

# A Joint Likelihood Alternative to WAIC for N-Mixture Models

Heather Gaya and Friends

2020

## 1 Introduction

Model selection is a critical part of analyzing ecological data. Often the importance of covariates for predicting a system is unknown and there is no available information on the most important variables in an ecological system. To further complicate the issue, observations are often biased by imperfect detection and human error, so direct data on the system is unavailable. To confront these challenges, a suitable method is needed to evaluate and compare available candidate models. While multiple methods have been proposed and thoroughly tested under a maximum likelihood framework, comparatively few methods have been discussed for a Bayesian framework and no consensus has been reached on which is the best option.

One popular method of Bayesian model selection is WAIC (Watanabe, 2010), a generalized version of AIC for Bayesian models. WAIC combines the estimated log predictive density of the data with a penalty for the estimated number of parameters in the model to produce a value that can be compared between alternative models. This method is valid in both hierarchical and mixture models (Watanabe, 2010) and has been shown to be an efficient alternative to cross-validation (Gelman et al., 2014). However, WAIC may fail when data are correlated (such as spatial or temporal models) (Hooten and Hobbs, 2015; Link et al., 2020).

## 2 Methods

### 2.1 WAIC

WAIC (Watanabe, 2010) estimates the expected log predictive densities for all data points and is defined as:

$$WAIC = -2\widehat{lpd} + 2\widehat{p}_{waic}$$

where  $\widehat{lpd}$  is the log predictive density of the data:

$$lpd = \sum_{i=1}^n \log(P(y_i|\theta_i))$$

and  $\widehat{p}_{waic}$  is the estimated effective number of parameters:

$$p_{waic} = \sum_{i=1}^n var_{post}(\log(P(y_i|\theta_i)))$$

If  $y_i$  and  $\theta$  are directly linked, and there is no error in the data, then WAIC is likely sufficient for finding the top model. However, in many wildlife models, error is introduced into the data in the form of imperfect detection. Consider a simple Binomial  $N$ -mixture model where both abundance,  $N$ , and detection probability,  $p$ , vary with site  $i$ .

$$\begin{aligned}\log(\lambda_i) &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots \\ N_i &\sim \text{Poisson}(\lambda_i) \\ \text{logit}(p_{ij}) &= \alpha_0 + \alpha_1 x_{i1} + \dots \\ y_{ij} &\sim \text{Binomial}(N_i, p_{ij})\end{aligned}$$

WAIC only considers the fit of the data  $y_{ij}$  to the latent abundance  $N_i$  as it relates to  $p_{ij}$ . As  $p_{ij}$  approaches 1, the log predictive density of the data approaches 0.

$$\begin{aligned}P(y_{ij}|p_{ij}, N_i) &= p_{ij}^{y_{ij}} (1 - p_{ij})^{(N_i - y_{ij})} \\ \lim_{p_{ij} \rightarrow 1} \log(P(y_{ij}|p_{ij}, N_i)) &= 0\end{aligned}$$

Therefore, as  $p_{ij}$  approaches 1, the state process model has no effect on WAIC. Similarly, as  $p_{ij}$  approaches 0:

$$\begin{aligned}P(y_{ij}|p_{ij}, N_i) &= p_{ij}^{y_{ij}} (1 - p_{ij})^{(N_i - y_{ij})} \\ \lim_{p_{ij} \rightarrow 0} \log(P(y_{ij}|p_{ij}, N_i)) &= 0\end{aligned}$$

This heavily constrains situations where the use of WAIC is appropriate.

## 2.2 WAICj

We suggest an adjustment to WAIC that considers the joint-likelihood of the data and relevant latent variables. We define our joint-likelihood approach to WAIC as:

$$WAIC = -2\widehat{jlpd} + 2\widehat{p}_{waicj}$$

with:

$$\begin{aligned}jlpd &= \sum_{i=1}^n \log(P(y_i|N_i)) + \log(P(N_i|\lambda_i)) \\ p_{waicj} &= \sum_{i=1}^n var_{post}(\log(P(y_i|N_i)) + \log(P(N_i|\lambda_i)))\end{aligned}$$

Under this formulation, model selection will account for the log likelihood in both the detection model and the underlying state process model.

## 2.3 Marginalized N

An alternative approach is to marginalize N in the model.

## 2.4 Simulated Case Study

To test our WAIC approaches, we simulated 300 bird abundance data sets at 50 sites from a poisson distribution. We simulated detection as four independent visits of each site with a negative effect of windspeed on detection probability. The covariate coefficients for the detection and state process models were drawn randomly for each simulation. We further tested the effect of different sample sizes (15, 25, and 50 sites) and detection functions (low detection and high detection) on WAIC performance. We analyzed each simulation's data using 4 candidate binomial N-mixture models (Table 1) and used all three formulations of WAIC to assess model performance.

## 3 Results

Across all simulations, WAIC correctly identified the top model in 67.5%, 81.2%, and 73.2% of simulations under the traditional, joint-likelihood, and marginalization approaches respectively. The joint-likelihood approach tended to rank the correct model as the top model more often than the other two methods (Figure 1). All three metrics had more accurate model selection and gave more weight to the correct model when detection was lower and sample size was highest 2)

## 4 Discussion

WAICj is better but it's still not perfect. Oof.

## 5 Literature Cited

### References

- Andrew Gelman, Jessica Hwang, and Aki Vehtari. Understanding predictive information criteria for bayesian models. *STATISTICS AND COMPUTING*, 24(6):997–1016, 2014. ISSN 0960-3174. doi: 10.1007/s11222-013-9416-2.
- M. B. Hooten and N. T. Hobbs. A guide to bayesian model selection for ecologists. *Ecological Monographs*, 85(1):3–28, 2015. doi: <https://doi.org/10.1890/14-0661.1>. URL <https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1890/14-0661.1>.
- William A. Link, John R. Sauer, and Daniel K. Niven. Model selection for the north american breeding bird survey. *Ecological Applications*, 30(6):e02137, 2020. doi: <https://doi.org/10.1002/eap.2137>. URL <https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1002/eap.2137>.

Sumio Watanabe. Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11(116):3571–3594, 2010. URL <http://jmlr.org/papers/v11/watanabe10a.html>.

Model Number	Formulation
Model 1*	$\lambda(tree_i)p(wind_{it})$
Model 2	$\lambda(.)p(.)$
Model 3	$\lambda(tree_i)p(.)$
Model 4	$\lambda(.)p(wind_{it})$

Table 1: Candidate Models. The asterisk denotes the model under which data was simulated. Periods represent intercept only models.

No. Sites	Low Detection			High Detection		
	WAIC	WAICj	Marginalized	WAIC	WAICj	Marginalized
15	0.59 (.27)	.69 (.31)	.57 (.31)	.57 (.32)	.74 (.33)	.73 (.30)
25	0.66 (.30)	.79 (.28)	.67 (.33)	.66 (.32)	.85 (.26)	.82 (.29)
50	0.77 (.27)	.83 (.29)	.75 (.30)	.60 (.34)	.86 (.27)	.85 (.26)

Table 2: Weight Given to the Correct Candidate Model. Each site and detection combination represents the mean (sd) weight from 100 simulations.

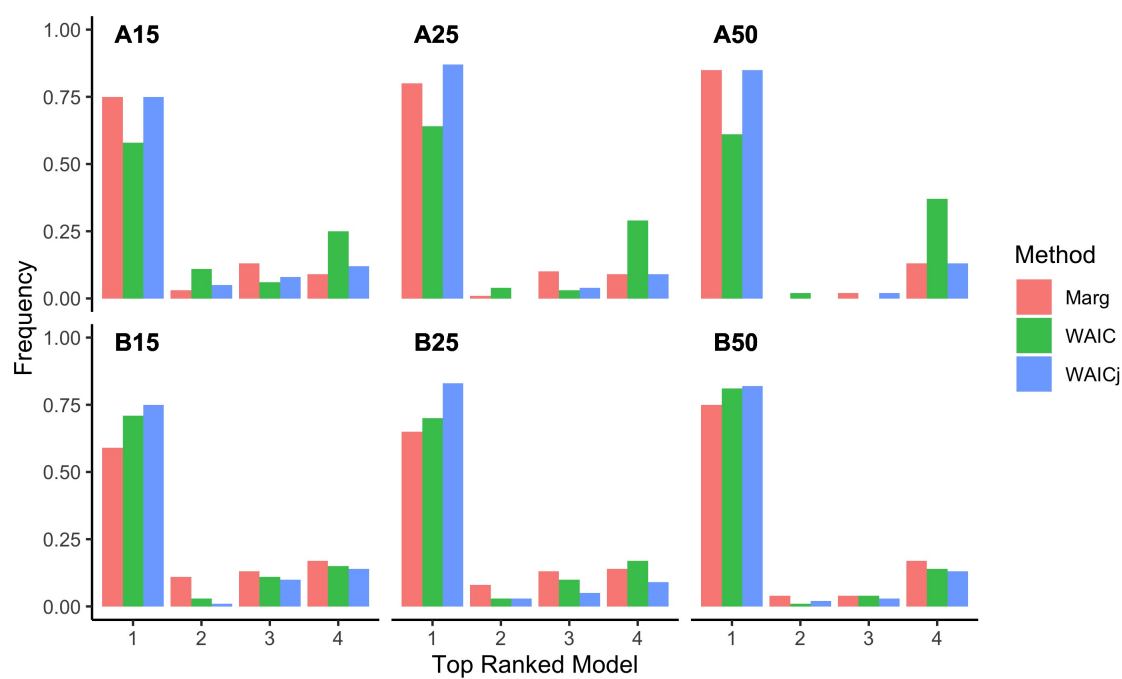


Figure 1: Model Ranking Under a A. High Detection and B. Low Detection Scenario. Each scenario was analyzed with data from 15, 25 and 50 sites.