

# Comparison of Hierarchical Bayesian Models for Overdispersed Count Data using DIC and Bayes' Factors

Russell B. Millar

Department of Statistics, University of Auckland, Private Bag 92019, Auckland, New Zealand  
*email:* r.millar@auckland.ac.nz

**SUMMARY.** When replicate count data are overdispersed, it is common practice to incorporate this extra-Poisson variability by including latent parameters at the observation level. For example, the negative binomial and Poisson-lognormal (PLN) models are obtained by using gamma and lognormal latent parameters, respectively. Several recent publications have employed the deviance information criterion (DIC) to choose between these two models, with the deviance defined using the Poisson likelihood that is obtained from conditioning on these latent parameters. The results herein show that this use of DIC is inappropriate. Instead, DIC was seen to perform well if calculated using likelihood that was marginalized at the group level by integrating out the observation-level latent parameters. This group-level marginalization is explicit in the case of the negative binomial, but requires numerical integration for the PLN model. Similarly, DIC performed well to judge whether zero inflation was required when calculated using the group-marginalized form of the zero-inflated likelihood. In the context of comparing multilevel hierarchical models, the top-level DIC was obtained using likelihood that was further marginalized by additional integration over the group-level latent parameters, and the marginal densities of the models were calculated for the purpose of providing Bayes' factors. The computational viability and interpretability of these different measures is considered.

**KEY WORDS:** Bayes' factors; Count data; DIC; Hierarchical model; Marginal likelihood; Negative binomial; Overdispersion; Poisson gamma; Poisson lognormal; **WinBUGS**; Zero inflation.

## 1. Introduction

The ease of fitting hierarchical models using Markov chain Monte Carlo (MCMC) methods has facilitated numerous Bayesian analyses of overdispersed count data across a wide variety of disciplines. For example, specification of gamma or lognormal latent parameters at the observation level corresponds to a negative binomial or Poisson-lognormal (PLN) model, respectively. Such models have been applied to machinery failure rates (George, Makov, and Smith, 1993), post-surgery mortality rates (Albert, 1999), species abundance (Martin et al., 2005; Golicher et al., 2006), hospital discharges (Van Den Heede et al., 2006), and traffic accident data (Miranda-Moreno, Labbe, and Fu, 2007) to list but a few. More generally, the model could include additional levels of hierarchy if, for example, the data are grouped (e.g., Dagne, 2004; Kuhnert et al., 2005; Fahrmeir and Echavarría, 2006).

When count data contain an excess of zeros then it may be necessary to employ a zero-inflated model whereby the probability of a zero count is explicitly inflated using a separate submodel for the zeros. This approach is particularly meaningful if there is reason to believe that the data are zero inflated due to some underlying processes. For example, in the modeling of abundance data, zero inflation could arise due to detection failure or habitat unsuitability (Martin et al., 2005). However, it has been argued that the negative binomial, and PLN models are usually sufficiently flexible to be able to pre-

dict large excesses of zeros (relative to the Poisson) without recourse to explicit zero inflation (Warton, 2005).

Here the focus is on comparison of the Poisson, negative binomial, and PLN models, and their zero-inflated extensions. It was of particular interest to investigate whether there was any substantive difference between the negative binomial and PLN models because of the implication for modeling of overdispersed multivariate count data. The PLN can easily and intuitively be extended to the multivariate case, simply by using Poisson-multivariate lognormal models (e.g., Bailey and Hewson, 2004). In contrast, multivariate extensions of the negative binomial are much more restrictive, with the most common forms permitting only nonnegative correlation (Kocherlakota and Kocherlakota, 1992), notwithstanding that Lee (1999) recently developed a more general bivariate negative binomial using copula models (Genest, 1987). That is, the option of choosing the better of the negative binomial and PLN models does not readily exist in the multivariate case, and so it would be reassuring to know that, at least in the univariate case, there was no substantive loss of model performance from using the PLN.

Model comparison is commonly based on the deviance information criterion (DIC) of Spiegelhalter et al. (2002) because of its widespread use and ease of application. However, it is demonstrated that the DIC is a potentially dangerous tool in the present context. For example, when fitting to simulated

PLN data, naive use of the DIC resulted in the negative binomial model always being preferred to the PLN. In contrast, a partially marginalized DIC was found to perform well.

There are numerous alternatives to the DIC, and diverse coverage of many of these is provided by the discussants of Spiegelhalter et al. (2002) and Celeux et al. (2006). The most familiar alternative is Bayes' factors, that is, the ratio of the marginal densities of the data under two competing models. However, Bayes' factors have two major drawbacks. Firstly, the marginal density corresponds to a high-dimensional integral over the parameter space. This integration is often not numerically feasible and other methods of evaluation are usually required (e.g., Gelman and Meng, 1998; Han and Carlin, 2001; Sinharay and Stern, 2005). Some approaches, such as the product-space approach, reversible-jump MCMC, and path sampling, are able to evaluate Bayes' factors without calculation of the marginal density, but nonetheless they often have extreme computational and programming requirements, with no general guarantees of success. Secondly, Bayes' factors can be sensitive to prior specification, notwithstanding that this sensitivity is primarily with respect to the prior specification of the parameters that differ between the two models (Kass and Raftery, 1995).

In the context of multilevel hierarchical models for count data, Section 2 presents the model formulations of six competing models, and three flavors of DIC. The conditional DIC uses likelihood conditioned on parameters at the observation level of the hierarchy, whereas the group- and top-level DICs use partially marginalized likelihoods. This section also introduces the Chib method for evaluation of the marginal density, from which Bayes' factors are obtained. Section 3 presents a case study that models fish abundance for three species that vary from solitary to highly aggregated. The case study suggests good behavior of the marginalized forms of DIC, and these provided conclusions similar to that of the Bayes' factors. However, use of the conditional DIC was misleading. These conclusions were verified by exploring the performance of the conditional and group-marginalized DICs using simulation (Section 4). The relative performance of the negative binomial and PLN models is addressed in Section 5, and explanation of the relative behavior of Bayes' factors and the three forms of DIC is provided.

## 2. Methods

### 2.1 General Model Formulation

The general model herein is specified for repeat-measurement data because it is motivated by the spatially grouped count data from the case study (Section 3). Let  $y_{ij}$  denote observation  $j$ ,  $j = 1, \dots, n_i$  in group  $i$ ,  $i = 1, \dots, M$ . Six different models for the within-group distribution of the replicate counts are considered, the Poisson (P), Poisson-gamma (PGA), PLN, and their zero-inflated extensions, ZIP, ZIPGA, and ZIPLN, respectively. The PGA and ZIPGA are (marginally) equivalent to the negative binomial and zero-inflated negative binomial models, respectively.

The P, PGA, and PLN models can be expressed using the generic notation

$$a, \sigma, \gamma \sim \pi(a, \sigma, \gamma), \quad (1)$$

$$\mu_i \sim h(\mu_i | a, \sigma), \quad (2)$$

$$\lambda_{ij} \sim g(\lambda_{ij} | \mu_i, \gamma), \quad (3)$$

$$y_{ij} \sim f(y_{ij} | \lambda_{ij}), \quad (4)$$

where  $f(y_{ij} | \lambda_{ij})$  is the Poisson( $\lambda_{ij}$ ) density. Parameter  $\mu_i$  is the expected count in group  $i$ , and is distributed as

$$\log \mu_i \sim N(a, \sigma^2) \quad i = 1, \dots, m.$$

Parameter  $\gamma \in \{\alpha, \tau\}$  is used generically in equations (1) and (3). For the Poisson model,  $\gamma$  is null and  $\lambda_{ij} = \mu_i$ . For the PGA model,  $\gamma = \alpha$ , and denotes the negative binomial overdispersion parameter whereby

$$\begin{aligned} \log \lambda_{ij} &= \log \mu_i + \log \epsilon_{ij}, \\ \epsilon_{ij} &\sim \Gamma(\alpha, \alpha). \end{aligned} \quad (5)$$

For the PLN model,  $\gamma = \tau$ , where

$$\begin{aligned} \log \lambda_{ij} &= \log \mu_i + \log \epsilon_{ij}, \\ \epsilon_{ij} &\sim LN\left(\frac{-\tau^2}{2}, \tau^2\right). \end{aligned} \quad (6)$$

Note that  $E[\epsilon_{ij}]$  and  $E[\epsilon_{ij}]$  are both equal to unity, and consequently,  $E[\lambda_{ij}] = \mu_i$  under all three models.

In the three zero-inflated models, the probability of zero inflation was assumed constant across all observations. Zero inflation can be implemented using latent Bernoulli parameters,  $B_{ij}$ , distributed Bernoulli( $p$ ). Then, conditional on  $\lambda_{ij}$  and  $B_{ij}$ ,  $y_{ij}$  is distributed Poisson( $(1 - B_{ij})\lambda_{ij}$ ). That is,  $y_{ij}$  is distributed Poisson( $\lambda_{ij}$ ) if  $B_{ij} = 0$  and is identically zero if  $B_{ij} = 1$ . However, it is well known that DIC is invalid under this parameterization of a finite mixture model (e.g., Lawson and Clark, 2002) because the "plug-in" estimates of  $\lambda_{ij}$  and  $B_{ij}$  used in the calculation of DIC in equation (9) are inappropriate. Instead, it is common practice to specify a zero-inflated density corresponding to implicit marginalization over the latent Bernoulli parameters. In the present context, the zero-inflated models replace equations (1) and (4) with

$$\begin{aligned} a, \sigma, \gamma, p &\sim \pi(a, \sigma, \gamma, p), \\ y_{ij} &\sim f(y_{ij} | \lambda_{ij}, p), \end{aligned} \quad (7)$$

where  $f(y | \lambda_{ij}, p)$  is the density of the zero-inflated Poisson. That is,

$$f(y; \lambda, p) = \begin{cases} p + (1 - p)e^{-\lambda} & y = 0 \\ (1 - p)\frac{e^{-\lambda}\lambda^y}{y!} & y = 1, 2, 3, \dots \end{cases} \quad (8)$$

In what follows, calculation of DICs and marginal densities will be presented for the zero-inflated models, with the implicit understanding that the models without zero inflation are obtained by removing  $p$  from the notation and formulae.

### 2.2 Deviance Information Criterion

The DIC of Spiegelhalter et al. (2002) has the form

$$\text{DIC} = 2\overline{D(\boldsymbol{\theta})} - D(\overline{\boldsymbol{\theta}}), \quad (9)$$

where  $D$  denotes deviance and the overlines denote posterior expectation. Here, the deviance is conditional on parameter

vector  $\theta$ , which is variously termed the “parameters of interest” or “parameters in focus” by Spiegelhalter et al. (2002). The value  $D(\bar{\theta})$  is the “plug-in” estimate of model deviance, and it can alternatively be calculated using other “plug-in” estimates of  $\theta$ , such as the posterior mode. Here, the standard form of DIC in equation (9) is used throughout.

The DIC has many weaknesses, as discussed by Spiegelhalter et al. (2002) and comments therein, and by Celeux et al. (2006) and the accompanying comments by Carlin (2006), Meng and Vaida (2006), and Plummer (2006). These weaknesses include dependence on the level of parameter specification upon which the model likelihood is conditioned (i.e., “parameter focus”). The DIC also lacks invariance to reparameterization, because of the use of the posterior mean  $\bar{\theta}$  in equation (9).

Spiegelhalter et al. (2002, p. 613) write “the parameters in the focus of a model should ideally depend on the purpose of the investigation, although in practice it is likely that the focus may be chosen on computational grounds as providing likelihoods that are available in closed form.” In the present context, it is computationally convenient to take  $\theta$  to be the parameters in the lowest (i.e., observation) level of the hierarchy, as specified by equation (7). This will be referred to as the conditional model and the corresponding deviance is

$$D(\lambda, p) = -2 \sum_i \sum_j \log f(y_{ij} | \lambda_{ij}, p), \quad (10)$$

where  $\lambda$  denotes the collection of  $\lambda_{ij}$  parameters. Note that the saturated model term is not required in this definition of deviance because it is eliminated in the calculation of the DIC in equation (9). DIC calculated using equation (10) will be called the conditional DIC.

However, in many situations (including the case study in Section 3), it will be the marginal distribution of the replicates at the group level that will be of interest. This group-marginalized distribution has density

$$s(y_{ij} | \mu_i, \gamma, p) = \int_0^\infty f(y_{ij} | \lambda_{ij}, p) g(\lambda_{ij} | \mu_i, \gamma) d\lambda_{ij}, \quad (11)$$

and the corresponding group-level deviance  $D(\mu, \gamma, p)$  is then

$$D(\mu, \gamma, p) = -2 \sum_i \sum_j \log s(y_{ij} | \mu_i, \gamma, p), \quad (12)$$

where  $\mu = (\mu_1, \dots, \mu_M)$ . The DIC calculated using equation (12) will be called the group-marginalized DIC.

In the case of the Poisson and zero-inflated Poisson models,  $\lambda_{ij} = \mu_i$  and the conditional and group-marginalized densities are identical. The group-marginalized forms of the PGA and zero-inflated PGA densities are the negative binomial and zero-inflated negative binomial, respectively. For the PLN and zero-inflated PLN, the integral in equation (11) does not reduce to a closed form.

It is also possible to calculate a top-level marginalized DIC using the deviance  $D(a, \sigma, \gamma, p)$  obtained from additional marginalization over the group level parameters  $\mu_i$  in equation (2). That is, by using likelihood given by the marginalized density

$$\begin{aligned} t(y_{ij} | a, \sigma, \gamma, p) &= \int_0^\infty s(y_{ij} | \mu_i, \gamma, p) h(\mu_i | a, \sigma) d\mu_i \\ &= \int_0^\infty \int_0^\infty f(y_{ij} | \lambda_{ij}, p) g(\lambda_{ij} | \mu_i, \gamma) \\ &\quad \times h(\mu_i | a, \sigma) d\lambda_{ij} d\mu_i. \end{aligned} \quad (13)$$

The top-level marginalized DIC is not relevant to the case study (see Section 3.1), but is included for completeness.

### 2.3 Marginal Density and Bayes' Factors

The marginal density, evaluated at the observed data  $\mathbf{y}$ , can be expressed as

$$m(\mathbf{y}) = \int t(\mathbf{y} | \theta) \pi(\theta) d\theta, \quad (14)$$

where  $\theta = (a, \sigma, \gamma, p)$  and  $t(\mathbf{y} | \theta)$  is the density of  $\mathbf{y}$  given  $\theta$ , obtained from equation (13).

Chib's method for calculating marginal densities (Chib, 1995) uses an identity that is obtained by a rearrangement of Bayes' formula,

$$m(\mathbf{y}) = \frac{t(\mathbf{y} | \theta) \pi(\theta)}{\pi(\theta | \mathbf{y})}. \quad (15)$$

This identity holds for any value  $\theta^*$  in the parameter space; however, evaluation of  $m(\mathbf{y})$  is more efficient if  $\theta^*$  is chosen to be a point with high posterior density. Given  $\theta^*$ , calculation of  $\pi(\theta^*)$  is immediate. However, evaluation of  $t(\mathbf{y} | \theta^*)$  and  $\pi(\theta^* | \mathbf{y})$  is computationally intensive.

## 3. Case Study

### 3.1 Background

A subset of the fish abundance data from Anderson and Millar (2004) was used. Briefly, these comprise  $M = 16$  sites at which  $n = 10$  transects of length 25 m were swum over subtidal rocky reefs using scuba, for a total of  $N = 160$  observations. Forty-six different fish species were recorded but here the focus is on three of the more prevalent species, spotty (*Notolabrus celidotus*), goatfish (*Upeneichthys lineatus*), and two-spot demoiselle (*Chromis dispilus*). These three species were chosen to span a range of schooling behaviors (Figure 1). Two-spot demoiselles are often found in large aggregations, goatfish often congregate in small groups, and spotty tend to be more solitary though females and juveniles can sometimes be found in small groups (Francis, 1996).

The data modeled here are part of a larger experiment to establish a long-term times series of fish community data. Prediction at the level of individually swum transects is not of interest. Indeed, at each site, the transects were haphazardly chosen by the divers according to the vagaries of the underwater terrain and benthic cover on the reefs. Prediction of fish numbers at other sites is also not relevant to this study. These considerations suggest that it is the group-marginalized form of DIC that is most appropriate for model comparison.

### 3.2 Choice of Priors

All six models include parameters  $a$  and  $\sigma$ . The PGA and ZIPGA also include overdispersion parameter  $\alpha$ , while the PLN and ZIPLN include the lognormal parameter  $\tau$ , and the

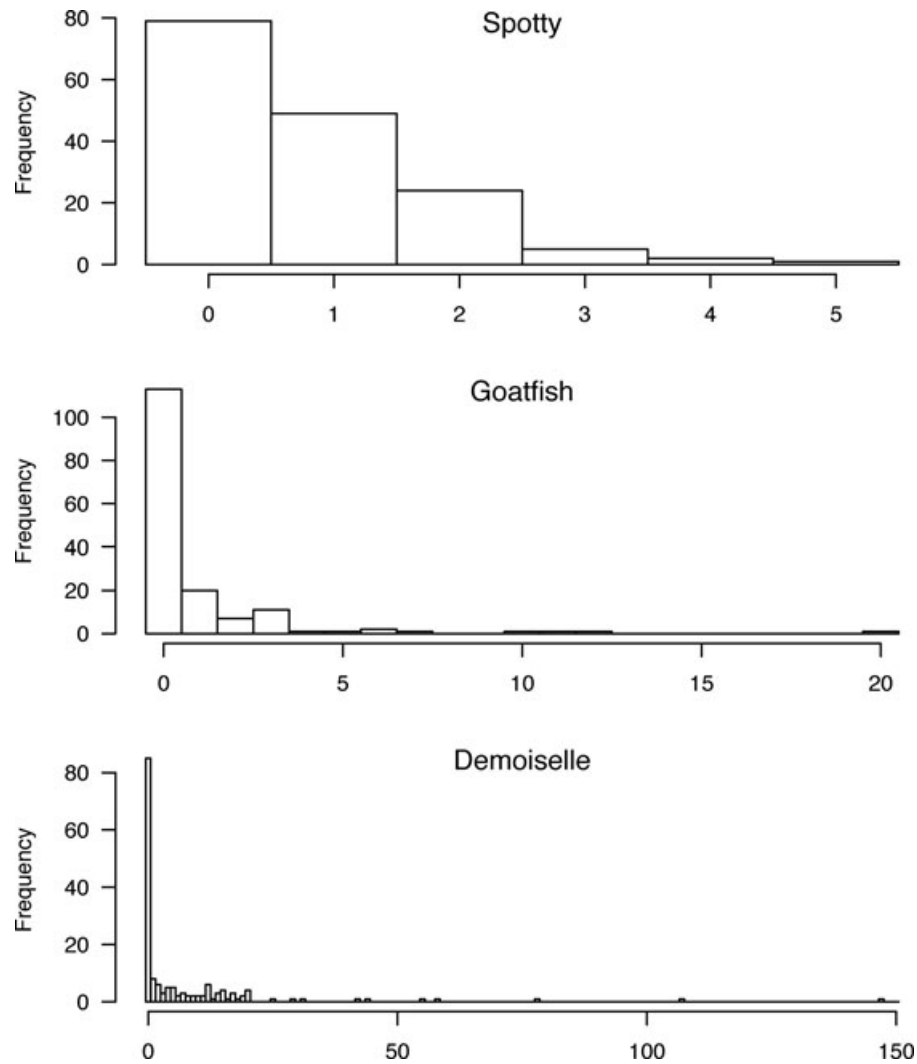


Figure 1. Histograms of fish abundance data from 160 transects swum using SCUBA.

three zero-inflated models include the zero-inflation probability parameter  $p$ . These parameters were assumed to be a priori independent.

The priors on  $a$  and  $p$  were

$$a \sim N(0, 10^6), \quad (16)$$

$$p \sim \text{beta}(0.5, 0.5). \quad (17)$$

A half standard-Cauchy prior was chosen for  $\sigma$  and  $\tau$  (Gelman, 2006). The prior on the negative binomial overdispersion parameter,  $\alpha$ , was specified such that the  $\Gamma(\alpha, \alpha)$  distribution in equation (5) would have the same a priori variance as the  $LN(-\tau^2/2, \tau^2)$  distribution in equation (6) (Echavarría, 2004). That is,

$$\alpha = 1/(\exp(\delta^2) - 1) \quad (18)$$

where  $\delta$  has a standard Cauchy distribution.

### 3.3 Implementation

**3.3.1 Sampling from the posterior.** The posterior density was sampled using the **OpenBUGS** package version 3.0.3 (freely available from <http://mathstat.helsinki.fi/openbugs/>).

The Cauchy is not a standard density in **OpenBUGS** 3.0.3 and hence the half-Cauchy priors on  $\sigma$  and  $\tau$  were implemented using the “zeros trick,” documented in the **OpenBUGS** online help. The prior on  $\alpha$  was also implemented using the zeros trick. Similarly, the three zero-inflated models were implemented using a form of the “zeros trick” for the specification of nonstandard likelihoods.

Trial runs of the models were used to check MCMC convergence using the Brooks-Gelman-Rubin (Gelman et al., 2003) diagnostic applied to four parallel chains, and it was determined that a burn-in of 10,000 was adequate. After discarding the burn-in, a further 50,000 values were sampled and thinned by a factor of 10 so that 5000 samples were retained to disc.

**3.3.2 Calculation of DIC.** The **OpenBUGS** code used for sampling of the six models was parameterized so that  $\lambda$  and  $p$  were specified explicitly as stochastic nodes rather than as logical functions of parent stochastic nodes. This ensured that  $(\lambda, p)$  was the parameter specification used by **OpenBUGS** for calculation of model deviance (i.e., equation 10) and hence of the conditional DIC.

The marginalized DICs were calculated within **R**, using the **BRugs** package to interface **OpenBUGS** with the **R** language. In the case of the PLN and ZIPLN models, the group-marginalized density in equation (11) is not available in closed form. Instead, the **integrate** function was used to numerically evaluate this one-dimensional integral. The top-level marginalized density,  $t(\mathbf{y} | \theta)$ , of the P, PGA, ZIP, and ZIPGA models required numerical integration with respect to  $\mu$ , while the PLN and ZIPLN models required a two-dimensional integration with respect to both  $\mu$  and  $\lambda$ . The **integrate** function returns an estimate of the relative error of the numerical integration. This relative error was required to be less than 1%, corresponding to an absolute error of less than 0.01 on the log scale.

**3.3.3 Calculation of the marginal density.** The parameter value  $\theta^*$  was taken to be the posterior mean of  $(a, \sigma, \gamma, p)$ . As noted above, calculation of  $t(\mathbf{y} | \theta^*)$  was performed using the **integrate** function. Evaluation of  $\pi(\theta^* | \mathbf{y})$  was performed using kernel density estimation via the **kde** function within the **ks** package.

The calculation of  $m(\mathbf{y})$  was repeated for several alternative values of  $\theta^*$ , to informally gauge the accuracy of estimation. This indicated that  $\log m(\mathbf{y})$  had estimation error less than 0.1 for the spotty and goatfish species, and less than 0.2 for demoiselle.

The **R** and **WinBUGS** codes for calculation of the three forms of DIC and  $m(\mathbf{y})$  are given in Web Appendix B. The posterior means of the top-level model parameters  $(a, \sigma, \gamma, p)$  for the three species and six models are provided in Web Table 1 in Web Appendix A.

#### 3.4 Model Comparison Using DIC and Bayes' Factors

Table 1 shows the conditional DIC ( $DIC_C$ ), group-marginalized DIC ( $DIC_G$ ), top-level marginalized DIC ( $DIC_T$ ), and  $-2 \log m(\mathbf{y})$  from fits of the six models to the three fish species. For these measures, a popular rule of thumb for model comparison (e.g., Kass and Raftery, 1995; Spiegelhalter et al., 2002), is that a difference of 2 or less is considered negligible, a difference between 2 and 6 provides positive support for the model with the lower value of the measure, and a difference in excess of 6 constitutes strong support. The log-Bayes' factor for comparison of two models is simply one half of the difference in their values of  $-2 \log m(\mathbf{y})$ , in favor of the model with smaller value of  $-2 \log m(\mathbf{y})$ .

In general, DIC and Bayes' factors are assessing different features of the model and cannot be expected to arrive at the same conclusion (Spiegelhalter et al., 2002). Nonetheless, in this case the six models differ by at most one or two parameters, and these parameters have moderately constrained reference priors, so it would be a concern if DIC and  $-2 \log m(\mathbf{y})$  suggest vastly different model preferences. (This point is taken up further in Section 5.)

The two marginalized DICs and  $-2 \log m(\mathbf{y})$  provide a reasonably consistent set of model preferences. For spotty, the least overdispersed of the three species,  $DIC_G$  and  $DIC_T$  are ambivalent to choice of the six models, while  $-2 \log m(\mathbf{y})$  suggests modest preference for the P and PGA models. For goatfish, the marginalized DICs and  $-2 \log m(\mathbf{y})$  all rank PGA, ZIPGA, and ZIPLN as the preferred three models, though not in the same order. For the highly overdispersed demoiselle,  $DIC_G$  and  $-2 \log m(\mathbf{y})$  both prefer ZIPGA, ZIPLN, and PGA in that order.  $DIC_T$  is much the same, except that ZIPGA and ZIPLN are preferred about equally well.

Conclusions from the conditional DIC are largely incompatible with those from the marginalized DICs and  $-2 \log m(\mathbf{y})$ . In particular, for all three species, the PGA model always had the lowest value of  $DIC_C$ . Moreover, zero inflation of the PGA and PLN models always resulted in an increase in  $DIC_C$ , and the increase in  $DIC_C$  due to zero inflation was more pronounced for the two most overdispersed species, goatfish and demoiselle.

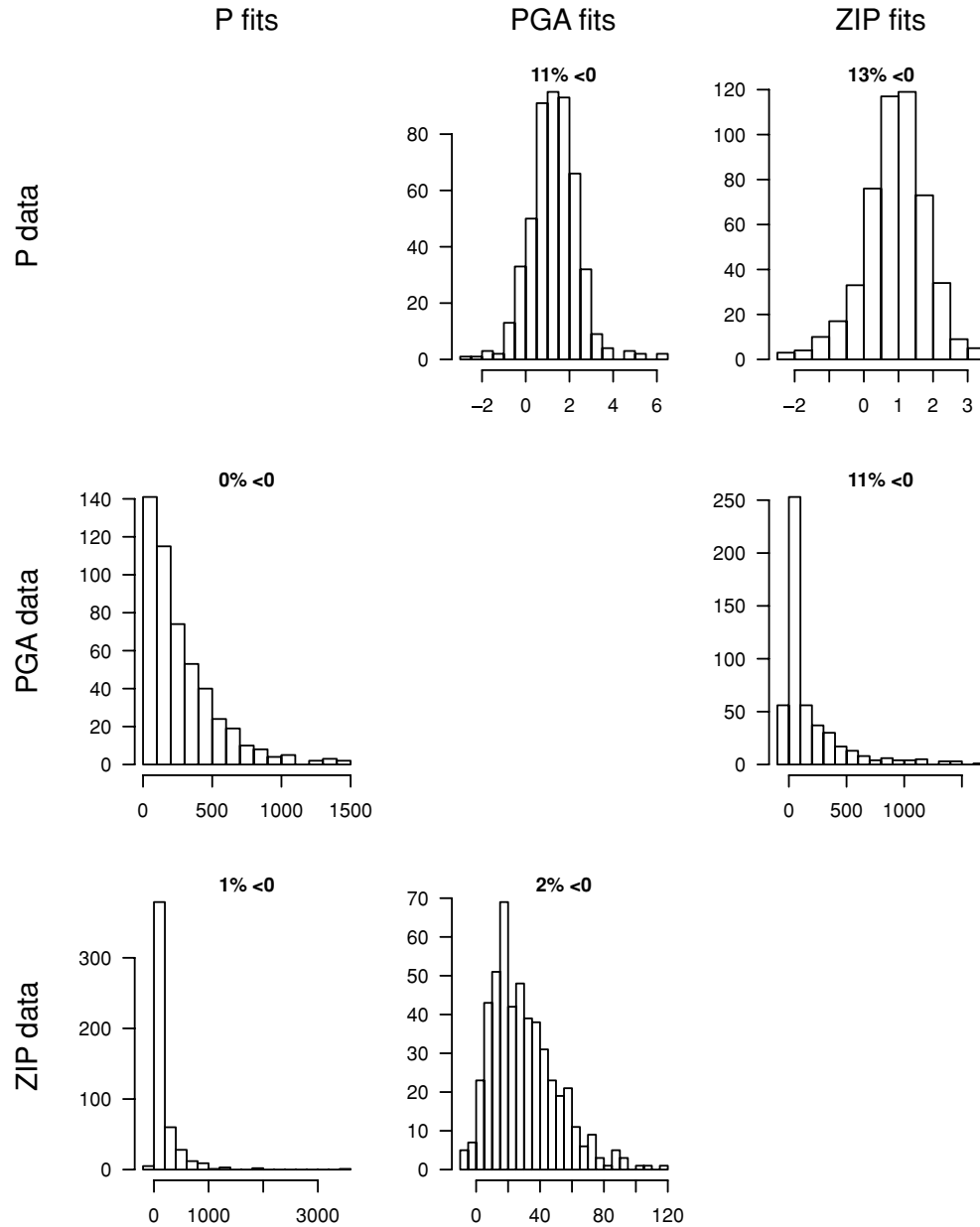
#### 4. Simulations

For meaningful examination of the behavior of DIC, the simulated data must be generated from a plausible model (Sinhara and Stern, 2003). Pure-blooded Bayesians would have used a plausible a priori model and simulated data would be from the marginal density  $m(\mathbf{y})$  (Box, 1980). However, the posited a priori models used herein were not (and never were intended to be) plausible because they made use of reference priors. For example, the prior on  $a$  (equation 16) would give rise to numerical overflow of  $\mu_i$  due to large values of  $\log \mu_i$ . (This parameter is common to all models, and hence Bayes' factors remain valid, despite the implausibility of the prior model.) Instead, data were simulated from the posterior-predictive distribution, that is, using the posterior distribution of  $(\mu, \gamma, p)$ .

Table 1

*DIC statistics and marginal densities.  $DIC_C$ ,  $DIC_G$ ,  $DIC_T$ , and  $-2l(\mathbf{y})$  denote the conditional DIC, group-marginalized DIC, top-level marginalized DIC, and  $-2 \log m(\mathbf{y})$ , respectively. Smaller values suggest a better model. A difference of less than 2 is considered negligible.*

Model	Spotty				Goatfish				Demoiselle			
	$DIC_C$	$DIC_G$	$DIC_T$	$-2l(\mathbf{y})$	$DIC_C$	$DIC_G$	$DIC_T$	$-2l(\mathbf{y})$	$DIC_C$	$DIC_G$	$DIC_T$	$-2l(\mathbf{y})$
P	385.2	385.2	387.2	405.8	449.2	449.2	480.1	496.4	2521.0	2521.0	2589.6	2614.5
PGA	379.7	385.8	387.0	405.5	265.7	346.6	364.5	381.1	481.5	734.3	764.7	784.2
PLN	384.5	385.7	387.0	408.3	280.2	351.4	369.4	386.7	493.7	743.1	773.9	791.7
ZIP	385.6	385.6	387.0	407.4	352.7	352.7	382.4	400.6	1469.8	1469.8	1541.3	1562.5
ZIPGA	381.1	386.1	387.1	408.1	278.4	344.0	363.4	381.4	507.3	728.0	762.8	782.1
ZIPLN	384.7	385.7	387.1	410.4	293.0	344.8	365.3	384.7	540.0	730.7	762.7	784.0

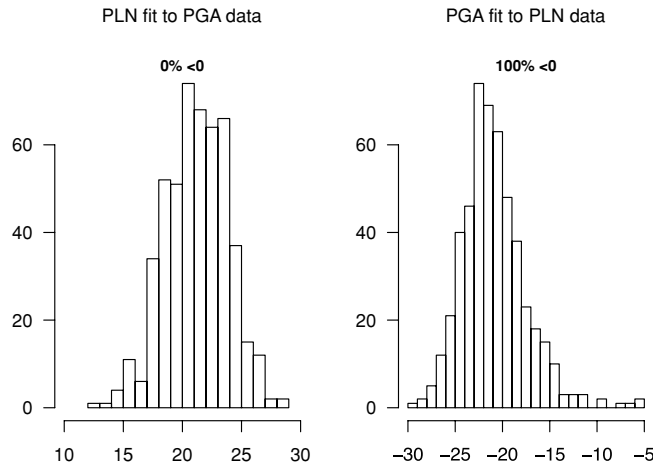


**Figure 2.** The group-marginalized DIC under the alternate model minus the group-marginalized DIC under the true model, from 500 simulated data sets of each type. Row 1, Poisson data; row 2, PGA (i.e., negative binomial) data; and row 3, zero-inflated Poisson data.

Simulated data were generated from the posterior-predictive fits of the P, PGA, PLN, and ZIP models to goat-fish data (Web Table 1). Five hundred simulated datasets were produced from each of these four models. Group-marginalized DIC was computed for P, ZIP, and PGA fits to the collection of 1500 datasets generated under the P, ZIP, and PGA models. It was not computationally viable to include the PLN model in this comparison due to the numerical integration required to calculate the marginalized likelihood, and also not possible to consider calculation of the top-level marginalized DIC. Conditional DIC was computed for PGA and PLN fits

to the collection of 1000 datasets generated under the PGA and PLN models.

The differences in  $DIC_G$  (between alternate model fit and true model fit) have mode close to unity for the PGA and ZIP fits to the Poisson data, reflecting the small penalty for inclusion of an unnecessary model parameter (Figure 2). The fits of the PGA and ZIP models had smaller  $DIC_G$  in 11% and 13% (respectively) of the fits to the simulated Poisson datasets. The P and ZIP fits to PGA data resulted in some extremely large increases in  $DIC_G$ , notwithstanding that in 11% of the datasets the ZIP model had lower  $DIC_G$  than the



**Figure 3.** The conditional DIC under the alternate model minus the conditional DIC under the true model, from 500 simulated data sets of each type. Left plot, PLN model fitted to PGA (i.e., negative binomial) data; right plot, PGA model fitted to PLN data.

PGA model. P and PGA fits to the ZIP data had lower  $DIC_G$  than the ZIP fit for just 1% and 2% of the simulated ZIP datasets, respectively.

From the fits of the PGA and PLN models (to PGA and PLN data), it was seen that the PGA fit always had the lower conditional DIC, regardless of whether the data were PGA or PLN (Figure 3).

## 5. Discussion

The case study and simulations showed that conditional DIC was not valid for model comparison.  $DIC_C$  gave conflicting conclusions compared to Bayes' factors, and for the highly overdispersed demoiselle species the  $DIC_C$  of the PGA and PLN models was seen to increase by about 26 and 46, respectively, when these models were zero inflated using the addition of one extra parameter. The simulations always resulted in a lower  $DIC_C$  for the PGA model than the PLN, even when the data were simulated from the PLN model (Figure 3). This is a timely warning against the blanket use of the  $DIC$ —several recent works have used the conditional DIC for comparison of PGA (i.e., negative binomial) and PLN models, including Golicher et al. (2006), Van Den Heede et al. (2006), and Miranda-Moreno et al. (2007), and all three preferred the PGA model because it had smaller  $DIC_C$ .

Group-marginalized DIC appears to perform well. In the case study,  $DIC_G$  behaved in a readily interpretable manner, and was generally in agreement with Bayes' factors. In the fits of PGA and ZIP models to Poisson data, the simulations (Figure 2) showed that  $DIC_G$  included a small penalty for the additional one-parameter increase in model complexity. Computational requirements prevented the PLN from being included in this simulation, due to the intractable form of the group-marginalized likelihood (equation 11) that requires numerical integration. However, it was seen that the  $DIC_G$  of the PGA model (i.e., using the negative binomial likelihood) behaved well. Within the context of the case study, the

top-level marginalized DIC performed similarly to the group-marginalized DIC and Bayes' factor.

The warnings by Celeux et al. (2006) and others suggest that zero inflation could be problematic for the DIC because it is a form of mixture model. It was noted in Section 2.1 that DIC should not be used with mixture model likelihoods that are specified using latent Bernoulli parameters. Indeed, the **OpenBUGS** software disables the menu option for calculation of DIC under this form of mixture model specification. The latent Bernoulli parameter was avoided here by using the familiar (implicitly marginalized) zero-inflated Poisson density in equation (8). That is, in the zero-inflated models, the marginalized DICs are also implicitly marginalized over latent Bernoulli parameters.

In the context of assessing the need for zero inflation, an intuitive explanation for the erroneous behavior of the conditional DIC is that in the uninflated model the posterior distribution of each  $\lambda_{ij}$  will have been "fitted" to  $y_{ij}$ . Hence the mixing of the  $\text{Poisson}(\lambda_{ij})$  model with the zero distribution will result in larger deviance. It is not immediately clear why  $DIC_C$  also fails for comparison between the PGA and PLN models. This may arise as a result of the asymptotic rationale behind DIC (Spiegelhalter et al., 2002; Meng and Vaida, 2006). In the case study and simulations used here, there were 10 replicate observations per group. This appeared to be sufficient for group-marginalized DIC to perform well, but due care should be taken if working with smaller group sizes.

Overall, the PGA model appears to be preferred to the PLN model when overdispersion is present, as indicated by the values of the marginalized DICs and  $-2 \log m(\mathbf{y})$  for the goatfish and demoiselle species. The strength of this preference is much weaker in comparison of the ZIPGA and ZIPLN models. One explanation for this is that, relative to the negative binomial, the PLN model may have difficulty in modeling excess zeros in overdispersed count data, and hence may benefit more from zero inflation.

In general, model comparison using DIC and Bayes' factors cannot be expected to lead to the same conclusions. The marginal density,  $m(\mathbf{y})$ , quantifies how well the prior model has predicted the observed data, whereas DIC is designed to measure how well the posterior model might predict future observations. Nonetheless, here marginalized DICs and Bayes' factors gave similar conclusions, notwithstanding that Bayes' factors appeared to incur a greater penalty for zero inflation (Table 1). The similar behavior of  $DIC_G$ ,  $DIC_T$  and Bayes' factors may be due to the weakly constrained priors that were applied to the parameters that differ between the six models. In particular, the PGA and PLN models differ only between the gamma and lognormal distributions in equations (6) and (5), and the a priori variance of these two distributions was made equal through use of the variance-matching prior on parameter  $\alpha$  (equation 18). It is more usual for  $\alpha$  to be given a highly dispersed gamma distribution (e.g., Van Den-Heede et al., 2006), in which case Bayes' factors for comparison between the negative binomial and PLN would not be meaningful.

These results suggest that, with sufficient within-group replication, the group-marginalized DIC performs well. It is straightforward to calculate for comparison of the Poisson and PGA (i.e., negative binomial) models and their zero-inflated extensions. The PLN is more problematic due

to the intractable group-level likelihood, and here, it did not generally perform as well as the negative binomial. However, the numerical calculation of the group-marginalized likelihood was far less demanding, in both programming effort and computational requirements, than calculation of the top-level marginalized DIC or marginal density,  $m(\mathbf{y})$ . The top-level marginalized DIC,  $\text{DIC}_T$  was particularly onerous to compute for the PLN and ZIPLN models, due to the two-dimensional numerical integrations required to compute  $t(\mathbf{y} | \boldsymbol{\theta})$  where, in the ZIPLN model,  $\boldsymbol{\theta} = (a, \sigma, \tau, p)$ . These numerical integrations were required for each of the 5000 sampled values from  $\pi(\boldsymbol{\theta} | \mathbf{y})$ . Using the `integrate` function in the R language, it required several days of CPU time on a 2.4 GHz PC to obtain the top-level marginalized DICs given in Table 1.

## 6. Supplementary Materials

The Web Appendices referenced in Sections 3.3.3 and 4 contain Web Table 1, and the `OpenBUGS` code, R code, and data for calculating DICs and the marginal density for the PLN fit to the goatfish data. They are available under the Paper Information link at the *Biometrics* website <http://www.biometrics.tibs.org>.

## ACKNOWLEDGEMENTS

The article was much improved by pertinent comments from the associate editor and two anonymous referees. Thanks also to Prof Marti Anderson (Department of Statistics, University of Auckland) for permission to use the fish abundance data of the case study.

## REFERENCES

- Albert, J. H. (1999). Criticism of a hierarchical model using Bayes factors. *Statistics in Medicine* **18**, 287–305.
- Anderson, M. J. and Millar, R. B. (2004). Spatial variation and effects of habitat on temperate reef fish assemblages in northeastern New Zealand. *Journal of Experimental Marine Biology and Ecology* **305**, 191–221.
- Bailey, T. C. and Hewson, P. J. (2004). Simultaneous modelling of multiple traffic safety performance indicators by using a multivariate generalized linear mixed model. *Journal of the Royal Statistical Society, Series A* **167**, 501–517.
- Box, G. E. P. (1980). Sampling and Bayes' inference in scientific modelling and robustness. *Journal of the Royal Statistical Society, Series A* **143**, 383–430.
- Carlin, B. P. (2006). Comment on article by Celeux et al. *Bayesian Analysis* **1**, 675–676.
- Celeux, G., Forbes, F., Robert, C. P., and Titterton, D. M. (2006). Deviance information criteria for missing data models (with discussion). *Bayesian Analysis* **1**, 651–706.
- Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association* **90**, 1313–1321.
- Dagne, G. A. (2004). Hierarchical Bayesian analysis of correlated zero-inflated count data. *Biometrical Journal* **46**, 653–663.
- Echavarría, L. E. O. (2004). Semiparametric Bayesian count data models. Ph.D. Dissertation, Ludwig-Maximilians-University, Munich.
- Fahrmeir, L. and Echavarría, L. O. (2006). Structured additive regression for overdispersed and zero-inflated count data. *Applied Stochastic Models in Business and Industry* **22**, 351–369.
- Francis, M. (1996). *Coastal fishes of New Zealand*. Auckland, New Zealand: Reed Books.
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (Comment on article by Browne and Draper). *Bayesian Analysis* **1**, 515–534.
- Gelman, A. and Meng, X.-L. (1998). Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical Science* **13**, 163–185.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003). *Bayesian Data Analysis*, 2nd edition. New York: Chapman and Hall.
- Genest, C. (1987). Frank's family of bivariate distributions. *Biometrika* **74**, 549–556.
- George, E. I., Makov, U. E., and Smith, A. F. M. (1993). Conjugate likelihood distributions. *Scandinavian Journal of Statistics* **20**, 147–156.
- Golicher, D. J., O'Hara, R. B., Ruíz-Montoya, L., and Cayuela, L. (2006). Lifting a veil on diversity: A Bayesian approach to fitting relative-abundance models. *Ecological Applications* **16**, 201–212.
- Han, C. and Carlin, B. P. (2001). Markov chain Monte Carlo methods for computing Bayes factors: A comparative review. *Journal of the American Statistical Association* **96**, 1122–1132.
- Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association* **90**, 773–795.
- Kocherlakota, S. and Kocherlakota, K. (1992). *Bivariate Discrete Distributions*. New York: Marcel Dekker.
- Kuhnert, P. M., Martin, T. G., Mengersen, K., and Possingham, H. P. (2005). Assessing the impacts of grazing levels on bird density in woodland habitat: A Bayesian approach using expert opinion. *Environmetrics* **16**, 717–747.
- Lawson, A. and Clark, A. (2002). Comment on article by Spiegelhalter et al. *Journal of the Royal Statistical Society, Series B* **64**, 624–625.
- Lee, A. (1999). Modelling rugby league data via bivariate negative binomial regression. *Australian and New Zealand Journal of Statistics* **41**, 141–152.
- Martin, T. G., Wintle, B. A., Rhodes, J. R., Kuhnert, P. M., Field, S. A., Low-Choy, S. J., Tyre, A. J., and Possingham, H. P. (2005). Zero tolerance ecology: Improving ecological inference by modelling the source of zero observations. *Ecology Letters* **8**, 1235–1246.
- Meng X.-L. and Vaida, F. (2006). What's missing for DIC with missing data? (Comment on article by Celeux et al.). *Bayesian Analysis* **1**, 687–698.
- Miranda-Moreno, L. F., Labbe, A., and Fu, L. (2007). Bayesian multiple testing procedures for hotspot identification. *Accident Analysis and Prevention* **39**, 1192–1201.
- Plummer, M. (2006). Comment on article by Celeux et al. *Bayesian Analysis* **1**, 681–686.
- Sinharay, S. and Stern, H. S. (2003). Posterior predictive model checking in hierarchical models. *Journal of Statistical Planning and Inference* **111**, 209–221.
- Sinharay, S. and Stern, H. S. (2005). An empirical comparison of methods for computing Bayes factors in generalized linear mixed models. *Journal of Computational and Graphical Statistics* **14**, 415–435.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B* **64**, 583–639.
- Van Den Heede, K., Sermeus, W., Diya, L., Lesaffre, E., and Vleugels, A. (2006). Adverse outcomes in Belgian acute hospitals: Retrospective analysis of the national hospital discharge dataset. *International Journal for Quality in Health Care* **18**, 211–219.
- Warton, D. I. (2005). Many zeros does not mean zero inflation: Comparing the goodness-of-fit of parametric models to multivariate abundance data. *Environmetrics* **16**, 275–289.

Received October 2007. Revised June 2008.

Accepted June 2008.