Finite Mixture, Zero-inflated Poisson and Hurdle models with application to SIDS

M.L. Dalrymple\textsuperscript{a,∗}, I.L. Hudson\textsuperscript{a}, R.P.K. Ford\textsuperscript{b}

\textsuperscript{a}Department of Mathematics and Statistics, University of Canterbury, Private Bag 4800, Christchurch, New Zealand
\textsuperscript{b}Community Paediatric Unit, Christchurch Public Hospital, Christchurch, New Zealand

Received 1 February 2002; received in revised form 1 March 2002

Abstract

This study examines the incidence of sudden infant death syndrome (SIDS) in Canterbury (1973–1989) in relation to climate. Three mixture models (Finite Mixture, Zero-inflated Poisson and Hurdle) are used as novel methods which are able to highlight differential effects of climatic covariates between months of SIDS and no SIDS. These methods accommodate the extra zeros, heterogeneity and autocorrelation found in the SIDS series. Mixture models are comprehensive methods applicable to many discrete chronological series including the Canterbury SIDS data. This analysis leads to a better understanding of the association between climate and SIDS deaths.

Results show a deviance-temperature (a measure of extreme change from the fortnightly average) is significantly associated with SIDS risk ($p < 0.005$). Months where there is a high deviance-temperature are associated with increased risk of SIDS, compared to months where the temperature has remained reasonably constant. This finding is consistent with the theory that hyperthermia, or overheating of infants leads to increased SIDS risk. In months where at least one SIDS death occurs, increased humidity leads to increased risk of SIDS ($p < 0.001$).

© 2002 Elsevier Science B.V. All rights reserved.

Keywords: Mixture models; Heterogeneity; Excess zeros; Sudden infant death syndrome; Climate

1. Introduction

Sudden infant death syndrome (SIDS or cot death) is characterised by the sudden and unexplained death of an infant and currently accounts for between 10% and 20% of all infant deaths in developed countries (Malloy and Freeman, 2000). Many factors in
an infant’s environment have been shown to contribute to the risk of SIDS (Mitchell et al., 1992), particularly infant sleep position, maternal smoking and seasonal temperature changes.

Previous studies have shown SIDS rates to have a general seasonal dependence, with more SIDS deaths occurring in winter (Mitchell et al., 1999; Gilman et al., 1995). Alongside this, short term changes in ambient temperature lead to increased SIDS risk (Campbell, 1989). The SIDS-climate relationship has been examined extensively in terms of seasonal effects and long or short-term ambient temperature (Gilman et al., 1995; Schluter et al., 1998). Few studies, however, have included additional weather variables, such as relative humidity, in the analysis. This study investigates temperature and humidity as climatic risk factors.

Traditionally Poisson regression has been used to model the rare count profile of SIDS. In this study, three forms of mixture modelling (Finite Mixture, Zero-inflated Poisson, and Hurdle models) are used to accommodate the extra zeros, heterogeneity and autocorrelation found in the SIDS series examined here. These methods are able to highlight differential effects of climatic covariates between months of SIDS and no SIDS, while keeping the underlying Poisson formulation. Analyses of this type lead to a better understanding of the association of climate with SIDS deaths. These covariate based models have not been applied previously to a temporal sequence of SIDS counts, and their application in this form is uncommon in the general time series framework.

2. Canterbury SIDS and climate data

A retrospective, complete ascertainment study was completed by the Christchurch Community Paediatric Unit which led to daily SIDS counts from 1968–1999. Problems caused by changing diagnostic policy over time were eliminated by collecting information from pathology records and autopsy reports, resulting in an unusually and uniquely accurate chronological profile of approximately 12,000 daily SIDS counts. With only 658 SIDS deaths in Canterbury over this time, there are approximately 94% of days with no SIDS occurring.

Change point analysis has shown two significant points of change in this series, at 1972 and 1989 (Dalrymple et al., 2001), effectively partitioning the series into three periods. The later change point corresponds to local publicity campaigns aimed at reducing the rate of prone sleeping (placing infants on their stomachs), which successfully reduced the SIDS rate by two-thirds (Mitchell et al., 1994). The analysis presented here examines aggregate monthly SIDS counts for the middle period, 1973–1989. Analysis of this period offers the best chance of highlighting differential climatic covariate effects due to its length (17 years) and its high SIDS rates (2.39 SIDS per month compared to 1.24 in the first period and 0.76 in the third period (Dalrymple et al., 2001)).

Table 1 summarises the covariates considered in this analysis, including the number of infants at risk, humidity, temperature, and interpretable temperature variants. There is no constant term used for the Poisson modelling as the covariate incorporating
Table 1
Summary of covariates to be considered in the mixture analyses

<table>
<thead>
<tr>
<th>Name</th>
<th>Definition (for month $i, i = 1, \ldots, n$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAR</td>
<td>$\log($number of infants at risk of SIDS/1000$)$</td>
</tr>
<tr>
<td>Humid</td>
<td>$\log(\text{mean(max daily relative humidity in HPa)})$</td>
</tr>
<tr>
<td>Temp</td>
<td>$\log(\text{mean(min daily temp in } ^\circ \text{C}))$</td>
</tr>
<tr>
<td>PosTemp</td>
<td>$\text{mean(all positive temp}_{14}(\text{day}))$</td>
</tr>
<tr>
<td>NegTemp</td>
<td>$\text{mean(all negative temp}_{14}(\text{day}))$</td>
</tr>
<tr>
<td>Sin</td>
<td>$\sin(2\pi/12)$</td>
</tr>
<tr>
<td>Cos</td>
<td>$\cos(2\pi/12)$</td>
</tr>
<tr>
<td>$Y_{i-t}$</td>
<td>Lagged SIDS counts (autoregressive covariate)</td>
</tr>
</tbody>
</table>

the number of infants at risk essentially acts as a surrogate constant, yet includes information about the population at risk (Rothman, 1986). All climate information was recorded at Christchurch International Airport by the New Zealand Meteorological Service.

Potential autocorrelation is incorporated into the models using an autoregressive covariate of lagged SIDS counts. The autoregressive structure of the residuals of any potential model is examined, and then, if necessary, the appropriate lagged variable is included in the model.

3. Poisson mixture model theory

The Poisson distribution, and specifically, Poisson regression, has traditionally been used to model SIDS data (for example, Knobel et al., 1995), but is limited in this application due to the extra variation in the SIDS series. The Poisson mixture methodology presented here, Finite Mixture, Zero-inflated Poisson and Hurdle models, account for both heterogeneity and the extra zeros in the data. This results in a more accurate analysis, with a more comprehensive understanding of the relationship between SIDS and climate.

3.1. Finite Mixture model

Finite Mixture (FM) models have become increasingly popular in the analysis of a wide range of data. The covariate adjusted mixture model was presented by Schlattmann et al. (1996) in a disease mapping context, while Wang et al. (1996) presented two medical examples (seizure frequency and Ames salmonella assay data) of mixed Poisson regression models with covariate dependent rates.

The FM model assumes an underlying partition of the population into $k$ homogeneous components, where each component has a different SIDS risk level ($\lambda_j$), dependent on possibly different covariates. This is a general model which allows mixing with respect to both zeros and positives. The FM formulation is given as follows:
Let $y_i$ represent the $i$th response variable ($i = 1, \ldots, n$); $y_i \sim \text{Poisson}(\lambda_j(z))$ with probability $p_j(x)$, where $x$ and $z$ are matrices of covariates. The model is then

$$P(y_i = q|x, z) = \sum_{j=1}^{k} \frac{p_j(x)\lambda_j(z)^q \exp(-\lambda_j(z))}{q!}$$

where $k$ represents the number of components in the model. Using logit and log-linear links to model $p_j(x)$ and $\lambda_j(z)$, respectively gives:

$$\text{logit}(p_j(x)) = \log \left(\frac{p_j(x)}{1 - p_j(x)}\right) = x_j^\top \beta_j,$$

$$\log(\lambda_j(z)) = z_j^\top \alpha_j$$

for unknown parameters $\beta_j$ and $\alpha_j$. Note that the covariates in the mixing proportions ($x$) are not restricted to being the same as those in the Poisson rates ($z$). The regression coefficients ($\alpha_j$) may vary between components, or be 0 for one or more covariates, that is $\alpha_{j} = 0$ for some $j$ ($j = 1, \ldots, k$) implying that the predictor corresponding to $z_j$ does not have an impact in the $j$th component.

### 3.2. Zero-inflated Poisson model

Zero-inflated Poisson (ZIP) models provide another way to model count data with excess zeros. The response variable is modelled as a mixture of a Bernoulli distribution and a Poisson distribution. Although ZIP models without covariates have a long history (for example, Johnson and Kotz, 1969), more recently, Lambert (1992) provided the general form of ZIP regression models incorporating covariates.

Lambert presented an application of defects in manufacturing where it was suggested that “one interpretation [of the excess zero-distribution] is that slight, unobserved changes in the environment cause the process to move randomly back and forth between a perfect state in which defects are extremely rare and an imperfect state in which defects are possible but not inevitable”. The existence of such a perfect state increases the number of zero counts in the data.

In ZIP regression (Böning et al., 1999; Dietz and Böning, 2000), a response vector of counts $y_i$, $i = 1, \ldots, n$, are independent; $y_i = 0$ with probability $1 - p(x)$ and $y_i \sim \text{Poisson}(\lambda(z))$ with probability $p(x)$, where $x$ and $z$ again represent covariate matrices. The model is then

$$P(y_i = 0|x, z) = 1 - p(x) + p(x) \exp(-\lambda(z)),$$

$$P(y_i = q|x, z) = \frac{p(x)\exp(-\lambda(z))\lambda(z)^q}{q!}, \quad q = 1, 2, \ldots .$$

Here $(1 - p(x))$ is the probability of the perfect state and $p(x)$ is the probability that the number of events has a Poisson distribution. Conditional on the imperfect state, $\lambda(z)$ is the mean number of events occurring. As with the FM model, a logit and log-linear link are used to model $p(x)$ and $\lambda(z)$, respectively, with covariates $x$ and $z$ potentially differing.
The ZIP model can be seen as a special case of a 2-component FM model, with no covariates in the mixing probabilities (Wang et al., 1998). One component is taken as a degenerate distribution, with mass 1 at \( y_i = 0 \), while the other component is a Poisson regression model. The ZIP model is more restrictive than the general FM model in that it only allows mixing with respect to zeros.

### 3.3. Hurdle model

The Hurdle model is a two part model; the first part is a binomial probability model which determines whether a zero or non-zero outcome occurs. A truncated count data distribution, which describes the positive outcomes is modelled as the second part. The idea behind this formulation is that given an event has occurred, that is, the “hurdle has been crossed”, the conditional distribution of this event is controlled by a truncated-at-zero distribution (Cameron and Trivedi, 1998).

The hurdle specification is defined as follows in the Poisson case: the counts \( y_i, i = 1, \ldots, n \) are independent; \( y_i = 0 \) with probability \( 1 - p(x) \) and \( y_i \sim \text{truncated Poisson} \left( \lambda(z) \right) \) with probability \( p(x) \), where \( x \) and \( z \) are matrices of covariates. The model is then

\[
P(y_i = 0|x) = 1 - p(x),
\]

\[
P(y_i = q|x,z) = \frac{p(x) \exp(-\lambda(z)) \lambda(z)^q}{q!(1 - \exp(-\lambda(z)))}, \quad q = 1, 2, \ldots.
\]

This formulation, presented by Welsh et al. (1996), increases the probability of the zero outcome and scales the remaining probabilities so that they all sum to one. Again, \( p(x) \) and \( \lambda(z) \) are modelled by logit and log-linear functionals, respectively, and correspondingly the covariates \( x \) are not restricted to being the same as the covariates \( z \).

The Hurdle model formulation is very similar to the ZIP model. Both models essentially combine binomial probabilities with Poisson distributions, but the Hurdle model keeps the zero-class disjoint from the non-zeros by modelling the non-zero \( y_i \)'s with a truncated Poisson distribution. This differs from the ZIP model where zeros occur both in the ‘perfect state’ (with probability \( 1 - p(x) \)) and the Poisson distribution (with probability \( p(x) \exp(-\lambda(z)) \)).

### 4. Parameter estimation and algorithms

This section outlines algorithms implemented to find parameter estimates for each method.

#### 4.1. Finite Mixture model

Estimation of model parameters for the FM models is achieved following Wang’s approach (1998), where maximum likelihood estimates (MLEs) are obtained using a combination of the EM and quasi-Newton algorithms. The maximum likelihood formulation is as follows;
Define $Z_{ij}$ as an unobserved indicator variable representing component membership. For example, $Z_i = (0, 1, 0, \ldots, 0)^T$ indicates that the $i$th observation belongs in the second mixture component. The log-likelihood ($LL$) for the complete data, where the $Z_{ij}$s are considered as missing values, is:

$$
LL_{FM} = \sum_{i=1}^{n} \sum_{j=1}^{k} Z_{ij} \left[ x_i \beta_j - \log \left( 1 + \sum_{l=1}^{k-1} \exp(x_i \beta_l) \right) \right]
$$

$$
+ \sum_{i=1}^{n} \sum_{j=1}^{k} Z_{ij} \left( y_i z_i x_j - \exp(z_i x_j) - \log(y_i!) \right).
$$

Wang’s algorithm ((1996) reproduced in the Appendix for convenience) and Fortran code\(^1\) are used to find MLEs. Initial values for the EM algorithm are chosen using his sequential approach, which is based on cluster analysis (Wang et al., 1998).

### 4.2. Zero-inflated Poisson model

Lambert (1992) developed an EM algorithm which maximises the log-likelihood for ZIP regression (see Appendix A). This involves fitting two general linear models (GLMs); a logistic and a Poisson regression model (see Appendix A for details). The log-likelihood is:

$$
LL_{ZIP} = \sum_{y_i = 0} \log(\exp(x_i \beta) + \exp(-\exp(z_i x)))
$$

$$
+ \sum_{y_i > 0} (y_i z_i x - \exp(z_i x)) - \sum_{i=1}^{n} \log(1 + \exp(x_i \beta))
$$

$$
- \sum_{y_i > 0} \log(y_i!).
$$

Implementation of this procedure was achieved using the ‘ZIP’ function in Stata (Intercooled Stata, Version 6; Stata Corporation).

### 4.3. Hurdle model

As with the FM and ZIP methods, a MLE procedure is used to model the SIDS data with the Hurdle model. The log-likelihood for the Hurdle model is:

$$
LL_{Hurdle} = \sum_{y_i > 0} x_i \beta - \sum_{i=1}^{n} \log(1 + \exp(x_i \beta))
$$

$$
+ \sum_{y_i > 0} \left[ y_i z_i x - \exp(z_i x) - \log(1 - \exp(-\exp(z_i x))) - \log(y_i!) \right]
$$

$$
= LL(\beta) + LL(x).
$$

\(^1\) Available at [http://coe.ubc.ca/users/marty/research.html](http://coe.ubc.ca/users/marty/research.html)
The two components of the Hurdle model may be fitted separately as (8) is the sum of two distinct likelihoods. The first component, $LL(\beta)$, is a likelihood based on a logistic model and, as such, can be fitted by standard GLM software. The second component, $LL(\alpha)$, is fitted by a specifically constructed Matlab macro which uses the Nelder–Mead algorithm to maximise $LL(\alpha)$ (Matlab, Version 6.0.0.88, Release 12; The MathWorks, Inc.).

4.4. Model selection

All three methods use a penalised-likelihood approach for model selection. Akaike’s information criterion (AIC) and the Bayesian information criterion (BIC) have been widely used previously for mixture model selection (for example Wang et al., 1996) and are used in the following form:

\[ \text{AIC} : \text{choose the model which minimises } \ -2LL + m, \]  

\[ \text{BIC} : \text{choose the model which minimises } \ -2LL + \log(n)m, \]  

where $LL$ is the log-likelihood, $m$ is the number of parameters in the model and $n$ the total number of observations.

The above statistics, AIC and BIC, are used in a two-step model selection procedure for FM models. The first step is to establish the number of components, $k$. This involves comparing saturated models (models containing all possible covariates) to determine which $k$ minimises AIC/BIC values (given by $\hat{k}$). Once $\hat{k}$ is determined, the appropriate SIDS model is selected from all $\hat{k}$-component mixture models.

Three statistics, AIC, BIC and the mean square error (MSE), are used for a comparative examination of the models resulting from each method. The MSE value is calculated as a goodness-of-fit measure with $\text{MSE} = E[(y - \hat{y})^2]$ (Cameron and Trivedi, 1998).

5. Results: SIDS and climate application

Table 2 reports LL, AIC and BIC values for the saturated FM models (without a constant or autoregressive term), assuming one, two or three components for the SIDS data. These penalised likelihood statistics suggest different $k$-component models.

<table>
<thead>
<tr>
<th>$k$</th>
<th>$m$</th>
<th>Log-likelihood</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>$-350.98$</td>
<td>709.96</td>
<td>747.70</td>
</tr>
<tr>
<td>2</td>
<td>24</td>
<td>$-342.48$</td>
<td>708.95</td>
<td>812.59</td>
</tr>
<tr>
<td>3</td>
<td>40</td>
<td>$-327.50$</td>
<td>695.00</td>
<td>867.72</td>
</tr>
</tbody>
</table>

Note: these models do not include a constant or an autoregressive term.
Finite Mixture model estimates for various two-component models

<table>
<thead>
<tr>
<th>Model number</th>
<th>Comp number</th>
<th>NAR</th>
<th>Humid</th>
<th>Sin</th>
<th>Cos</th>
<th>$Y_{i-1}$</th>
<th>Mixing probs</th>
<th>m</th>
<th>LL</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1.68</td>
<td></td>
<td></td>
<td></td>
<td>-0.47</td>
<td></td>
<td>3</td>
<td>-392.57</td>
<td>788.14</td>
<td>801.09</td>
</tr>
<tr>
<td>2</td>
<td>0.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>-2.92</td>
<td>1.48</td>
<td></td>
<td></td>
<td>-1.87</td>
<td></td>
<td>5</td>
<td>-386.36</td>
<td>777.72</td>
<td>799.31</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>19.94</td>
<td>-7.23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>-2.10</td>
<td>1.14</td>
<td>-0.31</td>
<td>-0.37</td>
<td>-2.96</td>
<td></td>
<td>9</td>
<td>-355.62</td>
<td>720.24</td>
<td>759.10</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>2.96</td>
<td>0.44</td>
<td>-0.41</td>
<td>-2.76</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>-2.37</td>
<td>1.33</td>
<td>-0.42</td>
<td>-0.47</td>
<td>-0.10</td>
<td>-2.93</td>
<td>2</td>
<td>-11.2</td>
<td>7</td>
<td>-350.76</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>-4.02</td>
<td>-0.98</td>
<td>-0.01</td>
<td>-6.49</td>
<td>-0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>-2.46</td>
<td>1.37</td>
<td>-0.39</td>
<td>-0.43</td>
<td>-0.09</td>
<td>-1.00</td>
<td>11</td>
<td>-349.44</td>
<td>709.88</td>
<td>757.38</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>3.65</td>
<td>0.71</td>
<td>-0.80</td>
<td>-3.12</td>
<td>-0.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>-25.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.81</td>
<td>7</td>
<td>-351.38</td>
<td>709.76</td>
<td>739.99</td>
</tr>
<tr>
<td>2</td>
<td>2.39</td>
<td>1.33</td>
<td>-0.41</td>
<td>-0.48</td>
<td>-0.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

as being optimal. For example, BIC is minimised by a one-component model with eight parameters, while a three-component model returns a minimum AIC value (40 parameters). The prime motivation of the study is to highlight differences in risk factors between months of no or low SIDS versus months of high SIDS; as such a $k > 1$ component model is required. A two-component model (24 parameters) is selected as it returns the second-lowest AIC and BIC values, and accommodates the risk model interpretation in a more parsimonious way than the $k = 3$ model (40 parameters).

Various two-component models are presented in Table 3. Whilst all covariate combinations were examined for $k = 2$ FM models, only a subset of models which contain significant covariates are reported. The predictors $Temp$, $PosTemp$ and $NegTemp$ were not useful in predicting the Poisson rates, while only a constant and $PosTemp$ were useful in modelling $p$, the mixing probability (refer to Table 1 for variable definitions). Models 1–5, presented in Table 3, show a consistent decrease in LL values as predictors are systematically added.

The FM model found to best fit the SIDS-climate data is Model 6 (Table 3), having the lowest $AIC$ (709.76) and $BIC$ (739.99) values. The first component contains only NAR, whilst the second component has $NAR$, $Humid$, $Sin$, $Cos$ and $Y_{i-1}$ as predictors.

A similar comprehensive model search was completed for both the ZIP and Hurdle methods (details not reported here). The ZIP and Hurdle methods resulted in best fitting models very similar to FM Model 6, and are presented in Table 4. Component 2 of the FM model has identical covariates, with comparable coefficients as the Poisson regression from the ZIP model and the truncated Poisson regression from the Hurdle model.

In all three models presented in Table 4, the probability of a month having at least one SIDS occurrence is dependent on only one climatic predictor, namely $PosTemp$, the positive deviance temperature covariate. The probability of SIDS increases as $PosTemp$ increases as shown in Fig. 1, where the asterisks highlight peaks in $PosTemp$ matching
with peaks in the probability of at least one SIDS occurring. Best fit models also show increasing $NAR$ and $Humid$ lead to increased SIDS. Also a significant seasonal component is evident with more SIDS in colder months, as accounted for by a sinusoid function (a combination of Sin and Cos).

Fig. 2 shows the predicted number of SIDS per month overlaying the observed numbers for the models presented in Table 4. As expected with such similar coefficients and covariates, the plots look essentially the same. The Hurdle model seems less able to predict extreme SIDS numbers, yet none of the models achieve good prediction of the extremes. This point is discussed in the next section.

Table 5 reports the estimated mean values and proportions of months with zero and non-zero SIDS counts for the three models presented in Table 4. The Poisson rate was calculated for each of the 204 months and an average of these estimates taken resulting in an estimated rate of SIDS of 2.4 per month, while the proportion of non-zeros is between 86% and 91%. Based on their theoretical formulation, both the ZIP and Hurdle methods model a zero/non-zero partition in the population. Such a partition is confirmed by the best fit FM model (Table 4), with estimated means of 0 and 2.44 per month for component 1 and component 2, respectively (Table 5). Given the extra hierarchical structure with component $j$ ($j=1,2$) modelled in the more general FM framework, Table 4 shows, that conditional on being in component 1, the
predicted number of SIDS is dependent only on the number of infants at risk \((NAR)\), and not significantly affected by any climatic factors.

6. Discussion

Mixture models performed well on the Canterbury SIDS-climate series, with the ZIP model best describing the SIDS-climate relationship. These methods highlighted differences in climatic dependencies between months of no SIDS and some SIDS. The probability of SIDS occurring increases as a positive temperature deviance \((\text{PosTemp})\) increases. This positive temperature deviance \((\text{PosTemp})\) is a contrived variable measuring extreme changes in temperature, compared to the past fortnightly average. A month of low \(\text{PosTemp}\) implies that the temperature has remained reasonably constant (on average, not more than a \(1\) °C shift), whilst a month with high \(\text{PosTemp}\) indicates there have been more extreme shifts away from the average temperature. This could imply overheating of an infant is a significant risk factor, in that, when it is cold parents will put more blankets on their infant, but often not remove those blankets again when a warmer day occurs. Overheating, or hyperthermia, has been shown to lead to increased SIDS risk (Wells, 1997).
Fig. 2. Observed (–) and predicted (■) number of SIDS per month for each model.

Table 5
Estimated mean values and proportions on months with zero and non-zero SIDS counts

<table>
<thead>
<tr>
<th></th>
<th>FM</th>
<th>ZIP</th>
<th>Hurdle</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Comp 1</td>
<td>Comp 2</td>
<td></td>
</tr>
<tr>
<td>Estimated mean (\hat{\mu})</td>
<td>0</td>
<td>2.44</td>
<td>2.44</td>
</tr>
<tr>
<td>Proportion of zeros ((1 - \hat{p}))</td>
<td>0.09</td>
<td></td>
<td>0.10</td>
</tr>
<tr>
<td>Proportion of non-zeros (\hat{p})</td>
<td>0.91</td>
<td>0.90</td>
<td>0.86</td>
</tr>
</tbody>
</table>

In more traditional analyses to date, temperature has been used as a surrogate for seasonality (Schluter et al., 1998), yet in all three methods presented here the seasonal component is best described by the sinusoid functional.

The non-zero component of all the models contains an autoregressive term \(Y_{t-1}\). This, and the systematic lack of fit at extreme SIDS counts, may indicate some further latent variable, whether climatically, environmentally or physiologically based, which still needs further investigation.
Acknowledgements

The authors would like to thank the referees and the editor for their many helpful comments. They would also like to thank Robin Turner for all her help on putting this paper together.

Appendix A. Estimation algorithms

A.1. Finite mixture model (Wang et al., 1998)

Step 0: Specify initial values $\alpha^{(0)}$ and $\beta^{(0)}$ and two tolerances $\varepsilon$ and $\varepsilon_0$.

Step 1: (E-Step) Compute

$$
\tilde{Z}_{ij}(\alpha^{(0)}, \beta^{(0)}) = \frac{p_j(x_i, \beta_j^{(0)}) \text{Po}(y_i | z_i, \alpha_j^{(0)})}{\sum_{q=1}^{k} p_q(x_i, \beta_q^{(0)}) \text{Po}(y_i | z_i, \alpha_q^{(0)})}, \quad i = 1, \ldots, n; \quad j = 1, \ldots, k, \quad (A.1)
$$

where

$$
p_j(x_i, \beta) = \frac{\exp(x_i; \beta_j)}{1 + \sum_{q=1}^{k-1} \exp(x_i; \beta_q)} \quad \text{for} \quad j = 1, \ldots, k - 1,
$$

$$
p_k(x_i, \beta_j) = 1 - \sum_{q=1}^{k-1} p_q(x_i, \beta_q)
$$

and

$$
\text{Po}(y_i | z_i, x_j) = \frac{\exp(z_i; x_j)^{y_i} \exp(-\exp(z_i; x_j))}{y_i!}.
$$

To avoid overflow in the calculation of $\tilde{Z}_{ij}$, both the numerator and denominator of (A.1) are divided by the largest term in the sum in the denominator.

Step 2: (M-Step) Find values of $\hat{\alpha}$ and $\hat{\beta}$ to solve (A.2) and (A.3) using the quasi-Newton algorithm (Nash, 1990).

$$
\sum_{i=1}^{n} \tilde{Z}_i \frac{\partial}{\partial \alpha} [\log(\text{Po}(y_i | z_i, x_j))] |_{\hat{z}_j} = 0, \quad j = 1, \ldots, k, \quad (A.2)
$$

$$
\sum_{i=1}^{n} \tilde{Z}_i \frac{\partial}{\partial \beta} [\log(p_j(x_i, \beta_j))] |_{\hat{\beta}_j} = 0, \quad j = 1, \ldots, k - 1, \quad (A.3)
$$

where $\hat{Z}_i \equiv \hat{Z}_i(\alpha^{(0)}, \beta^{(0)}) = (\hat{Z}_{i1}, \ldots, \hat{Z}_{ik})^\top; \quad i = 1, \ldots, n$.

Step 3: (a) If at least one of 1-3 below is true, set $\alpha^{(0)} = \hat{\alpha}$ and $\beta^{(0)} = \hat{\beta}$ and go to Step 1; otherwise go to (b).

1. $\|\hat{\alpha} - \alpha^{(0)}\| \geq \varepsilon$
2. $\|\hat{\beta} - \beta^{(0)}\| \geq \varepsilon$
3. $|LL(\hat{\alpha}, \hat{\beta}) - LL(\alpha^{(0)}, \beta^{(0)})| \geq \varepsilon_0$

(b) Maximise the observed likelihood function $LL(\alpha, \beta)$ using the quasi-Newton algorithm with $\hat{\alpha}$ and $\hat{\beta}$ as initial values. Then, stop.
A.2. Zero-inflated Poisson model (Lambert, 1992)

Step 0: Specify initial values \( x^{(0)} \) and \( \beta^{(0)} \).

Step 1: (E-Step) Compute
\[
Z^{(0)}_i(\chi^{(0)}, \beta^{(0)}) = \begin{cases} 
1 + \exp(-x_i\beta^{(0)} - \exp(z_i\chi^{(0)})) & \text{if } y_i = 0, \\
0 & \text{if } y_i > 0.
\end{cases}
\] (A.4)

Step 2: (M-Step) Find \( \hat{\chi} \) by maximising \( LL_c(\chi) \), and \( \hat{\beta} \) by maximising \( LL_c(\beta) \) given in (A.5) and (A.6).
\[
LL_c(\chi) = \sum_{i=1}^{n} (1 - Z_i)(y_i z_i \chi - \exp(z_i \chi)),
\] (A.5)
\[
LL_c(\beta) = \sum_{y_i=0} Z_i x_i \beta - \sum_{y_i=0} Z_i \log(1 + \exp(x_i \beta)) \]
\[- \sum_{i=1}^{n} (1 - Z_i) \log(1 + \exp(x_i \beta)).
\] (A.6)

Step 3: Set \( x^{(0)} = \hat{x} \) and \( \beta^{(0)} = \hat{\beta} \) and continue from Step 1 until convergence occurs, as given by Step 3(a) in the Finite Mixture model estimation algorithm above, for example.

References


