Posterior Predictive Checks for Model Assessment in Occupancy Modeling

Michael Lerch

Department of Mathematical Sciences
Montana State University

July 31, 2012

A writing project submitted in partial fulfillment
of the requirements for the degree

Master of Science in Statistics

APPROVAL

of a writing project submitted by

Michael Lerch

This writing project has been read by the writing project advisor and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the Statistics Faculty.

Date	Megan Higgs Writing Project Advisor
Date	Mark C. Greenwood Writing Project Coordinator

Abstract

Though methodology and motivation for the use of posterior predictive checks has been introduced in some textbooks (e.g. Gelman et al. 2004), the technique is not commonplace in, for example, applied ecological studies using Bayesian techniques. In many studies using Bayesian techniques, the only semblance of model checking is the use of DIC for model selection, but this does not provide practically meaningful information about the quality of the fit. In this paper, we demonstrate the use of posterior predictive checks for model assessment under an occupancy study of small land birds in the Helena National Forest. Occupancy of these birds was recorded at many different sites for several years before the onset of a massive pine beetle epidemic as well as for several years after the onset of the epidemic. We present two rather simple models to describe the occupancy of the bird species at the different sites and demonstrate the use of posterior predictive checks for model assessment and comparison of these two models.

1 Introduction

Model assessment is a critical part of any model-based statistical analysis. However, many ecological studies using Bayesian techniques tend to be curt in this regard. Often, no assessment of the model used is provided at all. This dearth of model assessment is likely due to a lack of knowledge of appropriate tools to perform model assessment and examples of their use. We hope to address this void by recommending the use of posterior predictive checks as a simple, intuitive, and interpretable technique to accomplish model assessment.

The use of posterior predictive checks has been previously proposed for model assessment (e.g. Gelman et al. 2004; Schofield and Barker 2011). We echo the sentiments of these authors and also provide an example of such use in this paper. Further, we demonstrate the value of posterior predictive checks in assessing the validity of assumptions when considering reduced versions of fuller models. In the case of multiple proposed models, criteria such as DIC are often used for selecting one of a set of competing models. However, omnibus measures such as DIC, as well as other classic goodness of fit tests, offer little interpretability and are only meaningful relative to other models fit on the same data. Bayesian model selection techniques–including Bayes factors and Bayesian hypothesis testing–can be complicated, computationally expensive, and ill-posed under certain conditions including diffuse priors (Gelfand and Ghosh 1998; Gelman et al. 2004). We argue that posterior predictive checking may aide in choosing among possible models under practically meaningful criteria.

In this paper, we present the use of posterior predictive checks for model assessment and comparison for an occupancy study of 46 small land birds in the Helena National Forest in Montana (Mosher 2011). Two simple models are proposed to explore the change in occupancy of these birds that may be attributable to the onset of a mountain pine beetle epidemic.

2 Avian response to mountain pine beetle

2.1 Study design

A study of small land birds was conducted in the Elkhorn Mountains in the Helena National Forest from 2003 to 2011. In all, 46 species of noncarnivorous birds were observed over the course of the study. Within the Elkhorn Mountain area, four study units (Figure 1) were selected, each between 130 and 260 hectares. These units were selected based on management priorities and are characteristic of dry mixed coniferous forests. Within these four study units, a total of 76 point count locations were established. These point count locations were randomly selected under the constraint that they are at least 250 meters apart and 150 meters from any study unit boundaries. In the years 2003 through 2006 and again from 2009 through 2011, the point count locations were visited approximately bi-weekly from May 22 to July 2; this amounts to about three to four visits per site per year. In a five minute period after the dawn chorus, researchers recorded all bird species detected within a radius of approximately 75 meters.

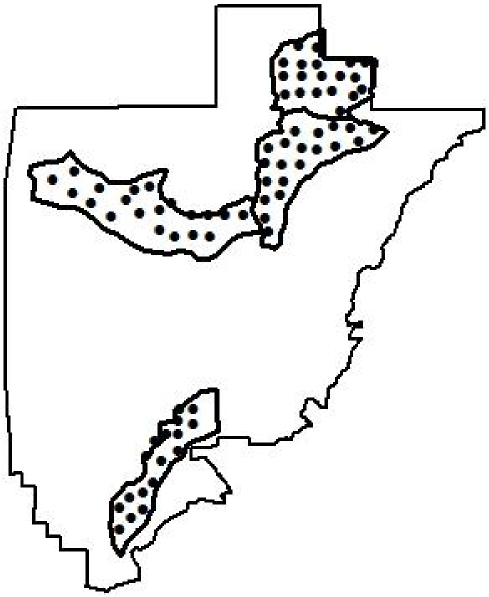


Figure 1: Four study units and 76 point count locations in the Elkhorn Wildlife Management Unit in the Helena National Forest

Much of the North American Rockies (including the Helena National Forest) experienced an exceptional mountain pine beetle epidemic beginning in about 2006 and peaking around 2008. The epidemic has been described as one of the worst such epidemics in the past 100 years (Kaufmann et al. 2008). The original objective of the researchers was to study avian response to prescribed fire. However, litigation in 2006 prevented the prescribed burning. The study was still biologically valuable, though, due to the onset of the beetle epidemic which presented a natural disturbance on the system. This data set is unique in that measurements before and after a natural phenomenon are available. The researchers are interested in changes in occupancy of the 46 species over time, particularly related to the timing of the pine beetle epidemic. The researchers believe that the influx of mountain pine beetle changes forest dynamics, which in turn influences the occupancy of bird species.

2.2 Occupancy model

The observable in this study can be modeled using a Bernoulli process: for a single visit, the species is either observed or not. We denote $y_{ijkt} = 1$ if the i^{th} species was observed at the j^{th} site on the k^{th} visit of the t^{th} year of the study and $y_{ijkt} = 0$ if not observed. We designate

$$y_{ijkt} \sim \text{Bernoulli}(\pi_{ijt}) \tag{2.1}$$

as the observation model ($i = 1, \ldots, n_{spec}$, $j = 1, \ldots, n_{site}$, $t = 1, \ldots, n_{year}$, $k = 1, \ldots, n_{visit,jt}$, where $n_{visit,jt}$ is the number of visits to the j^{th} site in the t^{th} year). The parameter π_{ijt} is the probability of making a detection of the i^{th} species at the j^{th} site for all visits during the t^{th} year of the study. Here, we use an implicit assumption that all visits to a site in a specified year are a set of replications; that is, we are using a primary period design (MacKenzie et al. 2003). A biologically motivated season defines a primary sampling period in which conditions are constant across visits within the specified primary period. In our case, all sampling was performed between May 22 and July 2, and we assume that the biological processes are constant over this time. Specifically, we employ the assumption of closure of sites over the primary sampling period: the true occupancy of the species is constant at a site over the entire primary sampling period. References to "years" in this study refer to the primary sampling periods between May 22 and July 2.

In our model, we will assume no false positive detections. Hence, in order to observe the species, the species must be present. However, we do allow for imperfect detection where the species is present but missed. In many studies, imperfect detection must be considered. Not accounting for imperfect detection can lead to inferences that are difficult to interpret or wrong (Royle and Dorazio 2008). Naively assuming a perfect detection rate leads to inferences that suggest an overly volatile population; common dynamic population measures such as extinction rates and survivability will be overestimated and underestimated, respectively. To account for this imperfect detection, we consider two sub-processes that contribute to π_{ijt}:

$$\pi_{ijt} = z_{ijt} \times p_{ijt} \ . \tag{2.2}$$

Here, z_{ijt} represents the true occupancy (1 if present and 0 if not present) for the i^{th} species at the j^{th} site during the t^{th} season, and p_{ijt} is the probability of detection *given* that the species is present. Simultaneous estimation of both z_{ijt} and p_{ijt} requires replication which we achieve with multiple visits during each primary sampling period and the assumption of closure during primary sampling periods..

The value z_{ijt} is partially observed under our assumption of no false positive detections. If a bird of species i is observed on at least one visit during year t at site j, $z_{ijt} = 1$. However, if a particular species is not observed at a particular site on any vists of a particular year, the true occupancy of that species at that site is unknown. We introduce the occupancy probability, ψ_{ijt}, to parameterize the Bernoulli distribution for true occupancy of the i^{th} species at the j^{th} site during the t^{th} season:

$$z_{ijt} \sim \text{Bernoulli}(\psi_{ijt}) \tag{2.3}$$

In many applications, this occupancy probability is the parameter of interest to the researchers. Researchers may be interested in how this probability changes over time or may model it based on some covariates of interest. In our case, we are interested in changes in the occupancy probability attributable to the pine beetle epidemic.

In cases where perfect detection is a reasonable assumption, Equation 2.1 reduces to Equation 2.3. Here, the true occupancy, z_{ijt} is observed and may be appropriately modeled with binary logistic regression techniques. This is a much simpler analysis, but is not appropriate our study.

2.3 Proposed models

We begin with the model used by Mosher (2011). This model supposes constant occupancy and detection probabilities before the pine beetle epidemic began (years 2003-2006) and separate constant occupancy and detection probabilities after the pine beetle epidemic began (years 2009-2011). We will refer to this model as the beetle indicator model. The occupancy and detection probabilities under this model are shown in Equation 2.4 and subsequently follow 2.1, 2.2, and 2.3 to completely specify the model. Mosher (2011) estimates the parameters of this model hierarchically, assuming exchangealility among all species and using non-informative priors for the hyperparameters:

$$\begin{aligned} \operatorname{logit}(p_{ijt}) &= \alpha_{1i} + \alpha_{2i} I(t \geq 2008) \\ \operatorname{logit}(\psi_{ijt}) &= \beta_{1i} + \beta_{2i} I(t \geq 2008) \\ \alpha_{1i} &\sim \mathrm{N}(\mu_{\alpha 1}, \sigma_{\alpha 1}) \\ \alpha_{2i} &\sim \mathrm{N}(\mu_{\alpha 2}, \sigma_{\alpha 2}) \\ \beta_{1i} &\sim \mathrm{N}(\mu_{\beta 1}, \sigma_{\beta 1}) \\ \beta_{2i} &\sim \mathrm{N}(\mu_{\beta 2}, \sigma_{\beta 2}) \\ \mu &\sim \mathrm{N}(0, 1000) \\ \sigma &\sim \mathrm{Unif}(0, 10) \end{aligned} \tag{2.4}$$

where I is the indicator function, μ represents each of $\mu_{\alpha 1}$, $\mu_{\alpha 2}$, $\mu_{\beta 1}$, $\mu_{\beta 2}$, and σ represents each of $\sigma_{\alpha 1}$, $\sigma_{\alpha 2}$, $\mu_{\beta 1}$, $\mu_{\beta 2}$.

The form of the beetle indicator model is clearly biologically motivated and provides simple inferences regarding the question of interest: the parameters β_{2i} contain the information of how the occupancy rates changed as associated with the timing of the beetle epidemic. However, this model still relies on a major assumption that all the pre-epidemic years are the same and all the post-epidemic years are the same. We would like to investigate the validity of this assumption, particularly for the post-epidemic years since the effects of the beetle epidemic on the system are likely dynamic over a period of years. We do this by fitting a richer model referred to as the *separate years model*. The separate years model (2.5) allows each species to have a different occupancy probability and detection probability each year. Like in the beetle indicator model, we estimate the parameters hierarchically assuming exchangeability among species and using uninformative priors for the hyperparameters:

$$\begin{aligned}
\mathrm{logit}(p_{ijt}) &= \alpha_{it} \\
\mathrm{logit}(\psi_{ijt}) &= \beta_{it} \\
\alpha_{it} &\sim \mathrm{N}(\mu_{\alpha}, \sigma_{\alpha}) \\
\beta_{it} &\sim \mathrm{N}(\mu_{\beta}, \sigma_{\beta}) \\
\mu &\sim \mathrm{N}(0, 1000) \\
\sigma &\sim \mathrm{Unif}(0, 10)\ .
\end{aligned} \tag{2.5}$$

It should be noted that neither of these models consider any site to site differences. This is a clear weakness of the models and will be addressed later, but we continue with this scenario to demonstrate posterior predictive checking.

3 Methods

3.1 Exploratory data analysis

No model assessment or selection technique renders an exploratory analysis of the data futile. However, the two processes we are modeling (the true occupancy and detection rate) are not explicitly observed. This makes an empirical evaluation difficult. However, the naive occupancy rates and naive detection rates may serve as suggestive replacements. The difference between the occupancy probability and occupancy rate is that the occupancy probability is a parameter describing the true process and the occupancy rate is a particular realization, much like the relationship between an observed proportion and a true probability. Likewise, detection probabilities and detection rates are parameters and realizations, respectively.

Because we assume there are no false positive observations, the true occupancy is partially observed. If a species is observed at any visit to a particular site during a particular primary sampling period, the occupancy of that species at that site for that year is known: $z_{ijt} = 1$. In the cases where the species is not observed during any visits to a particular site during a particular year, the occupancy is not known; either (1) the species is occupying the site but not detected or (2) the species is not present and therefore not detected. We define naive occupancy, z'_{ijt}, to be 1 if the i^{th} species is observed on any visit to the j^{th} site during the t^{th} season and 0 otherwise:

$$z'_{ijt} = \begin{cases} 1 & \text{if } \max_k(y_{ijkt}) = 1 \\ 0 & \text{if } \max_k(y_{ijkt}) = 0 \,. \end{cases} \tag{3.1}$$

Naive occupancy gives a lower bound on the true occupancy. Summation of naive occupancy over the sites for a given species in a primary sampling period gives the total number of sites at which the particular species is known to be present. It is customary to divide by the total number of sites in the study to obtain the naive occupancy rate:

$$\text{Naive occupancy rate}_{it} = \frac{1}{n_{site}} \sum_{j=1}^{n_{site}} z'_{ijt} \quad , \tag{3.2}$$

where n_{site} is the total number of sites. Investigation of the naive occupancy rate by year for some species suggests a non-constant occupancy rate which in turn may suggest a non-constant occupancy probability (Figure 2). However, this may also be a result of a non-constant detection rate.

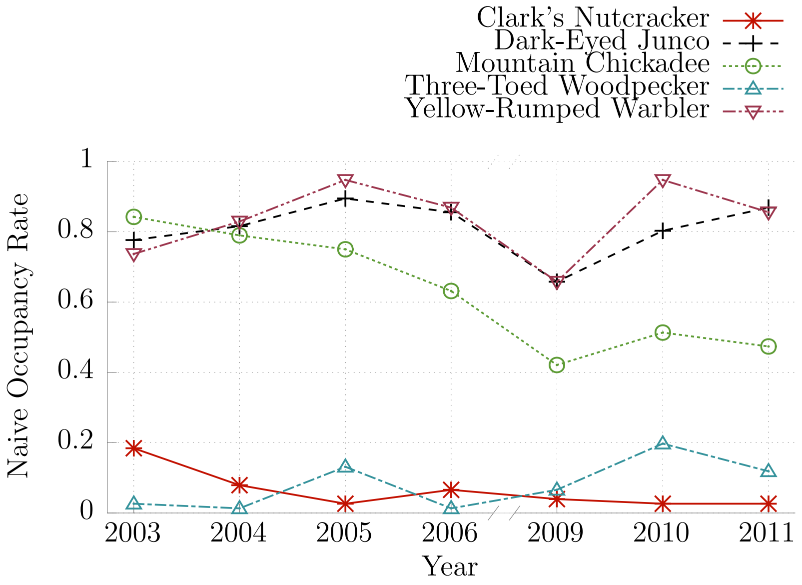


Figure 2: Naive occupancy rates for several species of particular interest

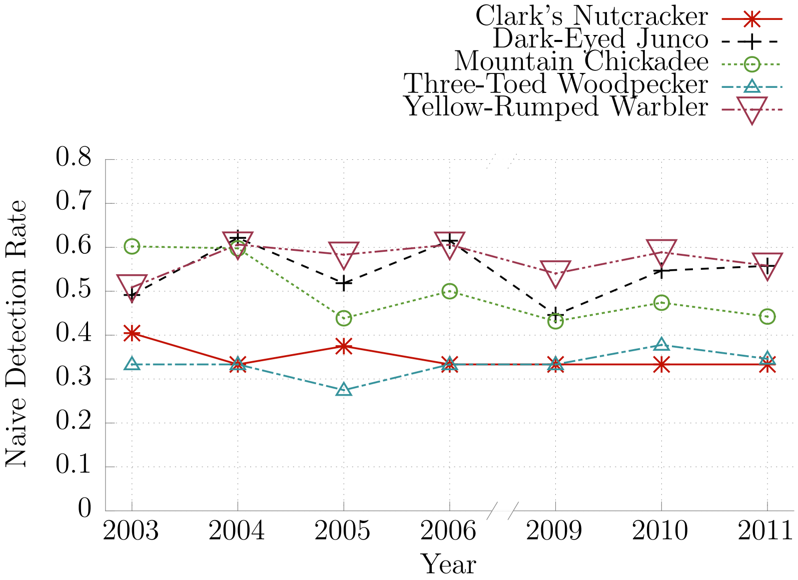


Figure 3: Naive detection rates for several species of particular interest. A value of zero indicates the species was not observed that year.

Similarly, we can calculate an upper bound on the true detection probability using a naive detection rate. We define naive detection rate as the total number of visits in which a particular species is observed in a particular primary sampling period divided by the total number of visits to sites in which the species is known to be present. This value is an upper bound because the calculation of the true detection rate would include additional sites in which the species was not observed in any visit but was actually present:

$$\text{Naive detection rate}_{it} = \frac{\displaystyle\sum_{j=1}^{n_{site}} \sum_{k=1}^{n_{visit,jt}} y_{ijkt}}{\displaystyle\sum_{j=1}^{n_{site}} \left(z'_{ijt} \times n_{visit,jt}\right)} \quad , \tag{3.3}$$

where $n_{visit,jt}$ is the number of visits to the j^{th} site in the t^{th} year. Plots of the naive detection rate by year for some species suggest yearly changes in detection, though for many, the naive detection rate appears relatively constant over time (Figure 3).

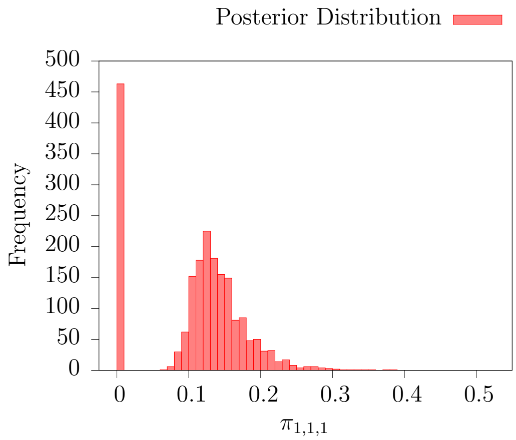
3.2 Posterior predictive distributions

3.2.1 Theory and application

A familiar use of statistical models is to predict new observations at a given set of covariate values. Many ecologists are acquainted with simulating data from a model under particular values of model parameters (e.g. MLEs) without accounting for any uncertainty of the values. Under a Bayesian approach, however, we can explicitly incorporate the uncertainty because the parameters have distributions. Hence, the simulation of data from a Bayesian model takes on an additional step in which the values of the model parameters themselves must be chosen at random from their distribution.

A posterior prediction is a simulation from a model where the values for the model parameters are drawn from the posterior distributions for the parameters. For example, in this study, we use Gibbs sampling to obtain random draws of our parameters from their posterior distributions. Graphically, the histogram of the chain of samples drawn for the parameter $\pi_{1,1,1}$ (the probability of making a detection of the first species on a visit to the first site in the first year of the study) for the beetle indicator model is shown in Figure 4. Rather than taking only the mean, median, or the mode, we take a random draw from these values and simulate from our model through Equation 2.1, to obtain a posterior prediction for whether or not the first species was observed on a visit to the first site in the first year of the study.

We can use posterior predictive checks to approach the question "could this model have generated the data we observed?" To do this, we use posterior predictions to construct predicted data sets (using different sets of parameter values) and compare them to the original data set. For example, in this study the first site was visited three times in the first year. Therefore, to replicate the observations for the first species at the first site during the first year, we draw a value for $\pi_{1,1,1}$ from its posterior distribution and make three draws from Bernoulli($\pi_{1,1,1}$) (2.1). This process is repeated for $\pi_{1,1,2}$, and so on, until we have replicated the entire original data set. Following Gelman et al. (2004) we denote this replicate data set as y^{rep}. The process of generating y^{rep} is then repeated many times to explore the entire space of possible posterior predictions under the different possible parameter values defined by the posterior distribution. To *check* our model with these posterior predictions, we simply compare these y^{rep} simulations to the original data.



$$\xrightarrow{\textit{sample}} \pi^*_{1,1,1}$$

$$y^{\text{rep}}_{1,1,k,1} \sim \text{Bernoulli}(\pi^*_{1,1,1})$$

Figure 4: A graphical demonstration of generating posterior predictive samples. Parameter values are drawn from posterior distributions and predictions are drawn utilizing these parameter values

It is possible to compare the simulations y^{rep} directly to the original data set (Laud and Ibrahim 1995), but we suggest an easier to interpret strategy. Researchers should determine a particular aspect of the observed data for which they wish to investigate the model's ability to reproduce. This aspect of the observed data must be captured in some statistic that can be calculated on the original data so that it can be calculated using both the original data and each of the replicate data sets, y^{rep}. The statistic calculated from the observed data is compared to the values calculated for each replicate data set to see if the model's predictions are consistent with the observed data. We may then assess the model's ability to predict data consistent with the original data as measured by this particular statistic that has been deemed important by the researchers.

Previously, we introduced the naive occupancy rate as a statistic providing a lower bound on the true occupancy probability. For this paper, we will use the naive occupancy rate as our statistic upon which to perform our posterior predictive checks. Since the naive occupancy rate is calculated each year for each species, we are actually dealing with 322 different statistics. This is a great departure from model selection criteria such as DIC where a single value summarizes the fit of the model. However, we may still investigate all of these statistics simultaneously. In fact, we argue that having multiple measures is a strength of posterior predictive checking in that we may assess the model fit under practically meaningful summaries. In this example, we can determine particular species and years in which the model performs adequately at predicting naive occupancy rates. We illustrate the difference between the simplified beetle indicator model and the separate year model which we expect to perform well for posterior predictive checks.

3.3 Posterior predictive checking and p-values

Under frequentist approaches, a p-value is typically defined as the probability of generating a test statistic under different random assignments or random selections as or more extreme than that observed *given* some null parameter value(s). Here, we are clearly not fixing parameter values, but the Bayesian p-values we will calculate are still analogous to those calculated in a one-sided hypothesis test. We define a Bayesian p-value as the probability of generating a test statistic from posterior predictions that is more extreme than that calculated from the observed data. In other words, the sidedness of the test is such that the smaller p-value is obtained.

Small p-values are indicative of poor model fit. Graphically, a histogram of the statistic calculated for many replicate data sets can be overlaid with a demarcation of the observed statistic (Figure 5). A small Bayesian p-value corresponds to an observed value in the tails of histogram (or off the histogram); the model rarely predicts data with statistics as extreme as the value actually observed.

After calculation of this Bayesian p-value, a heuristic approach to model assessment and comparison can be performed. As mentioned, our statistic of choice, the naive occupancy rate is calculated for each species for each year of the study. Therefore, the p-value for a single naive occupancy rate is not sufficient for model assessment. Instead, the p-values for all species and all years should be evaluated in concert. This helps address not only the quality of the model fit but indicates *where* a model fails, and specifically allows researchers to assess whether this failure is of practical importance. For example, researchers may find a particular bird guild (e.g. woodpeckers) to be of greater importance than other guilds, so they may weight the ability of the model to predict the naive occupancy of woodpeckers more highly than that for other species when assessing the fit of the model. It is important to realize that researchers should not feel confined to a single statistic or family of statistics. This heuristic approach allows great flexibility in assessment. For the purposes of this paper, we only investigate the naive occupancy rate.

In this study, we assess validity of the assumption that the pre-epidemic years may be merged and that the post-epidemic years may be merged. Since the separate years model is a richer model than the beetle indicator model, we expect the separate years model will fit better than the beetle indicator model. However, Bayesian p-values offer no "penalty" for richer models. The biological value of the beetle indicator model is inherent to the researchers in the structure of the model and this value should be taken into consideration while assessing the fit. Though the researchers should expect a better fit from the separate years model, they may find the fit of the beetle indicator model to be good enough to accept the assumption of grouped years.

An important consideration for our strategy is that the separate years model is explicitly modeling the occupancy rate and the detection probability, the two elements that contribute to the naive occupancy rate. Hence, posterior predictions of the naive occupancy rate will be consistent with the data. Our model comparison will assess the validity of the assumption that the pre-epidemic and post-epidemic years may be grouped. However, additional checks, such as on the validity of considering all sites to be equivalent, may also be warranted.

4 Results

4.1 Model Assessment

In assessing the fit of our models, we begin in demonstration by considering a single species in a single year: the American robin in 2011. Results are summarized using histograms of posterior predictive naive occupancy rates for the American robin species in 2011 for the two proposed models (Figure 5). The observed naive occupancy rate is denoted with a vertical line.

Graphically, it is clear that for the naive occupancy of the American robin in 2011, the separate years model is more consistent with the observed data than the beetle indicator model. Our model predicts lower naive occupancy in 2011 than is actually observed when years are lumped together in the beetle indicator model. The Bayesian p-values may be calculated for a more quantitative comparison. For the beetle indicator model, the p-value is 0.1075 (only 10.75% of posterior predictions had naive occupancies larger than that

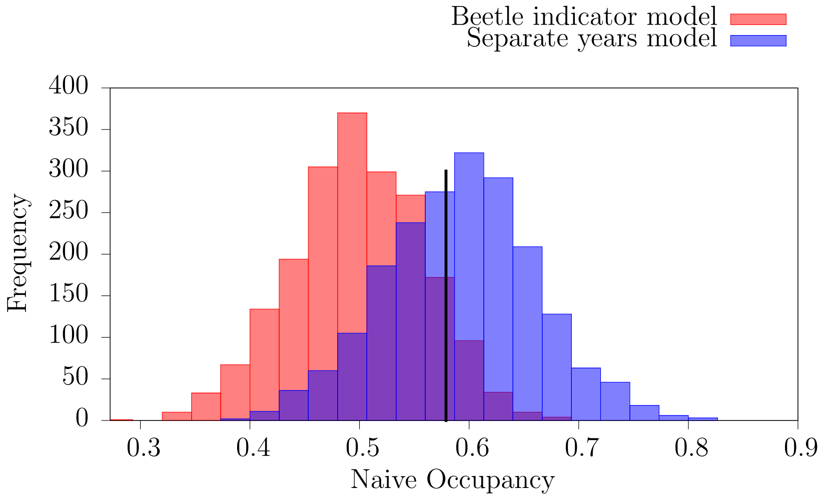


Figure 5: Posterior predictive draws of naive occupancy for robin in 2011 from the beetle indicator model and the separate years model. The vertical line denotes the observed value

actually observed); for the separate years model, the p-value is 0.4735. As we alluded to previously, there are many species and many years to be considered and we would like to investigate how our model fits for the many species over the multiple years of the study.

In our next step, we repeat the process of assessing naive occupancy of the American Robin and obtain posterior predictive p-values for all years (Figure 6). This gives a fuller comparison of the two models for this species. For the American Robin, we can see that the efficacy of the beetle indicator model waxes and wanes among the years due to heterogeneity across years beyond what is attributable to the onset of the beetle epidemic. Meanwhile, the separate years model consistently produces naive occupancy rates in congruence with the observed rate. While this is not surprising, it provides a nice illustration of a comparison of two models and the use of posterior predictive checks.

Had the American Robin been the only bird upon which we are interested in making inference, this may be sufficient for our model comparison and assumption checking. Since we are actually interested in multiple species simultaneously, we calculate Bayesian p-values for naive occupancy rate for all the species. For this paper, however, we only plot a subset of all 46 species that are of interest to the researchers (Figures 7 and 8) showing that

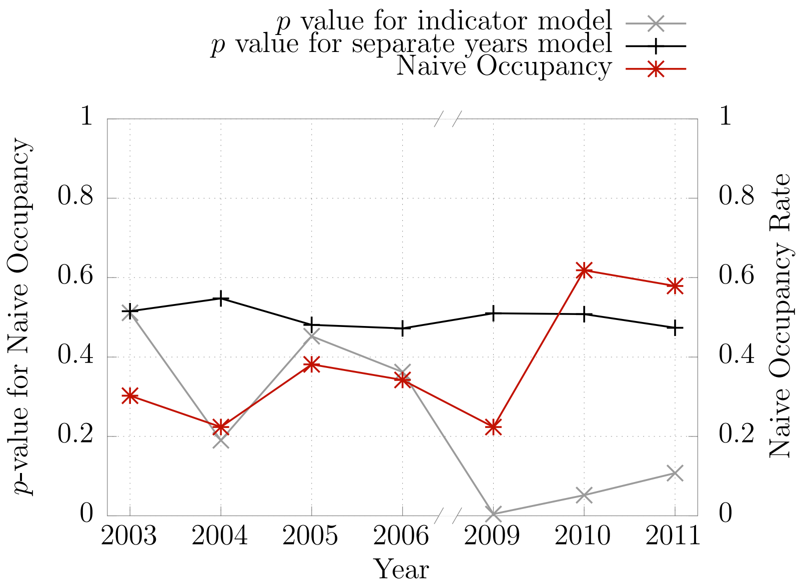


Figure 6: Bayesian p-values for robin across all years of the study under both models and observed naive detection rate

the beetle indicator model is simply not rich enough to adequately recreate the data, particularly in the years most of interest to the researchers. We believe that the benefit of the interpretability of the beetle indicator model does not outweigh its lack of fit and propose the separate years model to the researchers.

4.2 Answering the question of interest

The researchers' question of interest lies in the change in occupancy probability that may be attributable to the pine beetle epidemic. The allure of the beetle indicator model is quite clear. In Equation 2.4, this question is easily answered with the parameter β_{2i}. However, we have concluded that this beetle indicator model is an oversimplification that could possibly lead to misinterpretation of results. However, we further must require that the model is capable of answering the question of interest. To answer their question using the separate years model, the researchers may consider a weighted average of each species occupancy probability before and after the epidemic. If weighting is constant, then we will reproduce the inference from the beetle

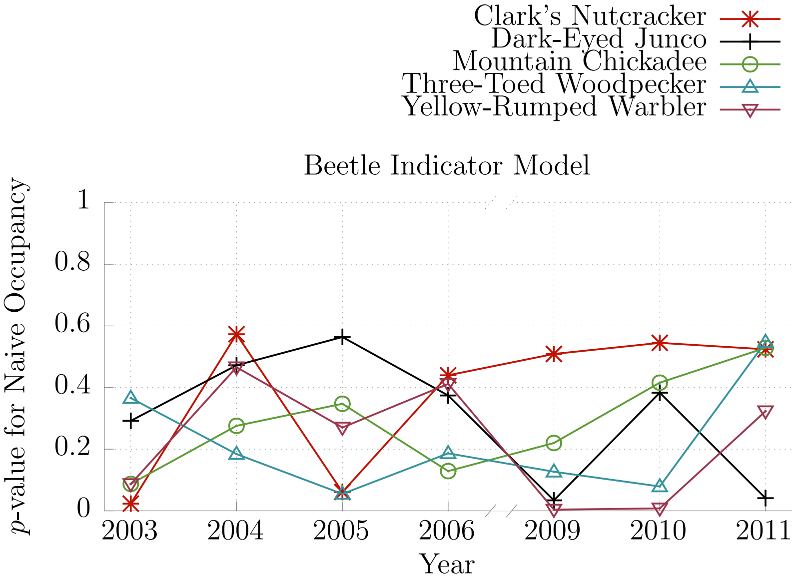


Figure 7: Bayesian p-values for several species of interest under beetle indicator model

indicator model. However, the researchers may wish to downweight the years 2006 and 2009 if they feel that the beetle epidemic status was actually transitional in these years. Additional information and strategies will be provided in Mosher et al. (in preperation).

5 Discussion

5.1 Assumption of additional model assumptions.

As we mentioned previously, neither of our models address the possibility of site to site differences. In this paper, we addressed the assumption of lumping the pre-epidemic years as well as the post-epidemic years. The assumption of identical sites is clearly a large assumption and should be addressed in development of a model. A posterior predictive check should be developed with the researchers to assess whether the models we used in this paper are capable of reproducing any variability observed among sites.

Further, we employed the assumption of closure of sites in utilizing the primary period design. The true occupancy of each species was assumed

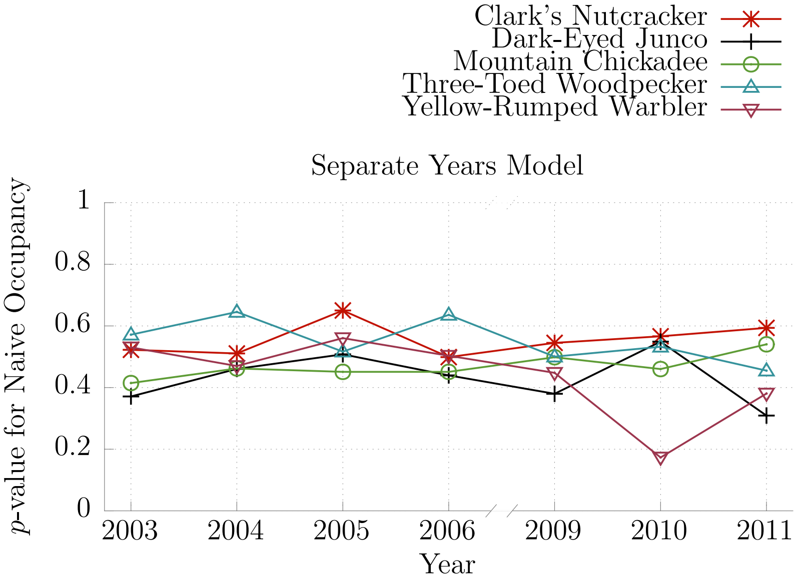


Figure 8: Bayesian p-values for several species of interest in study under separate years model

constant over each primary period at each site. Evaluation of this assumption may be difficult under the current study design, but this assumption must be appreciated by the researchers as it is fundamental to our inferences.

5.2 Discussion of other selection and assessment routines

There already exist a great number of omnibus criteria for use in model selection and assessment. Posterior predictive checks fill a different niche. Perhaps the most widely used tool in model selection for applied Bayesian statistical analyses is the deviance information criterion (DIC). However, the motivation in use may be more tightly related to the simplicity in calculating this measure in WinBUGS and JAGS rather than any theoretical basis. In an attempt to illuminate this basis, Plummer (2008) suggests that despite the current practical advantages (namely availability in Gibbs sampling packages), the theoretical foundations of DIC remain controversial. Further, some note that it may take a "zillion" iterations in a Gibbs sampling routine for DIC to converge (Gelman 2011). Users of DIC must be sure that convergence has been achieved for DIC in addition to their model parameters.

Model selection with DIC amounts to choosing the model with the smallest DIC value. DIC values are relative amongst models fit to a particular data set. As such, researchers may choose the “best” model as determined by DIC and may be unwittingly choosing the best of a collection of poorly fitting models. Further, like all omnibus statistics, researchers cannot determine the particular aspect of the study that leads to poor fit with DIC alone. We believe that even when using DIC for model selection, it is appropriate to perform model assessment. Posterior predictive checks can compliment DIC by providing tangible motivation for model selection.

Though DIC is widely used in applied Bayesian statistics, there are a host of similar criteria (e.g. Ando 2007; Watanabe 2010). There is no reason posterior predictive checking would be incompatible with any of these measures.

Some authors have developed model selection routines based explicitly on posterior predictive distributions. Laud and Ibrahim (1995) introduce an L-criterion based on posterior predictive error:

$$L_m^2 = \sum_{i,j,k,t} \left[\{E(\tilde{y}_{ijkt}) - y_{ijkt}\}^2 + \mathrm{var}(\tilde{y}_{ijkt}) \right] \tag{5.1}$$

where $\tilde{y}_{ijkt}$ are the posterior predictions (the individual components of y^{rep} of the replicate data) and y_{ijkt} are the original data. The authors posit that good models will have smaller values of L_m^2. Though this measure is based on posterior predictions, it is an omnibus measure more closely related to DIC than posterior predictive checks. Again, posterior predictive checks may add depth to model selection based on L_m^2, but by itself, L_m^2 does not offer much interpretability. Further, we would argue against this measure for occupancy study like presented in this paper. While the expectations of the posterior predictions are continuous from zero to one, the observations themselves are binary.

6 Conclusion

In this paper, we explored the use of posterior predictive checks in model assessment and in assessing the validity of assumptions. In a study of avian occupancy possibly related to the mountain pine beetle epidemic in the Elkhorn

Mountains of the Helena National Forest, we found that an assumption of constant occupancy probability in pre-beetle years and a second, constant occupancy probability in post-beetle years may not be valid. Based on these observations, we suggested a model that allowed a different occupancy probability for each year of the study.

7 Acknowledgements

We thank Brittany Mosher and Vicki Saab for their discussions, explanations, and gracious permission to use their data. Additionally, we thank Jay Rotella for his discussion and insight into ecological techniques. Megan Higgs provided her invaluable support and time in directing this project. The data was collected under funding from the USDA Forest Service Rocky Mountain Research Station and the Helena National Forest.

References

Ando, T. (2007), "Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models," *Biometrika*, 94, 443–458.

Gelfand, A. E. and Ghosh, S. K. (1998), "Model choice: A minimum posterior predictive loss approach," *Biometrika*, 85, 1–11.

Gelman, A. (2011), "Deviance, DIC, AIC, cross-validation, etc," *http://andrewgelman.com/2011/06/deviance_dic_ai/*.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004), *Bayesian data analysis*, CRC press.

Kaufmann, M. R., Aplet, G. H., Babler, M., Baker, W. L., Harrington, M., Hawkes, B. C., Huckaby, L. S., Michael, J., Kashian, D. M., Keane, R. E., Kulakowski, D., Negron, J., Popp, J., Romme, W. H., Schoennagel, T., Shepperd, W., Smith, F. W., Sutherland, E. K., and Veblen, T. T. (2008), "The Status of Our Scientific Understanding of Lodgepole Pine and Mountain Pine Beetles A Focus on Forest Ecology and Fire Behavior," *The Nature Coservancy, Arlington, VA*, GFI techni, 1–22.

Laud, P. and Ibrahim, J. (1995), "Predictive model selection," *Journal of the Royal Statistical Society. Series B*, 57, 247–262.

MacKenzie, D., Nichols, J., Hines, J., and Knutson, M. (2003), "Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly," *Ecology*, 84, 2200–2207.

Mosher, B. A. (2011), "Avian community response to a mountain pine beetle epidemic," Ms thesis, Montana State University.

Mosher, B. A., et. al. (In preperation), "TBA," .

Plummer, M. (2008), "Penalized loss functions for Bayesian model comparison." *Biostatistics (Oxford, England)*, 9, 523–39.

Royle, J. A. and Dorazio, R. M. (2008), *Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations and communities*, Academic Press.

Schofield, M. R. and Barker, R. J. (2011), "Full Open Population CaptureRecapture Models With Individual Covariates," *Journal of Agricultural, Biological, and Environmental Statistics*, 16, 253–268.

Watanabe, S. (2010), "Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory," *Journal of Machine Learning Research*, 11, 3571–3594.