New Mixture Models for Discrete Counts Time Series: with an Application to Modelling Mortality and Climate in NZ

I. L. Hudson[1], **M. L. Dalrymple**[2] **and M. J. Faddy**[3]

[1] School of Mathematics and Statistics, University of South Australia, Adelaide, Australia
[2]Department of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand
[3] School of Mathematical Sciences, Queensland University of Technology, Brisbane, Queensland, Australia
E-mail: irene.hudson@unisa.edu.au

Keywords: *Discrete counts time series, mixtures, Poisson distribution, SIDS, climate*

EXTENDED ABSTRACT

Mixture models have experienced a huge resurgence of interest and appeal over the last two decades (Böhning & Seidel, 2003). To date the development of mixture based models, for time series data, has mostly concentrated on modelling continuous data as mixtures of normals or as mixtures of ARMA and ARCH models (Bollerslev et al., 1992). Little work has been done, however, which adapts mixture methodology to a discrete time series context (Dalrymple, 2004). In the present paper five related regression models, based on mixtures of distributions and Poisson processes, are presented with adaptations to account for discrete counts time series data. A preliminary paper gave some suggestions for three of the five models in the context of a discrete series of SIDS counts and a subset of climatic predictors (Dalrymple et al., 2003). Mooney et al. (2003), in the only paper to date using mixtures to discrete counts time series, fitted mixtures of von Mises distributions to a case study of UK SIDS rates (1983–1998), with monthly rates analysed separately for each year and covariates not explicitly incorporated.

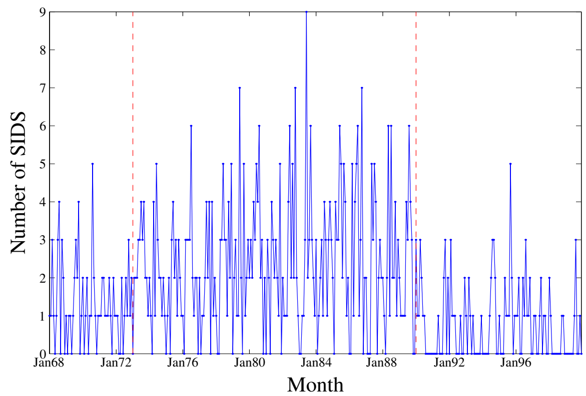

Figure 1. SIDS counts per month, Canterbury, NZ, 1968–2000.

Only recently have mixture-based models been extended to accommodate serial correlation in longitudinal studies. Booth et al. (2003) extended the negative binomial model to the case of dependent (repeated measures) counts. Min & Agresti (2005) recently developed a random effects model for repeated measures of zero-inflated count data. Dobbie & Welsh (2001) extended the hurdle model to incorporate serial correlation arising from repeated measures count data via construction of generalised estimating equations for each model component; Toscas & Faddy (2003) used transition methods to model correlated longitudinal data using the extended Poisson process (EPP) approach of Faddy (1997a). Note also a recent application of EPP models to study the dependence between the number of trials and success probabilities in binary trials by Faddy & Smith (2005).

The five models discussed here are the negative binomial (NB), finite mixture (FM), zero-inflated Poisson (ZIP), hurdle and Extended Poisson Process (EPP), based on mixtures of distributions, gamma, multinomial and binomial, and Poisson processes; developed so as to take into account possible serial dependence in time series data. This is achieved, in part, by the inclusion of an autoregressive based variable. All models accommodate for potential overdispersion and the EPP model allows also for underdispersion. The mixture models are illustrated with analyses of the effect of climate on sudden infant deaths (SIDS) rates, based on a unique total ascertainment study of SIDS incidence in Canterbury, NZ for the period 1968-2000 (Figure 1) (Dalrymple, 2004). This is the first study to examine dewpoint, wind direction and wind chill as possible SIDS risk factors.

For the SIDS data the ZIP and FM models performed best. The analysis suggested that the incidence of SIDS was associated with humidity and wind speed. This is not surprising, as variations in these factors may modify parental care practices. The methods presented in this paper are applicable to any study examining the relationship between discrete counts time series and multiple time series profiles of predictors.

1 INTRODUCTION

SIDS is still the most predominant cause of death in infants under one year of age, and currently accounts for between 10-20% of all infant deaths in developed countries (Malloy & Freeman, 2000). Epidemiological research has, for a long time, looked at climatic variables as contributors to SIDS risk. Many studies have related SIDS to various meteorological measures, yet the only consistently found relationship is between SIDS and seasonality, with more SIDS deaths occurring in winter (Dalrymple, 2004). Even though the SIDS-climate relationship has been examined extensively in terms of seasonal effects and also long/short term ambient temperature, there have only been a small number of studies relating SIDS to meteorological variables other than temperature (Macey et al., 2000). See also Dalrymple (2004).

Recently Dalrymple et al. (2003) used three mixture models (finite mixture, zero-inflated Poisson and the hurdle) to examine the relationship between SIDS incidence in Canterbury, NZ, with humidity and temperature for a seventeen year period (1973–1989). There have however, only been two other studies to date, that have examined SIDS incidence in relation to a more comprehensive set of meteorological variables, namely that of Auliciems & Barnes (1987) and Knöbel et al. (1995). This present study aims to undertake a most comprehensive analysis of SIDS incidence in relation to multiple meteorological measures in Canterbury, NZ. Unlike Auliciems & Barnes (1987), who examine climatic conditions centred around the day of death, this present study applies Poisson mixture methods to examine the monthly profile of Canterbury SIDS data to determine a profile of at-risk SIDS months. This study also investigates a more comprehensive set of climatic variables than studied to date, examining eleven climatic variables (with corresponding variable names) as follows; Temperature (°C) (Temp); Wind Direction (WindD); Wind Speed (knots) (WindS); Wind Velocity (knots) (WindV); Wind Chill (°C) (WindC); Humidity (%) (Humid); Pressure (Hpa) (Pres); Rainfall (mm/hr) (Rain); Sunshine (hrs) (Sun); Solar Radiation (Mj/m^2) (Rad); and Dewpoint (°C) (Dew). Lagged variants of the latter were also investigated. No study to date has included dewpoint, humidity, wind direction nor wind chill as possible SIDS risk factors (Dalrymple, 2004).

This current paper formulates five Poisson mixture (observation driven) models which incorporate a state-dependent structure. An overview of the negative binomial, finite mixture, zero-inflated Poisson, hurdle and extended Poisson process models is given. The advantages of the class of mixture-based, generalized Poisson models, suggested here, include, their accommodation of potential serial dependence in discrete count(s) time series data, the provision of additional insights into underlying dependencies relating to the outcome variable of interest, and allowance for possible overdispersion or underdispersion in contrast to the Poisson distribution. The five models are applied to an investigation of possible climatic impacts on a unique series of sudden infant death syndrome counts.

2 POISSON MIXTURE MODELS

The five formulations, presented in this paper are generalisations of the Poisson distribution. A summary of these models is presented in Table 1, which highlights the mixing formulation, distribution and peculiarities of each model. Table 2 presents details of the theoretical aspects of each model including the probability density function (pdf), mean, variance and log-likelihood, alongside the computational estimation procedures implemented. A summary of the notation, rationale for and comparison of models follows.

Notation for the model formulations in the subsequent section is defined as: Let y_i denote the response variable (the number of SIDS in month i), $i = 1, \ldots, n$; $\mathbf{y}$ is then the vector of (monthly) SIDS counts. Let $\mathbf{X}$ denote the matrix of (monthly) covariates with $\mathbf{x}_i = [1, x_{1i}, \ldots, x_{gi}]'$, with a total of g covariates. Also let $\mathbf{Z}$ denote the matrix of (monthly) covariates such that $\mathbf{z}_i = [1, z_{1i}, \ldots, z_{gi}]'$ giving a total of g covariates. The covariates in $\mathbf{X}$ and $\mathbf{Z}$ are associated with the non-zero and zero-SIDS components respectively and are not restricted to being the same.

A common alternative to the Poisson log-linear model, when the requirement of equality of the conditional mean and variance is unrealistic, is the negative binomial (NB) model. The NB model allows for overdispersion through flexible modelling of the variance. The formulation of the NB model as a mixture of Poisson and gamma distributions can be found in Greene (2000) (See Tables 1-2.).

Finite mixture (FM) models have been extensively discussed in the literature and applied to a wide range of data (McLachlan & Peel, 2000). The FM formulation as given in Table 2 follows that of Wang et al. (1998). This FM model assumes an underlying partition of the population into k homogeneous components. Each component having a different SIDS risk level $\lambda_j(\mathbf{x}_i)$, dependent on possibly different covariates. The FM model is a general model, which allows mixing with respect to both zeros and positives, with $\mathbf{Z}$ representing the covariates in the mixing proportions, and $\mathbf{X}$ the covariates in the Poisson rates (Table 2). In this study

Model	Mixing formulation	Distribution	Particular cases for certain parameter values
Negative binomial	Poisson + Gamma	$y_i \sim$ negative binomial	• reduces to Poisson
Finite mixture	Poisson + Multinomial	$y_i \sim$ Poisson(λ_j) with probability p_j $j = 1, \ldots, k$, where $k =$ number of components	• reduces to ZIP • reduces to Poisson
Zero-inflated Poisson	Poisson + Binomial	$y_i = 0$, with prob $(1 - p)$ $y_i \sim$ Poisson(λ), with prob p	• reduces to Poisson
Hurdle	Truncated Poisson + Binomial	$y_i = 0$, with prob $(1 - p)$ $y_i \sim$ truncated Poisson, with prob p	
Extended Poisson Process	Poisson process	$Y(t) =$ prob distn of SIDS counts	• reduces to Poisson • reduces to NB

Table 1. Overview of mixture models.

estimation of model parameters for the FM models was achieved by an adaptation of Wang's algorithm (Wang et al., 1998).

For the ZIP model the response variable is modelled as a mixture of a Bernoulli distribution and a Poisson distribution (Table 2). Only relatively recently did Lambert (1992) provide the general formulation for ZIP regression models incorporating covariates.

The hurdle specification for a truncated-at-zero Poisson distribution as presented in Table 2 follows the formulation of Welsh et al. (1996) and is a two component mixture approach to modelling count data, via a bivariate model. This hurdle model formulation increases the probability of the zero outcomes and scales the remaining probabilities to add to one.

The EPP model given here provides a time series extension of the EPP model of Faddy (1997a) developed recently by Dalrymple (2004). Other applications of the EPP model are due to Faddy (1997b). Note that recently Podlich et al. (2004) derived semi-parametric EPP models for (non-time series) counts, where the transition rates depend nonparametrically on the number of events, but parametrically on covariates The full EPP *time series* formulation is given in detail in Dalrymple (2004).

3 MODEL ADAPTATIONS

Testing for a significant serial dependence is typically a first step in discrete time series analysis. In the SIDS counts application, this was performed using the simple runs test for serial dependence, as discussed by Jung & Tremayne (2003). In the presence of a significant correlation, however, the following modification to model formulation and testing is implemented. Autoregressive covariates, (y_{i-k}) to lag k^*, $k = 1, 2, 3, \ldots, k^*$ based on past observations, are added to the model formulations to describe a possible state dependence. The autocorrelation function of the residuals of the model (without the AR regressor) is examined and the appropriate AR structure determined sequentially. The additional lag term(s) (up to order k^*) are then incorporated into the model, giving an exponential conditional mean of the following form:

$$E[y_i|\mathbf{x}_i, y_{i-1}, y_{i-2}, \ldots, y_{i-k}] = \exp(\mathbf{x}_i\boldsymbol{\beta} + \boldsymbol{\rho}\mathbf{y}_{i-k}), \quad (1)$$

where $\mathbf{y}_{i-k} = [y_{i-1}, y_{i-2}, \ldots, y_{i-k}]'$.

The conditional means and variances in Table 2 are then easily modified, via equation 1, to incorporate the addition of lagged dependent regressors. For example, in the NB case, the exponential conditional mean is

$E[y_i|\mathbf{y}_{i-k^*}, \mathbf{x}_i] = \lambda(\mathbf{y}_{i-k^*}, \mathbf{x}_i) = \exp(\boldsymbol{\beta}\mathbf{x}_i + \boldsymbol{\rho}\mathbf{y}_{i-k^*})$ and variance $Var[y_i|\mathbf{y}_{i-k^*}, \mathbf{x}_i] = \lambda(\mathbf{y}_{i-k^*}, \mathbf{x}_i) + \frac{\lambda(\mathbf{y}_{i-k^*}, \mathbf{x}_i)^2}{\theta} = \exp(\boldsymbol{\beta}\mathbf{x}_i + \boldsymbol{\rho}\mathbf{y}_{i-k^*}) + \frac{\exp(\boldsymbol{\beta}\mathbf{x}_i + \boldsymbol{\rho}\mathbf{y}_{i-k^*})^2}{\theta}$ $(\theta > 0)$ (See Dalrymple (2004)).

Testing for the order k and the influence of higher order lags, up to k^* was achieved following an extension based on Raftery's mixture transition distribution model. See Raftery & Tavare (1994), where the conditional distribution of $P(y_i|y_{i-1}, y_{i-2}, \ldots, y_{i-G})$ for some $G > 1$ can be approximated by a weighted combination of G conditional distributions,

$$P(y_i|y_{i-1}, y_{i-2}, \ldots, y_{i-G}) = \sum_{g=1}^{G} \pi_g P(y_i|y_{i-g}) \quad (2)$$

where $\sum_{g=1}^{G} \pi_g = 1$ and $\pi_g \geqslant 0$, $g = 1, \ldots, G$.

4 APPLICATION: SIDS AND CLIMATE

4.1 Diagnostic Accuracy & Change points

The SIDS counts derive from a unique retrospective, complete ascertainment study of SIDS incidence in Canterbury, which was completed by the Christchurch Community Paediatric Unit and provided daily SIDS counts (1968 to 2000).

Table 2. Theoretical overview of aspects of each of the five mixture model formulations.

	Negative Binomial (NB)	Finite Mixture (FM)	Extended Poisson Process (EPP)
Distribution	$y_i^m \sim NB(\theta, \lambda(\mathbf{x}_i^m))$	$y_i^m \sim \text{Poisson}(\lambda_j(\mathbf{x}_i^m))$ with probability $_j(\mathbf{z}_i^m)$	
PDF	$P(y_i^m = q \mid \mathbf{x}_i^m) = \frac{\Gamma(\theta+q)}{\Gamma(q+1)\Gamma(\theta)} r_i^q (1-r_i)^\theta$ $r_i = \frac{\lambda(\mathbf{x}_i^m)}{\lambda(\mathbf{x}_i^m)+\theta} \qquad q = 0, 1, 2, \ldots$	$P(y_i^m = q \mid \mathbf{x}_i^m, \mathbf{z}_i^m) =$ $\sum_{j=1}^k \frac{p_j(\mathbf{z}_i^m)\exp(-\lambda_j(\mathbf{x}_i^m))\lambda_j(\mathbf{x}_i^m)^q}{q!}$ $q = 0, 1, 2, \ldots$	
Link	log-linear	$p_j(\mathbf{z}_i)$ logit $(j = 1, \ldots, k)$ $\lambda_j(\mathbf{x}_i)$ log-linear	$\lambda_i = a(i+b)^c$ $a > 0, b > 0, c \leqslant 1$
Mean	$E[y_i^m \mid \mathbf{x}_i^m] = \lambda(\mathbf{x}_i^m)$	$E[y_i^m \mid \mathbf{x}_i^m, \mathbf{z}_i^m] = \sum_{j=1}^k p_j(\mathbf{z}_i^m)\lambda_j(\mathbf{x}_i^m)$	$E[Y(1) \mid \mathbf{X}^m] = b\left[(1+r)^{\frac{1}{1-c}} - 1\right]$
Variance	$Var[y_i^m \mid \mathbf{x}_i^m] = \lambda(\mathbf{x}_i^m) + \frac{\lambda(\mathbf{x}_i^m)^2}{\theta}$ $(\theta > 0)$	$Var[y_i^m \mid \mathbf{x}_i^m, \mathbf{z}_i^m] = E[y_i^m \mid \mathbf{x}_i^m, \mathbf{z}_i^m]\Big[1 + \lambda_j(\mathbf{x}_i^m)$ $- E[y_i^m \mid \mathbf{x}_i^m, \mathbf{z}_i^m]\Big]$	$Var[Y(1) \mid \mathbf{X}^m]$ $= \frac{b}{1-2c}(1+r)^{\frac{1}{1-c}}$ $\times\Big(1 - (1+r)^{\frac{2c-1}{1-c}}\Big)$ $r = \frac{a(1-c)}{b^{1-c}} = \frac{\exp(\mathbf{X}\beta)(1-c)}{b^{1-c}}$
LL	$L_{NB}(\theta, \boldsymbol{\beta}) = \sum_{i=1}^n \Bigg(\big[\sum_{j=0}^{y_i^m-1}\log(j-\theta)\big] - \log y_i^m!$ $-(y_i^m+\theta)\log(1+\frac{1}{\theta}\exp(\mathbf{x}_i^m\boldsymbol{\beta})) + y_i\log\theta + y_i\mathbf{x}_i^m\boldsymbol{\beta}\Bigg)$	$L_{FM}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{i=1}^n \sum_{j=1}^k \mathcal{Z}_{ji}\log[p_j(\mathbf{z}_i^m)\times$ $\text{Poisson}(\lambda_j(\mathbf{x}_i^m))]$	
Estimation Procedure	Maximum likelihood (ML) SAS GENMOD procedure	Wang (1998) Combination of EM and quasi-Newton	
	Zero-inflated Poisson (ZIP)	**Hurdle (H)**	
Distribution	$y_i^m = 0$ prob $1 - p(\mathbf{z}_i^m)$ $y_i^m \sim \text{Poisson}(\lambda(\mathbf{x}_i^m))$ prob $p(\mathbf{z}_i^m)$	$y_i^m = 0$ prob $1 - p(\mathbf{z}_i^m)$ $y_i^m \sim$ truncated $\text{Poisson}(\lambda(\mathbf{x}_i^m))$ prob $p(\mathbf{z}_i^m)$	
PDF	$P(y_i^m = 0 \mid \mathbf{x}_i^m, \mathbf{z}_i^m) = 1 - p(\mathbf{z}_i^m) + p(\mathbf{z}_i^m)\exp(-\lambda(\mathbf{x}_i^m))$ $P(y_i^m = q \mid \mathbf{x}_i^m, \mathbf{z}_i^m) = \frac{p(\mathbf{z}_i^m)\exp(-\lambda(\mathbf{x}_i^m))\lambda(\mathbf{x}_i^m)^q}{q!}$ $q = 1, 2, \ldots$	$P(y_i^m = 0 \mid \mathbf{z}_i^m) = 1 - p(\mathbf{z}_i^m)$ $P(y_i^m = q \mid \mathbf{x}_i^m, \mathbf{z}_i^m) = \frac{p(\mathbf{z}_i^m)\exp(-\lambda(\mathbf{x}_i^m))\lambda(\mathbf{x}_i^m)^q}{q!(1-\exp(-\lambda(\mathbf{x}_i^m)))}$ $q = 1, 2, \ldots$	
Link	$p_j(\mathbf{z}_i)$ logit $\lambda_j(\mathbf{x}_i)$ log-linear	$p_j(\mathbf{z}_i)$ logit $\lambda_j(\mathbf{x}_i)$ log-linear	
Mean	$E[y_i^m \mid \mathbf{x}_i^m, \mathbf{z}_i^m] = p(\mathbf{z}_i^m)\lambda(\mathbf{x}_i^m)$	$E[y_i^m \mid \mathbf{x}_i^m, \mathbf{z}_i^m] = \frac{p(\mathbf{z}_i^m)\lambda(\mathbf{x}_i^m)}{1-\exp(-\lambda(\mathbf{x}_i^m))}$	
Variance	$Var[y_i^m \mid \mathbf{x}_i^m, \mathbf{z}_i^m] = E[y_i^m \mid \mathbf{x}_i^m, \mathbf{z}_i^m]\Big[1 + \lambda(\mathbf{x}_i^m)$ $- E[y_i^m \mid \mathbf{x}_i^m, \mathbf{z}_i^m]\Big]$	$Var[y_i^m \mid \mathbf{x}_i^m, \mathbf{z}_i^m] = \frac{\lambda(\mathbf{x}_i^m)}{1-\exp(-\lambda(\mathbf{x}_i^m))}\Big[1 - \exp(-\lambda(\mathbf{x}_i^m))E[y_i^m \mid \mathbf{x}_i^m, \mathbf{z}_i^m]\Big]$	
LL	$L_{ZIP}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{y_i^m=0}\log(\exp(\mathbf{z}_i^m\boldsymbol{\alpha})$ $+\exp(-\exp(\mathbf{x}_i^m\boldsymbol{\beta})))$ $+\sum_{y_i^m>0}(y_i^m\mathbf{x}_i^m\boldsymbol{\beta} - \exp(\mathbf{x}_i^m\boldsymbol{\beta}))$ $-\sum_{i=1}^n\log(1+\exp(\mathbf{z}_i^m\boldsymbol{\alpha})) - \sum_{y_i^m>0}\log(y_i^m!)$	$L_{Hurdle}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{y_i^m=0}\log\left(\frac{1}{1+\exp(\mathbf{z}_i^m\boldsymbol{\alpha})}\right) + \sum_{y_i^m>0}\log\left(\frac{\exp(\mathbf{z}_i^m\boldsymbol{\alpha})}{1+\exp(\mathbf{z}_i^m\boldsymbol{\alpha})}\right)$ $+\sum_{y_i^m>0}\Big(y_i^m\mathbf{x}_i^m\boldsymbol{\beta} - \exp(\mathbf{x}_i^m\boldsymbol{\beta}) - \log(1-\exp(-\exp(\mathbf{x}_i^m\boldsymbol{\beta}))) - \log(y_i^m!)\Big)$ $= L_{Hurdle}(\boldsymbol{\alpha}) + L_{Hurdle}(\boldsymbol{\beta})$	
Estimation Proc.	Maximum likelihood STATA ZIP function	Maximum likelihood $L_{Hurdle}(\boldsymbol{\alpha})$ SAS LOGISTIC procedure $L_{Hurdle}(\boldsymbol{\beta})$ maximised via Nelder-Mead algorithm	

Problems caused by changing diagnostic policy were eliminated by collecting information retrospectively from pathology and autopsy reports. Full details of the SIDS data are given in Dalrymple (2004). Figure 1 shows the monthly SIDS series for the 32 years. The Canterbury climate data analysed in this study was obtained from NIWA, Taihoro Nukurangi [http://www.niwa.co.nz/]. This data is unusual in the field of SIDS-climate research, as SIDS deaths were localised around the site of the meteorological data collection.

It is widely known that there are annual trends in birth numbers and also changes in birth numbers over time. The logarithm of the number of infants at risk (NAR) of SIDS each month was defined to account for underlying population changes (Table 3). A change point analysis of the SIDS/1000 live births profile over the 32 years of the study, which implements a novel block bootstrap method (Dalrymple et al., 2001), found two significant change points; at December 1972 and May 1990. These points effectively partition the series into three epochs or periods: Period 1 (1968–1972), Period 2 (1973–1989) and Period 3 (1990–2000), and are illustrated by vertical (dashed) lines on Figure 1. For brevity only period 2 (1973–1989) results are reported here.

4.2 Climate Models: Period 2

The k=2 component FM model was found to be the best candidate FM model to fit the Period 2 SIDS-climate series. All best climatic models for each of five methods are presented in Table 3 (where $X_{mean(std)}(i)$ is the mean over month i of the daily standard deviation of X and $X_{min(min)}(i)$ the minimum over month i of X; for X any one of the eleven climate variables (Section 1)). All models are constructed similarly, with $WindS_{mean(std)}$ the only significant climate variable in the Poisson rate component, over and above a full baseline model (intercept term, seasonal sinusoid and NAR) for four of the five models. Covariate coefficients of significant climate factors are comparable across $\lambda(\mathbf{X})$ for all five models. The NB model contained only seasonality and $WindS_{mean(std)}$, while the second component of the FM model comprised only the seasonal sinusoid (Table 3). The FM, ZIP and hurdle models exhibit some differences in their covariate structure of the probability component $p(\mathbf{Z})$. The covariate found to best fit the mixing proportions for the FM model was the seasonal sinusoid. The best covariate combination for both the ZIP and hurdle model contained the climate variables, $Humid_{min(min)}$ and $WindS_{mean(std)}$ respectively, in addition to the seasonal sinusoid. These three models which contain $p(\mathbf{Z})$ also exhibited differing covariates. It is interesting to note that the climate-EPP model (containing $WindS_{mean(std)}$) has an index c is -0.044, which indicates slight underdispersion. Both the EPP model and NB models perform poorly in comparison to the FM, ZIP and hurdle models. Both the FM and ZIP models performed well in terms of model fit and residual diagnostics and their overall performance (Dalrymple, 2004). The FM model and ZIP model also performed well with respect to goodness-of-fit statistics (Table 3), indicating sufficient evidence that the FM model, or ZIP model, do adequately fit the monthly SIDS data. We observe that the ZIP model is, broadly speaking, the FM model with the additional humidity term. For conciseness only the ZIP profile of SIDS risk is detailed here.

The ZIP model includes two climate covariates: $WindS_{mean(std)}$ in the Poisson rate component and $Humid_{min(min)}$ in the probability component. The Poisson rate component contains a full baseline form ($intercept+\sin(\frac{2\pi t}{12})+\cos(\frac{2\pi t}{12})+NAR$), in addition to $WindS_{mean(std)}$ and the autoregressive covariate y_{i-1}. The probability component structure is defined by the seasonal sinusoid and also humidity (Table 3). Increasing $WindS_{mean(std)}$ corresponds to an increase in the risk of SIDS: the lowest estimated risk of 0.85 SIDS per month occurs at 2.9 knots in summer, whereas the highest estimated risk of 3.38 SIDS per month at 5.8 knots in winter. The risk equates to a 15% increase in the SIDS rates for every one knot increase in the monthly average of the daily standard deviation of wind speed. Essentially this implies that the more variable the wind speed is, on average, the higher the estimated SIDS risk. This same trend was evident with respect to wind speed in the FM model. Note that humidity appeared exclusively in the ZIP model and only in its probability component (Table 3). Trends in relation to humidity are as follows: an increase in the minimum monthly humidity ($Humid_{min(min)}$) corresponds to a decrease in probability of belonging to the 'perfect state' (no SIDS deaths occurring). None of the models however, achieved the necessary inflation of SIDS counts evident in the observed series in winter (observed $mean = 3.7$ SIDS per month) (Dalrymple, 2004). This feature and the systematic lack of fit at extreme SIDS counts (see Dalrymple (2004)) evident across all five methods; in addition to the need to include AR state-dependence, may indicate that some additional latent variable, whether environmental (pollution), climatic or physiological may be involved in the aetiology of SIDS.

5 DISCUSSION

Recent research has focussed on extending various forms of mixture models to serially correlated data arising from repeated measures studies (Hall, 2000). This paper has developed mixture models for the (discrete count(s)) time series context. By including

Table 3. Climate model parameter estimates (standard errors) for the five mixture methods, Period 2. Optimal models are bolded.

		NB	**FM**		**ZIP**	Hurdle	EPP
			Comp 1	**Comp 2**			($\log(a)$)
$\log(\lambda(\mathbf{x}))$	Intercept		**11.23**		**10.40**	13.63	8.89
			(3.76)		**(3.95)**	(1.97)	(3.98)
	$\sin(\frac{2\pi t}{12})$	-0.42	**-0.63**	**-0.34**	**-0.35**	-0.38	-0.42
		(0.07)	**(0.18)**	**(0.10)**	**(0.08)**	(0.04)	(0.07)
	$\cos(\frac{2\pi t}{12})$	-0.50	**-0.03**	**-1.01**	**-0.48**	-0.49	-0.51
		(0.07)	**(0.02)**	**(0.76)**	**(0.08)**	(0.04)	(0.07)
	NAR		**-1.12**		**-1.18**	-1.57	-1.00
			(0.43)		**(0.47)**	(0.24)	(0.48)
	$WindS_{mean(std)}$	0.22	**0.15**		**0.13**	0.15	0.13
		(0.02)	**(0.07)**		**(0.07)**	(0.04)	(0.07)
	y_{i-1}	-0.08	-0.10		-0.10	-0.10	-0.09
		(0.03)	(0.03)		(0.03)	(0.01)	(0.03)
$\text{logit}(p(\mathbf{z}))$	$\sin(\frac{2\pi t}{12})$		**3.76**		**-5.72**	-1.21	
			(0.89)		**(2.45)**	(0.36)	
	$\cos(\frac{2\pi t}{12})$		**-14.15**		**-1.86**	-1.26	
			(2.76)		**(1.72)**	(0.36)	
	$Humid_{min(min)}$				**0.11**		
					(0.06)		
	$WindS_{mean(std)}$	$d = 0.013$				0.51	$b = 1.220$
						(0.07)	$c = -0.044$
m		5	**10**		**9**	9	8
AIC		710.79	**708.62**		**710.12**	711.24	713.21
BIC		728.06	**741.80**		**739.99**	750.10	747.75
χ^2		191.16	**173.45**		**182.95**	189.07	180.92

an autoregressive-based regressor, serial dependence within an observed series can be successfully accommodated. The application of mixture models in this paper, to a discrete counts time series is unique, and presents potential important extensions for future research. The mixture models in this paper are also a significant extension of the general class of conditional linear AR(1) models (CLAR(1)) (Grunwald et al., 2000), to the case where covariates are allowed in the conditional mean and a generalised log link for the conditional mean is assumed.

The analysis of Canterbury SIDS counts suggested that the incidence of SIDS is associated with humidity and wind speed in Period 2. These relationships are identified, after accommodating for seasonality. The mixture methods in this paper allude to possible causal pathways, where climate possibly acts on changing parental care, which in turn, has an impact on SIDS risk. For example, wind has long been associated with ill health. Indeed the warm, dry alpine winds, such as Canterbury's Norwester have been related to cardiovascular problems, migraines and allergies (Dalrymple, 2004). It is thus of interest to note the significant wind variants in our best fit models. It is not surprising that variations in wind may have an effect on SIDS, in that wind variations, unlike air pressure changes, are easily perceived by humans, and may modify parental care practices. Similar hypotheses in regard to effects of increased temperature (or decreased dewpoint) with sustained overwrapping have recently been presented (Dalrymple et al., 2003; Dalrymple, 2004).

SIDS remains a leading cause of infant death in the western world, thus any comprehensive analysis that leads to a more refined understanding of the risk profile of a SIDS victim, may help in the identification of the underlying aetiology and cause(s) of this syndrome. Hudson et al. (2005) recently showed that climatic impacts on SIDS are not just a proxy for pollution impacts. Further NZ research of air pollution and SIDS is still needed to clarify the pollution-climate-SIDS interplay which remains controversial. Future work will also involve generalizing our methods to the Tweedie family of distributions (Dunn & Smyth, 2005).

6 REFERENCES

Auliciems, A. & Barnes, A. (1987). Sudden infant deaths and clear weather in a subtropical environment. *Social Science in Medicine*, 24(1), 51–56.

Böhning, D. & Seidel, W. (2003). Editoral: Recent Developments in Mixture Models. *Computational Statistics and Data Analysis*, 41(3-4), 349–357.

Bollerslev, T., Chou, R., & Kroner, K. (1992). ARCH Modeling in Finance. A Review of the Theory and Empirical Evidence. *J. Econom.*, 52, 5–59.

Booth, J., Casella, G., Friedl, H., & Hobert, J. (2003). Negative binomial loglinear mixed models. *Statistical Modelling*, 3, 179–191.

Dalrymple, M. (2004). *Poisson Mixture Methods and Change Point Analyses to Study the Relationship Between Temporal Profiles of Sudden Infant Death Syndrome and Climate*. PhD thesis, University of Canterbury, Christchurch, New Zealand.

Dalrymple, M., Hudson, I., & Barnett, A. (2001). Survival, block bootstrap and mixture methods for detecting change points in discrete time series data with application to SIDS. In B. Klein & L. Karsholm (Eds.), *Proc. of the 16th International Workshop on Statistical Modelling* (pp. 135–146).

Dalrymple, M., Hudson, I., & Ford, R. (2003). Finite mixture, zero-inflated Poisson and hurdle models with application to SIDS. *Computational Statistics and Data Analysis*, 41(3-4), 491–504.

Dobbie, M. & Welsh, A. (2001). Modelling Correlated Zero-Inflated Count Data. *Australian & New Zealand Journal of Statistics*, 43(4), 431–444.

Dunn, P. & Smyth, G. (2005). Series evaluation of Tweedie exponential dispersion model densities. *Stats & Computing*, 15, 267–280.

Faddy, M. (1997a). Extended Poisson process modelling and analysis of count data. *Biometrical Journal*, 39(4), 431–440.

Faddy, M. (1997b). On extending the negative binomial distribution, and the number of weekly winners of the UK national lottery. *Mathematical Scientist*, 22, 77–82.

Faddy, M. J. & Smith, D. M. (2005). Modelling the dependence between the number of trials and success probabilities in binary trials. *Biometrics*, 61(4), 1112–1114.

Greene, W. (2000). *Econometric Analysis*. Upper Saddle River, New Jersey: Prentice Hall, 4th edition.

Grunwald, G., Hyndman, R., Tedesco, L., & Tweedie, R. (2000). Non-Gaussian conditional linear AR(1) models. *Australian and New Zealand Journal of Statistics*, 42(4), 479–495.

Hall, D. (2000). Zero-inflated Poisson and binomial regression with random effects: A case study. *Biometrics*, 56, 1030–1039.

Hudson, I., Fukuda, K., & Dalrymple, M. L. (2005). Climate-pollution impacts on sudden infant death (SIDS). In *16th congress of the Modelling and Simulation Society of Australia and New Zealand, MODSIM 2005* (pp. 1–7).

Jung, R. & Tremayne, A. (2003). Testing for Serial Dependence in Time Series Models of Counts. *Journal of Time Series Analysis*, 24(1), 65–84.

Knöbel, H., Chen, D., & Liang, K. (1995). Sudden infant death syndrome in relation to weather and optimetrically measured air pollution in Taiwan. *Pediatrics*, 96(6), 1106–1110.

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

Macey, P., Schluter, P., & Ford, R. (2000). Weather and the risk of sudden infant death syndrome: the effect of wind. *Journal of Epidemiology and Community Health*, 54, 333–339.

Malloy, M. & Freeman, D. (2000). Birth weight- and gestational age-specific sudden infant death syndrome mortality: United States, 1991 versus 1995. *Pediatrics*, 105, 1227–1231.

McLachlan, G. & Peel, D. (2000). *Finite Mixture Models*. Wiley Series in Probability and Statistics, Applied Probability and Statistics Section. John Wiley & Sons, Inc.

Min, Y. & Agresti, A. (2005). Random effect models for repeated measures of zero-inflated count data. *Statistical Modelling*, 5, 1–19.

Mooney, J., Helms, P., & Jolliffe, I. (2003). Fitting Mixtures of Von Mises Distributions: A Case Study Involving Sudden Infant Death Syndrome. *Computational Statistics and Data Analysis*, 41(3-4), 505–513.

Podlich, H. M., Faddy, M. J., & Smyth, G. K. (2004). Semi-parametric extended poisson process models. *Statistics and Computing*, 14, 311–321.

Raftery, A. & Tavare, S. (1994). Estimation and Modelling Repeated Patterns in High Order Markov Chains with the Mixture Transition Distribution Model. *Appl. Statist.*, 43, 179–199.

Toscas, P. & Faddy, M. (2003). Likelihood based analysis of longitudinal count data using a generalised Poisson model. *Statistical Modelling*, 3, 99–108.

Wang, P., Cockburn, I., & Puterman, M. (1998). Analysis of patent data – a mixed-Poisson-regression-model approach. *Journal of Business & Economic Statistics*, 16(1), 27–40.

Welsh, A., Cunningham, R., Donnelly, C., & Lindenmayer, D. (1996). Modelling the abundance of rare species: statistical models for counts with extra zeros. *Ecological Modelling*, 88, 297–308.