, the heat maps for individual flies were used as inputs to a convolutional neural network to build a classification model. The estimated model provides a reasonably accurate way to group flies into lifespan categories. Grouping flies into such categories would facilitate the discovery of biochemical markers of aging by assaying groups of short and long-lived flies.

Keywords: Convolutional Neural Networks; Sleep Models; Fruit Fly; Integer-valued Time Series; Poisson Processes.

1. INTRODUCTION

Sleep has been found in every animal that has been rigorously evaluated. Adequate sleep is essential for good health [1] and poor sleep can lead to significant health issues, both physical and mental [2]. The study of sleep is not only important to understanding the relationship between inadequate sleep on both individual and population health [3], but is also central to determining its effect on lifespan [4]. For example, Gallicchio & Kalesan [5] employed meta-analysis to establish a relationship between sleep duration and mortality. Additional evidence has accumulated that both short and long durations of sleep contribute to a higher risk of mortality [6, 7]. Moreover, it is known that sleep patterns are affected by aging in the central nervous system [8]. Ohayon et al. [9] added to this evidence and showed that total sleep time and sleep efficiency decrease with age, but the reverse occurred with sleep latency. Unfortunately, studying how age-related changes in sleep architecture relate to human mortality is hampered by the decades' long lifespan. Alternatively, such investigations can be conducted using model organisms with short lifespans and high molecular and biochemical homology to human sleep regulation. The use of model organisms can also enhance the search for chemical biomarkers that are linked to biological aging. In this paper, we present the results of a study of the model organism *Drosophila melanogaster*, the fruit fly, to investigate if a single sleep characteristic of the flies derived from their per-minute activity counts can be utilized to classify study specimens into long- and short-lived categories. This is the first step in a series of studies, with the ultimate goal of building a predictive model based on multiple sleep characteristics to isolate flies into short- and long-lived categories, based on the sleep data in the early part of their

lives. The classified flies can be assayed to determine biochemical markers of aging, thus helping to understand the role sleep plays in affecting the lifespan of this model organism and ultimately that of humans.

The fruit fly, *Drosophila melanogaster*, has emerged as an ideal model organism to study sleep [10, 11]. Sleep has been defined in the fly by measuring the quiescent episodes [12], and findings in the fly are relevant to understanding human sleep and health. Flies have their primary sleep period at night, the same neurotransmitters control sleep and wakefulness in both organisms, and proteins conserved between humans and flies both regulate sleep and wakefulness [12, 13]. The relatively short lifespan of the fruit fly allows one to evaluate sleep data continuously over their entire lifespan. In addition, *Drosophila* has been used to generate and evaluate numerous novel molecules and hypotheses in the aging field [14]. Preliminary analyses have demonstrated that features of sleep architecture associated with lifespan in the fly also apply to the understanding of human sleep and health [15]. However, quantitative studies of how sleep patterns affect lifespan in the fly are currently limited, and the results reported herein would be valuable in understanding how to improve sleep to limit the health impacts of inadequate sleep or to indirectly identify the onset of a decline in health.

In this study, per-minute activity counts of each fly are modelled using an autoregressive conditional Poisson (ACP) formulation with zero-inflation, with the zero-inflation parameter allowed to vary from hour to hour across each day, thus accounting for circadian rhythms and other cyclical activity patterns. The experiment was conducted under light and dark periods, each lasting 12 hours, which coordinates with the circadian rhythm of the *Canton Special* (*CS*) genotype studied. We hypothesize that the zero-inflation parameter is a proxy for the propensity to sleep during a given hour and thus models a particular aspect of the sleep architecture. This study has two intertwined goals. The first is to develop a predictive model that can be utilized to separate individual flies into short- and long-lived groups, and the second is to determine if the daily profile of zero-inflation probabilities

contain information associated with aging. To this end, a convolutional neural network (CNN) is trained on 'heat maps' generated from the estimated daily profile of zero-inflation probabilities to obtain a classification model.

In Section 2, we introduce a temporally dependent Poisson model with a zero-inflation parameter that varies with time. A brief introduction to CNNs and the reasons for its adoption in this study are presented in Section 3, followed by a discussion of the data collection and analysis given in Section 4. The results of the analyses are presented in Section 5. In Section 6, a discussion of the results is presented together with suggestions for future work. The paper is concluded with a summary given in Section 7.

2. ZERO-INFLATED POISSON MODEL

The first step in our approach is to build a count data model that incorporates zero-inflation, with the zero activity counts indicating possible periods of sleep. The *CS* genotype flies under study have a regular sleep pattern that repeats with approximately a 24-hour period. Thus, a model that accounts for both a cyclically varying zero-inflation and serial dependence is needed.

Heinen [16] introduced an ACP model to deal with time series of count data with serial correlation. Building on Heinen's work, the theoretical properties of the general ACP model were derived by Ghahramani and Thavaneswaran [17]. In brief, the count X_t observed during the interval (t-1, t] considered a realization from a Poisson distribution with a conditional mean that is dependent on past observations and past conditional means. The process is formally defined as follows. Let the count data series be denoted by $\{X_t : t \in \mathbb{N}\}$ and let $\mathcal{F}_t$ denote the sigma field generated by the set of random variables $\{X_i : i \leq t\}$. The counts are assumed to be realizations from a Poisson distribution with a conditional mean λ_t, whose dependence on past conditional means and counts parallels the GARCH model of Bollerslev [18]:

$$X_t|\mathcal{F}_{t-1} \sim Pois(\lambda_t)$$

$$E\,(X_t \mid \mathcal{F}_{t-1}) = \lambda_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j},$$

under the condition that all of $\alpha_i's$, $\beta_j's$ are positive.

The above formulation is suitable for modelling discrete time series with overdispersion while taking serial correlation into consideration, with the latter property being of interest given that temporal dependence is not uncommon in times series of count data.

Zhu [19] proposed a generalized version of the ACP model by replacing the Poisson assumption with a zero-inflated Poisson (*ZIP*) distribution that has a zero-inflation parameter ω. A $ZIP\,(\lambda, \omega)$ distribution can be characterized by the probability mass function:

$$P\,(X=k) = \omega\delta_{k,0} + (1-\omega)\,\frac{e^{\lambda}\lambda^{-k}}{k!},\ k \in \mathbb{N} \cup \{0\},$$

where $0 < \omega < 1$ and $\delta_{k,0}$ is the Kronecker delta such that

$$\delta_{k,0} = \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{if } k \neq 0 \end{cases}.$$

Thus, under the formulation of Zhu [19], the time series $\{X_t : t \in \mathbb{N}\}$, conditional on $\mathcal{F}_{t-1}$, satisfy

$$X_t | \mathcal{F}_{t-1} \sim ZIP\,(\lambda_t, \omega),$$

$$\lambda_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j},$$

where $0 < \omega < 1$, $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, for $i = 1, 2, \ldots, p$ and $j = 1, 2, \ldots, q$.

The above model assumes that the zero-inflation probability is a constant over time, and this condition was relaxed by Ratnayake and Samaranayake [20]. This relaxation allows the zero-inflation probabilities to be influenced by an exogenous random variable or a function of time. Employing this concept, the following model is defined to allow zero-inflation probabilities to vary from hour to hour, over the 24 hours of a given day, to accommodate the circadian rhythm associated with sleep found in the *CS* fly. Defining

$\{X_t : t \in \mathbb{N}\}$ and $\mathcal{F}_t$ as before, we let

$$P\,(X_t{=}k \mid \mathcal{F}_{t-1}) = \omega_t \delta_{k,0} + (1-\omega_t)\,\frac{\lambda_t^k e^{-\lambda_t}}{k!},$$

where

$$\lambda_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j},$$

under the constraints $0 < \omega_t < 1$, $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, for $i = 1, 2, \ldots, p$, $j = 1, 2, \ldots, q$, and $\delta_{k,0}$ is the Kronecker delta defined previously, with $\omega_t = \sum_{i=1}^{24} \omega_i\, I_i(t)$. Note that $I_i(t)$, $i = 1, 2, \ldots, 24$ is a set of indicator functions that are zero except when $i = t \pmod{24}$. The above formulation will be referred to hereafter as the variable zero-inflated ACP model or the VZI-ACP model.

When applied to the activity count data obtained from the flies, we assume that for a given fly, the zero-inflation probability, ω_i, for the hour i of a day remains constant over that 1-hour period. Since the count time series of each fly is modelled individually, these probabilities may vary not only across the hours but also across flies as well. In fact, the other parameters of the model may also differ across flies. Thus the counts could be denoted by $X_{l,t}$, where l signifies the fly, and all the parameters of the above model can also be indexed by the subscript l, but for simplicity this additional indexing is avoided.

To allow for the variation of the zero-inflation probabilities across the lifespan of a single fly, the above model is fitted to 1-minute activity count data over a 5-day window, with this window shifted forward 1 day at a time. This approach yields parameter estimates that vary from day to day within a fly, including the 24 zero-inflation parameters. Results from fitting the VZI-ACP model to the fly activity data are discussed in Section 5. Note that each zero-inflation probability is estimated based on 300 data points.

The zero-inflation probabilities estimated from the VZI-ACP model are converted to images, as described in Sections 3 and 5, and used as input to a CNN trained to carry out the classification of the flies into short and long-lived groups.

The question of whether the above VZI-ACP model is parsimonious is a natural one due to the presence of 24 zero-inflation parameters. For an individual fly, there may be adjoining hours on a particular day where a single zero-inflation parameter could be used. Using a common zero-inflation parameter across a particular multihour period would only work if this is a reasonable assumption across all 5 days used for modelling and the data show that this is not the case. In addition, the zero-inflation-based heat map inputs to the CNN must contain identical features for all the flies used in this study. Note that it is assumed that the other parameters remain constant while the zero-inflation probabilities, reflecting the circadian nature of the sleep-wake process, cycles over a given 24-hour period. Thus, the information about sleep architecture is captured only by the zero-inflation probabilities and not by the other model parameters. A potential alternative to the VZI-ACP model is found in Xu *et al.* [21]. While this model has advantages, its adaptive property allows both the zero-inflation and other model parameters to vary over the 24-hour cycle. Thus, sleep -related information will be captured by more than one variable and some may also capture other dynamics unrelated to sleep. In its current form, it is not suitable for our immediate objective of isolating a single variable that can be directly attributed to an aspect of the sleep architecture, namely, the propensity to sleep. As indicated before, this is the first in a series of studies that are planned, and an adaptation of Xu's model will be considered in future studies.

3. CONVOLUTIONAL NEURAL NETWORKS (CNN)

In the past few decades, CNNs have attracted wide attention, especially in the context of image and video classification [22, 23, 24]. Applications of CNNs in medical image analysis have become common [25, 26, 27] and their utility has extended to modelling financial time series [28, 29]. In another application, researchers used handwritten digits (available in the MNIST database) as input and then classified the observations into corresponding numbers using a CNN successfully [30, 31]. The use of machine learning algorithms has

extended to the study of sleep. For example, Oropesa et al. [32] combined a wavelet transform and an artificial neural network (ANN) for sleep stage classification. Recently, machine learning (including CNN) has been applied to classify sleep stages [33, 34, 35], which provides new and unambiguous insights into sleep organization. However, there has been no study that employs a CNN to classify *Drosophila melanogaster* into age categories using activity count-based information.

The typical input to a CNN is a two-dimensional image or a tensor, and when trained on an appropriate data set, it can differentiate input images or tensors into different categories. For a detailed description of CNNs the reader is referred to Aggarwal [36]. In addition, Sewak *et.al.* [37] provide a practical guide on the implementation of CNNs.

The architecture of CNNs is, in a sense, based on the visual cortex of the human brain, where each portion of an image is processed by a small set of neurons, called receptive fields, with multiple such neuron sets processing overlapping segments of an image [36]. In general, a CNN consists of an input layer, multiple inner layers, and an output layer. The first set of inner layers act similar to the receptive fields, processing overlapping segments of the image. These are called convolution layers, and the convolution of the input image pixels occurs by the application of a moving linear filter to segments of the image. The objective of this processing is to extract high-value features (e.g., edges). The convolution operations can be done using one or more layers. Following the convolution layers is the pooling layer, which reduces the dimension of the image. In max pooling, the maximum value of a rectangular section of the convolved image is returned, while in average pooling, the average of the values observed in that segment is used. The resulting outputs are then fed into a traditional neural network (NN) to obtain the output [36].

In the present study, a 24x24 pixel image that may be viewed as a 'heat map' is constructed using the hourly zero-inflation probabilities for a given day displayed along each row, and each of the 24 columns representing a particular hour of the day. There are exactly 24 rows, representing the first 24 days for which the activity data are available for

each fly. The number of days in the study is not extended beyond Day 24 so as to enable the inclusion of data for flies with short lifespans. Note that to estimate the zero-inflation probabilities for Day 24, one needs data from Days 24-28.

A CNN was selected as the platform for building the classification model because of two reasons. The image inputs required by the CNN allow us to construct images that display the daily 24 zero-inflation probabilities in rows, preserving the temporal relationship between them within a day, while at the same time allowing the display of the zero-inflation probabilities for a specific hour across 24 days as a column, not only preserving the day-to-day temporal relationship, but also highlighting any age related changes in this parameter. More importantly, subtle changes that may be related to biological aging may occur at different points in the lifespan and on different hours of the day for different flies. That is, the image pixels that relate to lifespan may appear at slightly different positions of the 24x24 image for each fly. The CNNs have what is known as the translational invariance property (see [36]), which in lay terms means that key features of the image need not be at a fixed position for the CNN to perform well as a classifier. For example, images of dogs and cats need not be centered exactly when a CNN is used to classify them into dog and cat categories. This is very important in the current context because key sleep related features that change with biological aging are not found at exactly the same set of hours in each fly, and these changes may not occur at exactly the same point in time relative to their age. In addition, the duration between a fly's eclosing and the first full day in which the data were collected differed by several hours across the flies because of different lab conditions, and thus the translational invariance property becomes essential for the proper analysis of the data.

In addition, CNNs have fewer connections between nodes compared to fully connected NNs, require less memory, and the pooling allows for only the valuable features to be

passed to the next layer. Thus, they can be trained fairly quickly relative to fully connected NNs, and are regularized so that overfitting is reduced. Thus, CNNs also suit the current task, where only a small number of flies are available for model training and testing.

4. DATA

After eclosing, the fruit flies are housed in groups of 40 for 2 days, and then individuals are transferred to glass tubes that are 65 mm of length with 3 mm of inside diameter. The tube is placed in a monitor in which an infrared beam bisects the tube near the centre with food at one end and an air permeable barrier at the other. An activity count is registered when the fly breaks the beam [12]. The activity counts that are measured are the number of times the fly crossed the light beam. Each day, the flies are subject to 12 hours of light and dark periods to simulate the day-night cycle.

In this study, 1-minute activity counts until their death were collected from 531 *CS* male flies. The lifespan of each fly was also recorded, to an accuracy of 1 day. The minimum lifespan of the flies in the sample is 32 days, while the maximum is 74 days. The data set was divided into three groups based on days lived: (1) the short-lived group (labelled as Group 0), representing 151 flies with a maximum of 51 days lived; (2) the long-lived group (Group 2), consisting of 106 flies with a minimum of 60 days lived; and (3) the remaining falling into the middle group (Group 1). Note that an attempt was made to have Groups 0 and 2 be of equal size and represent approximately 25% of the flies in each category, but this could not be achieved due to ties at the Group 2 cut-off. A decision was made to exclude these ties from Group 2. To provide the CNN a good contrast between groups, the middle group was eliminated from our study. This left 257 flies in total for analyses and model building.

An actogram displaying 15 days' of activity counts for a selected fly is given in Figure 1. Each row shows activity counts of the fly in half-hour bins over the 24-hour day, beginning with the first day the fly's activities were recorded. For the convenience of

comparison, 2 days' worth of data are displayed in each row, say for day *i* and *i*+1, and the following row displays data for day *i*+1 and *i*+2. This is normally referred to as double plotting of the data. The actogram in Figure 1 clearly displays periods of high and low activity, with some periods showing the absence of any activity, indicating sleep. The main activity cycle coincides with the light/dark 12 hour periods flies are exposed to in order to maintain their circadian rhythm. Based on the inspection of the actograms for all flies, it can be observed that, in general, flies are more active when young and that their sleep patterns change as they age [38].

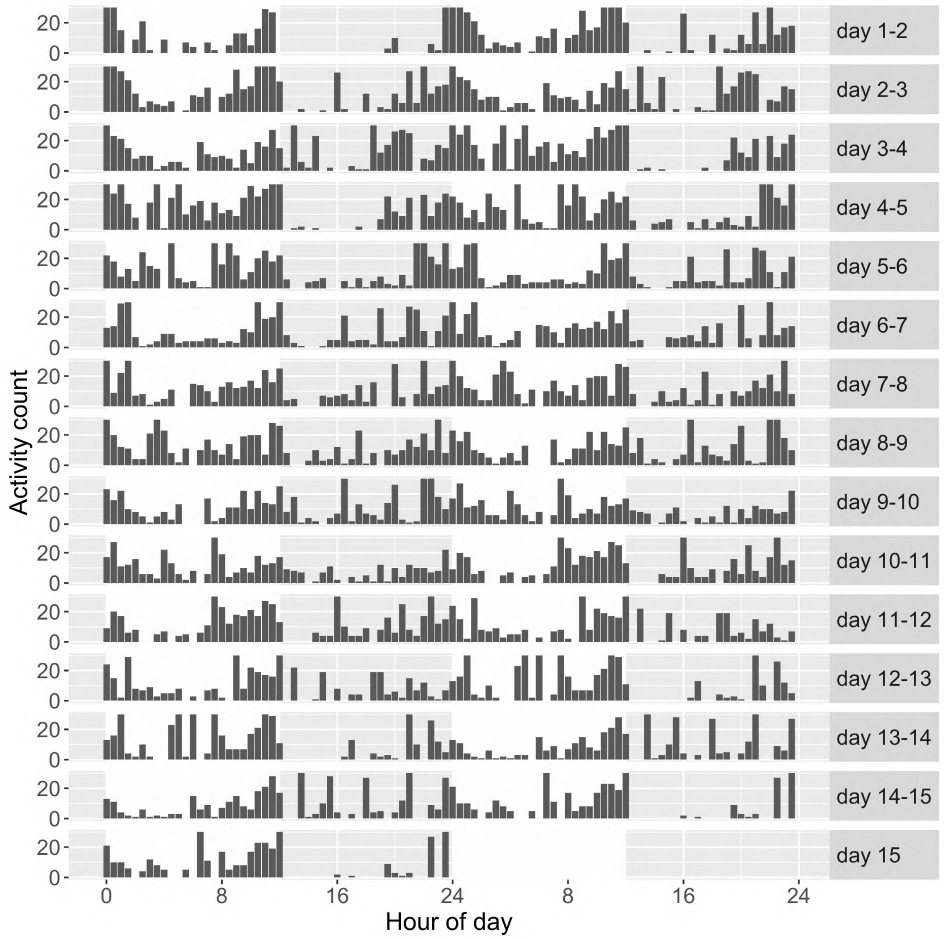


Figure 1. Actogram of activity data of an individual fly over fifteen days.

Note on data accessibility: The data used in this study are gathered for a much broader and ongoing investigation on the relationship between sleep and aging in *Drosophila*. As such, the data relevant to this study will not be available until the broader research project is completed. Readers are welcome to contact the last author who is leading the broader study with any specific questions about the data.

5. ANALYSIS AND RESULTS

The implementation of and the results from the two modelling stages, namely, fitting the VZI-ACP model, and the use of a CNN for classification are described in Sections 5.1 and 5.2, respectively.

5.1. RESULTS FROM FITTING THE VZI-ACP MODEL

In the VZI-ACP model, the zero-inflation probability was allowed to vary from hour to hour over the 24 hours of a day to reflect the circadian and other sleep cycles present in a fly's sleep architecture. A moving window of 5 days was used to model data from 5 days at a time, allowing the model parameters to vary over the course of the fly's life. The estimates of the 24 zero-inflation probabilities from days i through i+4 were recorded as belonging to day i, $i = 1, 2, \ldots, 24$. Thus, the moving window used is a right-sided one. Observe that whether the window is centred or not does not matter because the classifier is only influenced by the order in which the rows in the heat map are arranged rather than the label given to each row. Results from the zero-inflated ACP model fitting are illustrated in Figure 2, which provides a plot of these probabilities for two flies, one short-lived and one long-lived. Note that rather than illustrating these probabilities using step functions, the probabilities were plotted centred at each hour and then connected to illustrate the typical 'm' type shape of the sleep density exhibited by *CS* flies.

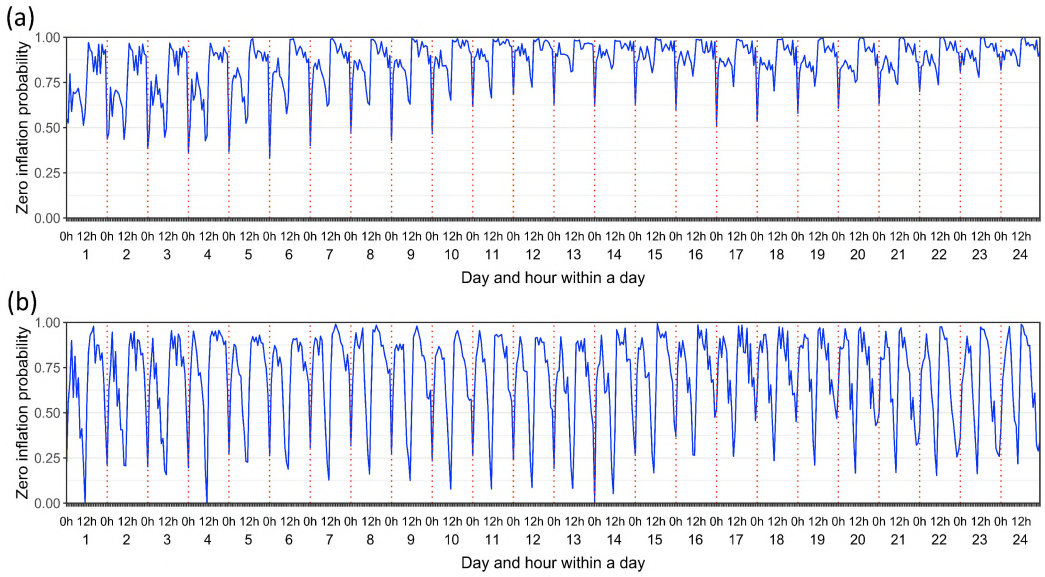


Figure 2. Estimated zero-inflation probabilities over 24 days. Figure (a) is for a fly that lived for 44 days and figure (b) is for a fly that lived 62 days.

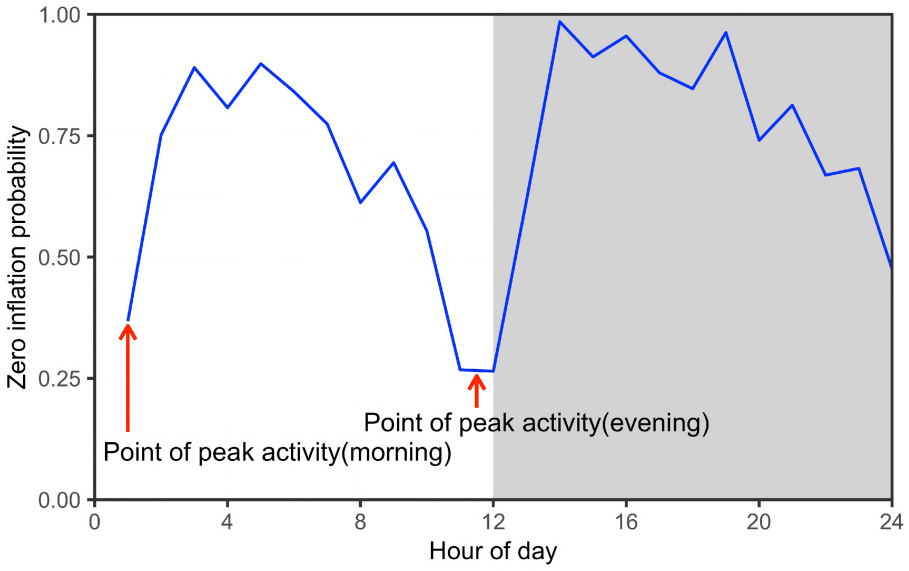


Figure 3. The graph of estimated zero-inflation probabilities over a day for an individual fly, with the 12-hour dark period shaded in grey.

It is apparent from Figure 2 that the 'm' shape evolves with a fly's aging process. These probabilities reflect the morning and evening points of peak activity (Figure 3) and suggest a change in the peaks as fly gets older. The underlying assumption in this study is that the zero-inflation probabilities capture information about how sleep patterns change with aging and they are hypothesized to contain features that will help categorize flies into short- and long-lived.

The resulting zero-inflation probabilities for each day, over a 24-day period (within the first 30 days of a fly's life) were utilized to create a heat map as described in Section 3, which contains information on the 24-hour sleep cycle for each day as well as how this profile changes across the 24-day observation period. The colours closer to red indicate high zero-inflation probabilities and those closer to purple signify low probabilities. Each row of the map represents the zero-inflation probability values for 1 day. Each column represents these probabilities for a specific hour across the 24 days over which the data are analysed. Examples of two such heat maps, one for a short-lived fly and the other for a long-lived fly, are displayed in Figure 4. The heat maps are used as input to the CNN to build the classification model.

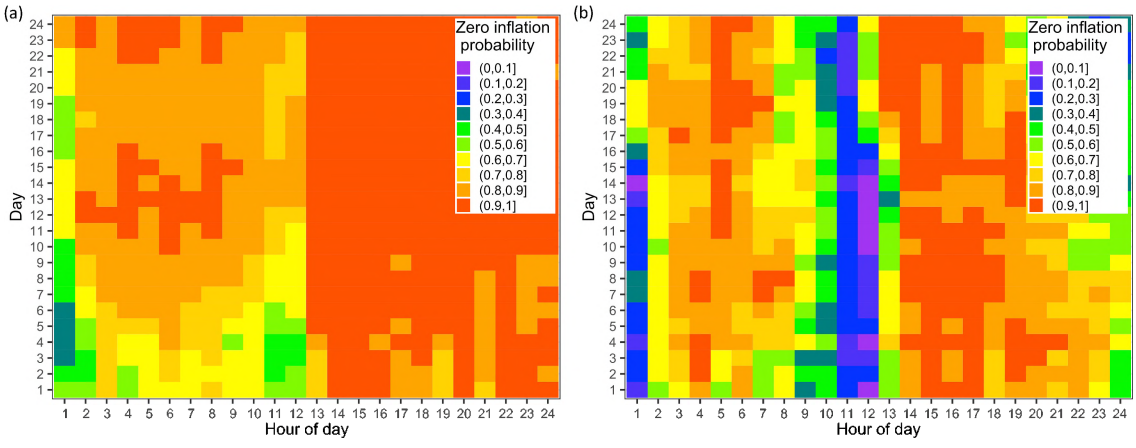


Figure 4. Heat map of zero-inflation probabilities for (a) short-lived and (b) long-lived flies.

5.2. RESULTS FROM THE CONVOLUTIONAL NEURAL NETWORK

The CNN used in this study consists of three layers. The first layer is a convolutional layer with 16 kernels of size of 3x3. This layer will get diverse information on different parts of the heat maps while decreasing noise from the original input. A pooling layer employed local pooling of 2x2 clusters, and max pooling was used in this second layer. It reduces the amount of information for each characteristic obtained from the convolutional layer but will keep the most important features and drop the others. This layer also helps to align key spatial information found in approximately the same area to a fixed location. Finally, a fully connected layer with the rectified linear unit (ReLU) activation function was selected to calculate the final probabilities for falling into each lifespan category. These hyperparameters were determined based on a few initial trials using a subset of the data. We tested convolutional kernels of size 2x2, 3x3, and 4x4 and determined that 3x3 is best. In the pooling layer, local pooling cluster sizes tested included 2x2, 3x3, and 4x4, with the smallest cluster size producing the best results. Both max pooling and average pooling were employed, but average pooling did not produce reasonable results. The activation functions that were tested are ReLU, Logistic, TanH, and Leaky ReLU, with ReLU outperforming the alternatives. The learning rate and decay rates were varied and we settled on 1E-4 for the learning rate and 1E-5 for the decay rate. Note that we did not include multiple hidden layers between the pooling layer and out put layer in order to avoid over-fitting. We plan to re-estimate the parameters, with additional layers included, using additional data when they become available through future experiments.

To compare the CNN performance with an alternative classifier, an ANN was trained on the same data. Since the CNN we employed had one inner layer, the ANN was restricted to one hidden layer. The number of hidden nodes in the inner layer was varied from 144 (which is comparable to the number of parameters in the CNN) to 288. Optimal results were obtained for an ANN with 216 hidden nodes, and the results are reported in Table 1.

Table 1. Confusion table based on test samples.

	Actual Group		Column Total
Predicted Group	0	1	
0	109	26	135 (80.74%)
1	42	80	122 (65.57%)
Row Total	151 (72.18%)	106 (75.47%)	257

Since the number of observations in the data set is limited, K-fold cross-validation was used instead of testing the estimated model employing a hold-out data set. Four disjoint subsets of data were created using stratified random sampling from the two lifespan groups (K=4). Three subsets were used as the training data and the remaining set as the test set. The process was repeated until all observations in each of the subsets were used in the test samples. The confusion table (Table 1) above provides the prediction results obtained by using probabilities greater than or equal to 0.5 to signify membership in a particular group. There are 151 short-lived flies and around 72% of these were predicted as short-lived by the CNN, while 80 out of 106 (around 75%) long-lived flies were predicted into the correct category. The prediction model yields reasonable accuracy, especially in light of the fact that sleep is not the only factor that affects the lifespans of the fruit flies. Results also show that the CNN outperformed the ANN across the board.

To provide a better view of how these two groups of flies are separated, a histogram of the difference in the probabilities of falling into Groups 0 and 2 are provided in Figure 5. Most short-lived flies have a positive difference, while most long-lived flies have a negative difference. This is expected as the flies belonging to the short-lived group have a higher probability of falling into Group 0 than falling into Group 2, resulting in a positive difference. On the contrary, long-lived files have a negative difference.

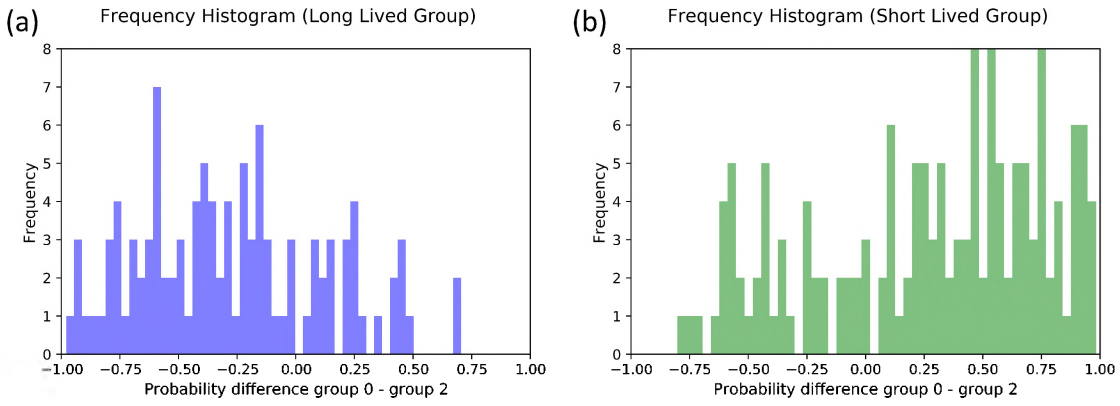


Figure 5. Frequency histogram of probability difference (probability of falling into Group 0 minus the probability of falling into Group 2).

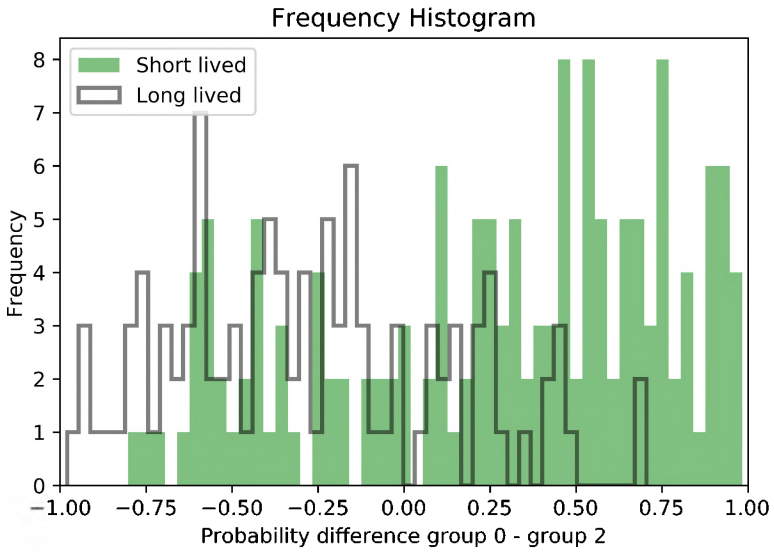


Figure 6. Overlap of the two histograms displayed in Figure 5.

Overlaying the two graphs (Figure 6) shows that there is a substantial overlap between short- and long-lived flies. This is expected because sleep is just one factor that plays a role in determining lifespan. If we do not set the probability for membership in a group at 0.5

or greater but set it at 0.75 or greater, there will be less mixing of these two groups. Table 2 shows the result of this updated separation for the CNN classifier. This, however, leaves a large number of flies without a classification. Results in Tables 1 and 2 show that the zero-inflation probabilities, which can be regarded as related to sleep, do have information regarding lifespan that can be utilized to build a classification model.

Table 2. Confusion table based on test samples after changing criteria (Unclassified samples not shown)

	Actual Group		Column Total
Predicted Group	0	1	
0	45	7	52 (86.54%)
1	11	24	35 (68.57%)
Row Total	56 (80.36%)	31 (77.42%)	87

6. DISCUSSION

Our study can be viewed as a proof of concept exercise, with the goal of determining if the per-minute activity data can be utilized to extract sleep related information and used to build a classification model to identify flies that fall into short- and long-lived categories. The results show reasonable success in this attempt. Results are promising, especially since sleep is not the only factor that influences lifespan. Studies on humans have identified sleep duration as a sleep-related factor that influences mortality [6, 7] , but preliminary studies by the authors and their research team have not found compelling evidence that this is the case in *Drosophila*. The primary reason for this may be that at the individual level there are compensatory parameters, such as consolidation, that may decrease the influence of absolute duration. Thus, the finding that the zero-inflation probabilities contain information linked to lifespan is a significant discovery. One major significance of our finding is the

generation of new hypotheses, testing of which can further our knowledge of the role sleep plays in health. For example, can we identify changes in the zero-inflation probabilities associated with a shorter or a longer lifespan? What do these changes signify with respect to restorative sleep? Are there biochemical markers that can be associated with such changes? In addition, our model will form the foundation for building even more accurate classifiers by incorporating other sleep characteristics. This will enable us to separate flies into short- and long-lived groups with even greater accuracy, with the predicted groups assayed to determine biochemical markers associated with biological aging. The results also show that image-based CNNs provide an appropriate platform for building a classification model for biological specimens where features useful for discrimination may vary slightly from one specimen to another.

These results, however, must be taken with a caveat. Our successful model was built using a CNN that was trained using only short- and long-lived flies. In addition, the training of the CNN was carried out using a relatively small data set. Once a sufficient number of observations become available to better tune the hyperparameters and train the CNN, we plan to train the CNN on three groups, short-, middle-, and long-lived. It is anticipated that future models will not only use the zero-inflation probability based heat maps, but images based on other statistics, such as the number of sleep-wake transitions. The addition of new input images when only a limited number of observations are available is problematic, and thus these advances will have to wait until a large number of additional experiments are conducted to incorporate more data.

7. CONCLUSIONS

A multi-stage modelling approach was employed to build a prediction model to identify fruit flies with short and long lifespans based on a sleep characteristic. Results show that the prediction model provides a reasonably accurate way to group flies into the two lifespan categories. In addition, the finding that the zero-inflation probabilities have

information related to lifespan is an important discovery. Further improvements to the model may be possible when considerably more observations become available, allowing additional model training as well as the inclusion of heat maps based on other statistics.

ACKNOWLEDGEMENTS

This research was solely supported by the National Institute of General Medical Sciences of the National Institute of Health under award number R15GM117507. We thank Josh Lisse, Elizabeth Park and Lauren Francis for their contributions in data collection and Luyang Wang for assistance with raw data formatting. We also like to thank Austin Vandegriffe for his assistance.

REFERENCES

[1] Daniel J Buysse. Sleep health: can we define it? does it matter? *Sleep*, 37(1):9–17, 2014.

[2] Kathryn J Reid, Zoran Martinovich, Sanford Finkel, Judy Statsinger, Robyn Golden, Kathryne Harter, and Phyllis C Zee. Sleep: a marker of physical and mental health in the elderly. *The American journal of geriatric psychiatry*, 14(10):860–866, 2006.

[3] Avi Sadeh, Amiram Raviv, and Reut Gruber. Sleep patterns and sleep disruptions in school-age children. *Developmental psychology*, 36(3):291, 2000.

[4] Daniel Bushey, Kimberly A Hughes, Giulio Tononi, and Chiara Cirelli. Sleep, aging, and lifespan in drosophila. *BMC neuroscience*, 11(1):1–18, 2010.

[5] Lisa Gallicchio and Bindu Kalesan. Sleep duration and mortality: a systematic review and meta-analysis. *Journal of sleep research*, 18(2):148–158, 2009.

[6] Francesco P Cappuccio, Lanfranco D'Elia, Pasquale Strazzullo, and Michelle A Miller. Sleep duration and all-cause mortality: a systematic review and meta-analysis of prospective studies. *Sleep*, 33(5):585–592, 2010.

[7] Michael A Grandner, Lauren Hale, Melisa Moore, and Nirav P Patel. Mortality associated with short sleep duration: the evidence, the possible mechanisms, and the future. *Sleep medicine reviews*, 14(3):191–203, 2010.

[8] Irwin Feinberg. Changes in sleep cycle patterns with age. *Journal of psychiatric research*, 10(3-4):283–306, 1974.

[9] Maurice M Ohayon, Mary A Carskadon, Christian Guilleminault, and Michael V Vitiello. Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: developing normative sleep values across the human lifespan. *Sleep*, 27(7):1255–1273, 2004.

[10] Joan C Hendricks, Stefanie M Finn, Karen A Panckeri, Jessica Chavkin, Julie A Williams, Amita Sehgal, and Allan I Pack. Rest in drosophila is a sleep-like state. *Neuron*, 25(1):129–138, 2000.

[11] Paul J Shaw, Chiara Cirelli, Ralph J Greenspan, and Giulio Tononi. Correlates of sleep and waking in drosophila melanogaster. *Science*, 287(5459):1834–1837, 2000.

[12] Rozi Andretic and Paul J Shaw. Essentials of sleep recordings in drosophila: moving beyond sleep time. *Methods in enzymology*, 393:759–772, 2005.

[13] Ravi Allada, Chiara Cirelli, and Amita Sehgal. Molecular mechanisms of sleep homeostasis in flies and mammals. *Cold Spring Harbor perspectives in biology*, 9(8):a027730, 2017.

[14] Matthew DW Piper and Linda Partridge. Drosophila as a model for ageing. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1864(9):2707–2717, 2018.

[15] Meredith L Wallace, Katie Stone, Stephen F Smagula, Martica H Hall, Burcin Simsek, Deborah M Kado, Susan Redline, Tien N Vo, Daniel J Buysse, and Osteoporotic Fractures in Men (MrOS) Study Research Group. Which sleep health characteristics predict all-cause mortality in older men? an application of flexible multivariable approaches. *Sleep*, 41(1):zsx189, 2018.

[16] Andréas Heinen. Modelling time series count data: an autoregressive conditional poisson model. *Available at SSRN 1117187*, 2003.

[17] M Ghahramani and A Thavaneswaran. On some properties of autoregressive conditional poisson (acp) models. *Economics Letters*, 105(3):273–275, 2009.

[18] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.

[19] Fukang Zhu. Zero-inflated poisson and negative binomial integer-valued garch models. *Journal of Statistical Planning and Inference*, 142(4):826–839, 2012.

[20] Samaranayake V.A. Ratnayake, I. A garch type poisson model for time series of counts with cyclically varying zero inflation. *In JSM Proceedings, Business and Economic Statistics Section, Alexandria, VA, American Statistical Association*, pages 1760–1770, 2017.

[21] Xiaofei Xu, Ying Chen, Cathy WS Chen, Xiancheng Lin, et al. Adaptive log-linear zero-inflated generalized poisson autoregressive model with applications to crime counts. *Annals of Applied Statistics*, 14(3):1493–1515, 2020.

[22] Steve Lawrence, C Lee Giles, Ah Chung Tsoi, and Andrew D Back. Face recognition: A convolutional neural-network approach. *IEEE transactions on neural networks*, 8(1):98–113, 1997.

[23] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012.

[24] Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei. Large-scale video classification with convolutional neural networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 1725–1732, 2014.

[25] Qing Li, Weidong Cai, Xiaogang Wang, Yun Zhou, David Dagan Feng, and Mei Chen. Medical image classification with convolutional neural network. In *2014 13th international conference on control automation robotics & vision (ICARCV)*, pages 844–848. IEEE, 2014.

[26] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565–571. IEEE, 2016.

[27] Nima Tajbakhsh, Jae Y Shin, Suryakanth R Gurudu, R Todd Hurst, Christopher B Kendall, Michael B Gotway, and Jianming Liang. Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE transactions on medical imaging*, 35(5):1299–1312, 2016.

[28] Anastasia Borovykh, Sander Bohte, and Cornelis W Oosterlee. Conditional time series forecasting with convolutional neural networks. *arXiv preprint arXiv:1703.04691*, 2017.

[29] Jou-Fan Chen, Wei-Lun Chen, Chun-Ping Huang, Szu-Hao Huang, and An-Pin Chen. Financial time-series data analysis using deep convolutional neural networks. In *2016 7th International conference on cloud computing and big data (CCBD)*, pages 87–92. IEEE, 2016.

[30] Yann LeCun. The mnist database of handwritten digits. *http://yann. lecun. com/exdb/mnist/*, 1998.

[31] Patrice Y Simard, David Steinkraus, John C Platt, et al. Best practices for convolutional neural networks applied to visual document analysis. In *Icdar*, volume 3. Citeseer, 2003.

[32] Edgar Oropesa, Hans L Cycon, and Marc Jobert. Sleep stage classification using wavelet transform and neural network. *International computer science institute*, 1999.

[33] Tomasz Wielek, Julia Lechinger, Malgorzata Wislowska, Christine Blume, Peter Ott, Stefan Wegenkittl, Renata Del Giudice, Dominik PJ Heib, Helmut A Mayer, Steven Laureys, et al. Sleep in patients with disorders of consciousness characterized by means of machine learning. *PloS one*, 13(1):e0190458, 2018.

[34] Arnaud Sors, Stéphane Bonnet, Sébastien Mirek, Laurent Vercueil, and Jean-François Payen. A convolutional neural network for sleep stage scoring from raw single-channel eeg. *Biomedical Signal Processing and Control*, 42:107–114, 2018.

[35] Orestis Tsinalis, Paul M Matthews, Yike Guo, and Stefanos Zafeiriou. Automatic sleep stage scoring with single-channel eeg using convolutional neural networks. *arXiv preprint arXiv:1610.01683*, 2016.

[36] Charu C Aggarwal et al. Neural networks and deep learning. *Springer*, 10:978–3, 2018.

[37] Mohit Sewak, Md Rezaul Karim, and Pradeep Pujari. *Practical convolutional neural networks: implement advanced deep learning models using Python*. Packt Publishing Ltd, 2018.

[38] Konstantin G Iliadi and Gabrielle L Boulianne. Age-related behavioral changes in drosophila. *Annals of the new York Academy of Sciences*, 1197(1):9–18, 2010.

SECTION

3. CONCLUSION

In this work, several variations of the autoregressive conditional Poisson (ACP) model were developed to allow for greater applicability in modeling real-life data. First, a periodic autoregressive conditional Poisson (PACP) model was proposed as a natural generalization of the autoregressive conditional Poisson model to explain the periodicity inherent in some count data series. By letting cyclical variations of the parameters in the model, it provided a way to explain such periodicity. The maximum likelihood estimation (MLE) method was utilized to estimate parameters of the model and a small-scale Monte Carlo simulation study was conducted with results showing good performance with respect to parameter estimation. The Akaike information criterion (AIC) value was calculated and used as a criterion to distinguish the true structure of the underlying data generating mechanism. The model was fitted to a real-life data set consisting of epidemiological disease counts and the periodic model was shown to have a smaller AIC compared to the non-periodic version, illustrating the necessity of such a generalization.

Second, an autoregressive conditional Poisson hidden Markov model (ACP-HMM) was developed to deal with count data time series whose dependence structure is governed by a hidden process. This model addressed the frequently observed serial correlation and the clustering of high or low counts inherent in many time series of count data, while at the same time allowing the underlying data generating mechanism to change according to a hidden Markov process. A small-scale Monte Carlo simulation study demonstrated that MLE method was a trustworthy way of estimating the model parameters. The AIC and Bayesian information criterion (BIC) were examined to determine their appropriateness for distinguishing the true structure of the model. Results show that both the AIC and the

BIC do a good job distinguishing the underlying model from an alternative. Application to daily number of deaths in Evora, Portugal, showed the model provided a better fit than the standard Poisson hidden Markov model.

Lastly, a modification of a zero-inflated Poisson model, whose zero-inflation parameters vary hourly within a day, was used to analyze activity counts of the fruit fly. The model captures the dynamic structure of activity patterns and the fly's propensity to sleep. The results were then fed to a convolutional neural network, and the life span categories (flies with short and long lifespans) were successfully identified by this multi-stage model.

REFERENCES

[1] Siddhartha Chib and Rainer Winkelmann. Markov chain monte carlo analysis of correlated count data. *Journal of Business & Economic Statistics*, 19(4):428–435, 2001.

[2] Iain L MacDonald and Walter Zucchini. *Hidden Markov and other models for discrete-valued time series*, volume 110. CRC Press, 1997.

[3] Tunny Sebastian, Visalakshi Jeyaseelan, Lakshmanan Jeyaseelan, Shalini Anandan, Sebastian George, and Shrikant I Bangdiwala. Decoding and modelling of time series count data using poisson hidden markov model and markov ordinal logistic regression models. *Statistical methods in medical research*, 28(5):1552–1563, 2019.

[4] Ben Cooper and Marc Lipsitch. The analysis of hospital infection data using hidden markov models. *Biostatistics*, 5(2):223–237, 2004.

[5] Ed McKenzie. Some simple models for discrete variate time series 1. *JAWRA Journal of the American Water Resources Association*, 21(4):645–650, 1985.

[6] Rong Zhu and Harry Joe. Modelling count data time series with markov processes based on binomial thinning. *Journal of Time Series Analysis*, 27(5):725–738, 2006.

[7] A Colin Cameron and Pravin K Trivedi. *Regression analysis of count data*, volume 53. Cambridge university press, 1998.

[8] Rainer Winkelmann and Klaus F Zimmermann. Recent developments in count data modelling: theory and application. *Journal of economic surveys*, 9(1):1–24, 1995.

[9] Stefany Coxe, Stephen G West, and Leona S Aiken. The analysis of count data: A gentle introduction to poisson regression and its alternatives. *Journal of personality assessment*, 91(2):121–136, 2009.

[10] Tracy DeHart, Howard Tennen, Stephen Armeli, Michael Todd, and Glenn Affleck. Drinking to regulate negative romantic relationship interactions: The moderating role of self-esteem. *Journal of Experimental Social Psychology*, 44(3):527–538, 2008.

[11] Andreas Lindén and Samu Mäntyniemi. Using the negative binomial distribution to model overdispersion in ecological count data. *Ecology*, 92(7):1414–1421, 2011.

[12] Paul S Albert. A two-state markov mixture model for a time series of epileptic seizure counts. *Biometrics*, pages 1371–1381, 1991.

[13] Patricia A Jacobs and Peter AW Lewis. Discrete time series generated by mixtures. i: Correlational and runs properties. *Journal of the Royal Statistical Society: Series B (Methodological)*, 40(1):94–105, 1978.

[14] Patricia A Jacobs and Peter AW Lewis. Stationary discrete autoregressive-moving average time series generated by mixtures. *Journal of Time Series Analysis*, 4(1):19–36, 1983.

[15] Andréas Heinen. Modelling time series count data: an autoregressive conditional poisson model. *Available at SSRN 1117187*, 2003.

[16] René Ferland, Alain Latour, and Driss Oraichi. Integer-valued garch process. *Journal of Time Series Analysis*, 27(6):923–942, 2006.

[17] Fukang Zhu. A negative binomial integer-valued garch model. *Journal of Time Series Analysis*, 32(1):54–67, 2011.

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

[19] Ming Yang, Gideon KD Zamba, and Joseph E Cavanaugh. Markov regression models for count time series with excess zeros: A partial likelihood approach. *Statistical Methodology*, 14:26–38, 2013.

[20] Tobias A Möller, Christian H Weiß, Hee-Young Kim, and Andrei Sirchenko. Modeling zero inflation in count data time series with bounded support. *Methodology and Computing in Applied Probability*, 20(2):589–609, 2018.

[21] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.

[22] Tim Bollerslev and Eric Ghysels. Periodic autoregressive conditional heteroscedasticity. *Journal of Business & Economic Statistics*, 14(2):139–151, 1996.

[23] Anthony Hopkins, Philip Davies, and Charles Dobson. Mathematical models of patterns of seizures: their use in the evaluation of drugs. *Archives of neurology*, 42(5):463–467, 1985.

[24] M Ghahramani and A Thavaneswaran. On some properties of autoregressive conditional poisson (acp) models. *Economics Letters*, 105(3):273–275, 2009.

[25] Fukang Zhu. Zero-inflated poisson and negative binomial integer-valued garch models. *Journal of Statistical Planning and Inference*, 142(4):826–839, 2012.

[26] PoC Consul and Felix Famoye. Generalized poisson regression model. *Communications in Statistics-Theory and Methods*, 21(1):89–109, 1992.

[27] Joseph M Hilbe. *Modeling count data*. Cambridge University Press, 2014.

[28] John Hinde. Compound poisson regression models. In *GLIM 82: Proceedings of the International Conference on Generalised Linear Models*, pages 109–121. Springer, 1982.

[29] Chester Ittner Bliss and Ronald A Fisher. Fitting the negative binomial distribution to biological data. *Biometrics*, 9(2):176–200, 1953.

[30] Scott L Zeger. A regression model for time series of counts. *Biometrika*, 75(4):621–629, 1988.

[31] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[32] Patricia A Jacobs and Peter AW Lewis. Stationary discrete autoregressive-moving average time series generated by mixtures. *Journal of Time Series Analysis*, 4(1):19–36, 1983.

[33] Eddie McKenzie. Discrete variate time series. *Handbook of statistics*, 21(573606):1, 2003.

[34] Fukang Zhu. Modeling overdispersed or underdispersed count data with generalized poisson integer-valued garch models. *Journal of Mathematical Analysis and Applications*, 389(1):58–71, 2012.

[35] Cathy WS Chen, Mike KP So, Jessica C Li, and Songsak Sriboonchitta. Autoregressive conditional negative binomial model applied to over-dispersed time series of counts. *Statistical Methodology*, 31:73–90, 2016.

[36] Adrian E Raftery. A model for high-order markov chains. *Journal of the Royal Statistical Society: Series B (Methodological)*, 47(3):528–539, 1985.

[37] Robert C Jung and Andrew R Tremayne. Coherent forecasting in integer time series models. *International Journal of Forecasting*, 22(2):223–238, 2006.

[38] Leonard E Baum and Ted Petrie. Statistical inference for probabilistic functions of finite state markov chains. *The annals of mathematical statistics*, 37(6):1554–1563, 1966.

[39] Leonard E Baum and John Alonzo Eagon. An inequality with applications to statistical estimation for probabilistic functions of markov processes and to a model for ecology. *Bulletin of the American Mathematical Society*, 73(3):360–363, 1967.

[40] Biing Hwang Juang and Laurence R Rabiner. Hidden markov models for speech recognition. *Technometrics*, 33(3):251–272, 1991.

[41] Stacia M DeSantis and Dipankar Bandyopadhyay. Hidden markov models for zero-inflated poisson counts with an application to substance use. *Statistics in medicine*, 30(14):1678–1694, 2011.

[42] Daniel J Buysse. Sleep health: can we define it? does it matter? *Sleep*, 37(1):9–17, 2014.

[43] Kathryn J Reid, Zoran Martinovich, Sanford Finkel, Judy Statsinger, Robyn Golden, Kathryne Harter, and Phyllis C Zee. Sleep: a marker of physical and mental health in the elderly. *The American journal of geriatric psychiatry*, 14(10):860–866, 2006.

[44] Avi Sadeh, Amiram Raviv, and Reut Gruber. Sleep patterns and sleep disruptions in school-age children. *Developmental psychology*, 36(3):291, 2000.

[45] Daniel Bushey, Kimberly A Hughes, Giulio Tononi, and Chiara Cirelli. Sleep, aging, and lifespan in drosophila. *BMC neuroscience*, 11(1):1–18, 2010.

[46] Lisa Gallicchio and Bindu Kalesan. Sleep duration and mortality: a systematic review and meta-analysis. *Journal of sleep research*, 18(2):148–158, 2009.

[47] Francesco P Cappuccio, Lanfranco D'Elia, Pasquale Strazzullo, and Michelle A Miller. Sleep duration and all-cause mortality: a systematic review and meta-analysis of prospective studies. *Sleep*, 33(5):585–592, 2010.

[48] Michael A Grandner, Lauren Hale, Melisa Moore, and Nirav P Patel. Mortality associated with short sleep duration: the evidence, the possible mechanisms, and the future. *Sleep medicine reviews*, 14(3):191–203, 2010.

[49] Irwin Feinberg. Changes in sleep cycle patterns with age. *Journal of psychiatric research*, 10(3-4):283–306, 1974.

[50] Maurice M Ohayon, Mary A Carskadon, Christian Guilleminault, and Michael V Vitiello. Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: developing normative sleep values across the human lifespan. *Sleep*, 27(7):1255–1273, 2004.

[51] Joan C Hendricks, Stefanie M Finn, Karen A Panckeri, Jessica Chavkin, Julie A Williams, Amita Sehgal, and Allan I Pack. Rest in drosophila is a sleep-like state. *Neuron*, 25(1):129–138, 2000.

[52] Paul J Shaw, Chiara Cirelli, Ralph J Greenspan, and Giulio Tononi. Correlates of sleep and waking in drosophila melanogaster. *Science*, 287(5459):1834–1837, 2000.

[53] Rozi Andretic and Paul J Shaw. Essentials of sleep recordings in drosophila: moving beyond sleep time. *Methods in enzymology*, 393:759–772, 2005.

[54] Ravi Allada, Chiara Cirelli, and Amita Sehgal. Molecular mechanisms of sleep homeostasis in flies and mammals. *Cold Spring Harbor perspectives in biology*, 9(8):a027730, 2017.

[55] Matthew DW Piper and Linda Partridge. Drosophila as a model for ageing. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1864(9):2707–2717, 2018.

[56] Meredith L Wallace, Katie Stone, Stephen F Smagula, Martica H Hall, Burcin Simsek, Deborah M Kado, Susan Redline, Tien N Vo, Daniel J Buysse, and Osteoporotic Fractures in Men (MrOS) Study Research Group. Which sleep health characteristics predict all-cause mortality in older men? an application of flexible multivariable approaches. *Sleep*, 41(1):zsx189, 2018.

[57] Samaranayake V.A. Ratnayake, I. A garch type poisson model for time series of counts with cyclically varying zero inflation. *In JSM Proceedings, Business and Economic Statistics Section, Alexandria, VA, American Statistical Association*, pages 1760–1770, 2017.

[58] Xiaofei Xu, Ying Chen, Cathy WS Chen, Xiancheng Lin, et al. Adaptive log-linear zero-inflated generalized poisson autoregressive model with applications to crime counts. *Annals of Applied Statistics*, 14(3):1493–1515, 2020.

[59] Steve Lawrence, C Lee Giles, Ah Chung Tsoi, and Andrew D Back. Face recognition: A convolutional neural-network approach. *IEEE transactions on neural networks*, 8(1):98–113, 1997.

[60] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012.

[61] Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei. Large-scale video classification with convolutional neural networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 1725–1732, 2014.

[62] Qing Li, Weidong Cai, Xiaogang Wang, Yun Zhou, David Dagan Feng, and Mei Chen. Medical image classification with convolutional neural network. In *2014 13th international conference on control automation robotics & vision (ICARCV)*, pages 844–848. IEEE, 2014.

[63] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565–571. IEEE, 2016.

[64] Nima Tajbakhsh, Jae Y Shin, Suryakanth R Gurudu, R Todd Hurst, Christopher B Kendall, Michael B Gotway, and Jianming Liang. Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE transactions on medical imaging*, 35(5):1299–1312, 2016.

[65] Anastasia Borovykh, Sander Bohte, and Cornelis W Oosterlee. Conditional time series forecasting with convolutional neural networks. *arXiv preprint arXiv:1703.04691*, 2017.

[66] Jou-Fan Chen, Wei-Lun Chen, Chun-Ping Huang, Szu-Hao Huang, and An-Pin Chen. Financial time-series data analysis using deep convolutional neural networks. In *2016 7th International conference on cloud computing and big data (CCBD)*, pages 87–92. IEEE, 2016.

[67] Yann LeCun. The mnist database of handwritten digits. *http://yann. lecun. com/exdb/mnist/*, 1998.

[68] Patrice Y Simard, David Steinkraus, John C Platt, et al. Best practices for convolutional neural networks applied to visual document analysis. In *Icdar*, volume 3. Citeseer, 2003.

[69] Edgar Oropesa, Hans L Cycon, and Marc Jobert. Sleep stage classification using wavelet transform and neural network. *International computer science institute*, 1999.

[70] Tomasz Wielek, Julia Lechinger, Malgorzata Wislowska, Christine Blume, Peter Ott, Stefan Wegenkittl, Renata Del Giudice, Dominik PJ Heib, Helmut A Mayer, Steven Laureys, et al. Sleep in patients with disorders of consciousness characterized by means of machine learning. *PloS one*, 13(1):e0190458, 2018.

[71] Arnaud Sors, Stéphane Bonnet, Sébastien Mirek, Laurent Vercueil, and Jean-François Payen. A convolutional neural network for sleep stage scoring from raw single-channel eeg. *Biomedical Signal Processing and Control*, 42:107–114, 2018.

[72] Orestis Tsinalis, Paul M Matthews, Yike Guo, and Stefanos Zafeiriou. Automatic sleep stage scoring with single-channel eeg using convolutional neural networks. *arXiv preprint arXiv:1610.01683*, 2016.

[73] Charu C Aggarwal et al. Neural networks and deep learning. *Springer*, 10:978–3, 2018.

[74] Mohit Sewak, Md Rezaul Karim, and Pradeep Pujari. *Practical convolutional neural networks: implement advanced deep learning models using Python*. Packt Publishing Ltd, 2018.

[75] Konstantin G Iliadi and Gabrielle L Boulianne. Age-related behavioral changes in drosophila. *Annals of the new York Academy of Sciences*, 1197(1):9–18, 2010.

VITA

Yi Zhang was born in Qingdao, a beautiful coastline city in China. She received her bachelor's degree in Bio-science from Ocean University of China. After she completed her degree, she worked as a graduate researcher at Ocean University of China to develop new medicine for cancer treatment. There she gained experience in bio-data collection and data analysis.

She obtained her Doctor of Philosophy in Mathematics with Statistics Emphasis from Missouri University of Science and Technology in July 2021, under the supervision of Dr. V. A. Samaranayake. Her research focused mainly on developing and using time series models for count data processes. She also applied machine learning techniques to extract biologically relevant information from experimental data and establish relationships between bio-information and life characteristics of model organisms. This latter research was funded by NIH. She published three papers in conference proceedings and in a journal.

Yi Zhang was good at playing piano and her hobbies also included badminton and chess.