

Modeling Unobserved Sources of Heterogeneity in Animal Abundance Using a Dirichlet Process Prior

Robert M. Dorazio,^{1,*} Bhramar Mukherjee,² Li Zhang,³ Malay Ghosh,⁴
Howard L. Jelks,⁵ and Frank Jordan⁶

¹U.S. Geological Survey and Department of Statistics, University of Florida, Gainesville, Florida 32611-0339, U.S.A.

²Department of Biostatistics, University of Michigan

³Department of Quantitative Health Sciences, Cleveland Clinic

⁴Department of Statistics, University of Florida

⁵U.S. Geological Survey, Florida Integrated Science Center, 7920 NW 71 Street, Gainesville, Florida 32653, U.S.A.

⁶Loyola University New Orleans, Department of Biological Sciences, New Orleans, Louisiana 70118, U.S.A.

**email*: bdorazio@ufl.edu

SUMMARY. In surveys of natural populations of animals, a sampling protocol is often spatially replicated to collect a representative sample of the population. In these surveys, differences in abundance of animals among sample locations may induce spatial heterogeneity in the counts associated with a particular sampling protocol. For some species, the sources of heterogeneity in abundance may be unknown or unmeasurable, leading one to specify the variation in abundance among sample locations stochastically. However, choosing a parametric model for the distribution of unmeasured heterogeneity is potentially subject to error and can have profound effects on predictions of abundance at unsampled locations. In this article, we develop an alternative approach wherein a Dirichlet process prior is assumed for the distribution of latent abundances. This approach allows for uncertainty in model specification and for natural clustering in the distribution of abundances in a data-adaptive way. We apply this approach in an analysis of counts based on removal samples of an endangered fish species, the Okaloosa darter. Results of our data analysis and simulation studies suggest that our implementation of the Dirichlet process prior has several attractive features not shared by conventional, fully parametric alternatives.

KEY WORDS: Abundance heterogeneity; Bayesian nonparametrics; Empirical Bayes; Model uncertainty; Okaloosa darter; Removal sampling.

1. Introduction

In studies of animal populations, various sampling protocols, such as mark–recapture, multiple-observer sampling, or removal sampling, are used to estimate abundance when animals cannot be captured or detected with absolute certainty. In these surveys the sampling protocol is often repeated at different locations (sample units) that are selected to be representative of the entire population of individuals. The spatial referencing inherent in these surveys can sometimes be exploited to improve estimates of animal abundance and occurrence, particularly in species of individuals that exhibit territorial behavior or other forms of site fidelity. For example, Royle and Dorazio (2006) describe a flexible hierarchical modeling framework for estimating animal abundance and occurrence from spatially referenced counts. In this framework, heterogeneity in abundance or detection may be specified by partitioning the variation among sites into a systematic component, using site-level covari-

ates, and a stochastic component, using simple distributional assumptions.

This approach, though often satisfactory, is vulnerable in some instances to errors in model specification that may be difficult or impossible to assess. For example, with some animal populations, heterogeneity in abundance is thought to exist among sites but the sources of heterogeneity are either poorly understood or unobservable. Many species of salamanders, for example, are highly abundant at some locations and less abundant, or even absent, at others. Site-level covariates, such as soil pH or water temperature, that are thought to be informative measures of habitat at the beginning of a survey, may provide little help in accounting for the differences in abundance among sample locations during data analysis (Jung et al., 2000; Hyde and Simons, 2001; Dodd and Dorazio, 2004). In this situation, an analyst may be tempted to specify the variation in abundance using a simple distributional assumption (say, log normal) for the stochastic component of

activities (e.g., stream discharge, forest cover, erosion control) that would allow the quantity and quality of darter habitat to be calculated at locations throughout the darter's range. However, this work is ongoing, and there is no guarantee that such calculated measures of habitat will prove to be associated with site-level abundances. Therefore, in this article, we develop a model for predicting site-level abundances of darters while accounting for substantial, but unmeasured, sources of variation in abundance that are thought to exist in the population.

The remainder of this article is organized as follows: In Section 2, we develop a hierarchical modeling framework for estimating animal abundance and occurrence from the spatially referenced counts observed in commonly used sampling protocols. This includes a general description of the Dirichlet process prior and a more specific description of its use in our model. Section 3 contains important guidelines for model fitting and prediction of abundances. In Section 4, we fit the model to Okaloosa darter counts and compare the results with fully parametric predictions to illustrate the advantages of the Dirichlet process prior. We also use simulation studies to explore the effects of sample size and the possibility of detecting spurious heterogeneity when none is present. Section 5 contains our conclusions and a description of potential uses of the Dirichlet process prior in other ecological problems.

2. Hierarchical Model

2.1 Sampling Protocol and Data Model

Suppose a particular sampling protocol is applied at each of the n locations (sample units) that constitute a representative sample of the entire population of animals. We assume that these locations and the method and duration of sampling can be selected to ensure that the local abundance of animals within each sample unit is demographically closed during the survey. However, we do not mean to imply that individual animals cannot move among locations once sampling is completed. For our purposes, the sampling frame may be regarded as a metapopulation of animals (Hanski and Simberloff, 1997), a collection of spatially distinct subpopulations (local populations) that occasionally intermix.

Let N_i ($i = 1, \dots, n$) denote the unknown number of animals that are available to be detected or captured within sample unit i , and let \mathbf{x}_i denote the vector of counts observed when a particular sampling protocol is used to detect the N_i animals. As noted by Royle and Dorazio (2006), the sampling protocol determines the precise structure of the observation vector \mathbf{x}_i and its corresponding model. Many common protocols (e.g., capture–recapture, removal sampling, double-observer sampling) induce a dependence among the elements of \mathbf{x}_i that may be specified with a multinomial probability mass function (PMF). The counts observed in simpler protocols, such as independent point counts, can be modeled as a product of binomial PMFs with common index. Without loss of generality, let $[\mathbf{x}_i | N_i, \theta_i]$ denote the conditional distribution of the observed sequence of counts given a multinomial (or binomial) index N_i and given a parameter θ_i , which denotes the probability of observing an individual at site i with the sampling protocol. For example, if removal sampling is used, the number of animals observed in the j th of J_i successive removals is modeled using the multinomial cell probability

$\pi_{ij} = \theta_i(1 - \theta_i)^{j-1}$, and the PMF of the observed counts is

$$f(\mathbf{x}_i | N_i, \theta_i) = \frac{N_i!}{(N_i - x_i)! \prod_{j=1}^{J_i} x_{ij}!} \times \left(\prod_{j=1}^{J_i} \pi_{ij}^{x_{ij}} \right) \left(1 - \sum_{j=1}^{J_i} \pi_{ij} \right)^{N_i - x_i}, \quad (1)$$

where $x_i = \sum_{j=1}^{J_i} x_{ij}$ is the total number of animals observed at site i . In Web Appendix A, we provide the conditional distributions needed to fit this model of removal counts using Gibbs sampling. In general, however, the exact form of the conditional distributions used in Gibbs sampling will depend on the sampling protocol.

2.2 Abundance Model

The abundance model we describe here is motivated by an all-too-common problem: substantial differences in abundance are thought to exist among sites (say, owing to differences in habitat) but the biologically relevant covariates of abundance are either unknown or unmeasured. Therefore, we begin by developing a model of abundance without spatial covariates; later, we describe how this model can be extended to incorporate site-level covariates that are thought to account for some, but not all, of the heterogeneity in abundance.

The Poisson distribution is a natural choice for modeling local abundance; therefore, we assume $[N_i | \lambda_i] \sim \text{Poisson}(\lambda_i)$, where λ_i denotes the mean abundance of animals within the i th sample unit. Royle and Dorazio (2006) showed that stochastic sources of variation in the mean abundance (which imply extra-Poisson variation in local abundances) may be specified using simple distributional assumptions. For example, assuming a normal prior distribution for $\phi_i = \log \lambda_i$ allows the variation in abundance among sites to be specified with only two parameters ($\phi_i \sim N(\alpha, \sigma^2)$). This prior, though often useful, is not always adequate for characterizing spatial variation in local abundances. In fact, the precise form of the prior may be unknown. For example, suppose differences in mean abundance are associated with differences in habitat. If the sample locations include several different habitats, the distribution of local abundances could have multiple modes. Unfortunately, in the absence of habitat covariate measurements, the number of modes and their corresponding levels of abundance are neither known nor easily specified.

The Dirichlet process provides a robust alternative for modeling such unobserved sources of variation in abundance. We describe it briefly as follows. Let $[\phi_i | G] \stackrel{\text{iid}}{\sim} G$, where G denotes a random measure on the parameter space, say, Ω . The Dirichlet process, as defined by Ferguson (1973), is a probability measure defined on the space of all measures G such that for every finite partition B_1, \dots, B_m of Ω , the vector of random probabilities $(G(B_1), \dots, G(B_m))$ has a Dirichlet distribution. Specifically, $[G(B_1), \dots, G(B_m) | \tau, G_0] \sim \text{Dirichlet}(\tau, G_0(B_1), \dots, G_0(B_m))$, where $\tau (> 0)$ is the Dirichlet precision parameter, and G_0 is a base measure defined on Ω . Then for any measurable set B , $E[G(B)] = G_0(B)$. The above Dirichlet process is formally denoted by $DP(\tau, G_0)$.

In practice, the base measure G_0 specifies one’s “best guess” of an underlying model of the variation in ϕ , and τ specifies the extent to which G_0 holds. In our model of mean abundances, the base measure is assumed to be a normal distribution, say $N(\alpha, \sigma^2)$, parameterized by mean α and variance σ^2 ; this yields the following hierarchical model of abundance:

$$\begin{aligned} N_i \mid \phi_i &\sim \text{Poisson}(\exp(\phi_i)); \\ \phi_i \mid G &\stackrel{\text{iid}}{\sim} G; \\ G \mid \tau, N(\alpha, \sigma^2) &\sim DP(\tau, N(\alpha, \sigma^2)). \end{aligned} \tag{2}$$

Loosely speaking, the Dirichlet process may be thought of as a prior on a function space, the space of all prior distribution functions with common support. In this sense the Dirichlet process specifies prior uncertainty in G . While the hierarchy in (2) is conceptually appealing, a marginalization step is often used in actual implementation. Blackwell and MacQueen (1973) proved that by integrating over G the joint distribution of $\phi = (\phi_1, \dots, \phi_n)$ may be factored into a product of successive conditional distributions of the following form:

$$\begin{aligned} \phi_1 \mid \alpha, \sigma^2 &\sim N(\alpha, \sigma^2); \\ \phi_i \mid \phi_1, \dots, \phi_{i-1}, \tau, \alpha, \sigma^2 &\sim \frac{1}{i-1+\tau} \sum_{j=1}^{i-1} \delta(\phi_j) \\ &\quad + \frac{\tau}{i-1+\tau} N(\alpha, \sigma^2), \end{aligned}$$

where $\delta(\phi)$ denotes a degenerate distribution of unit mass centered at ϕ . This factorization implies that ϕ is randomly partitioned into *classes* of distinct values such that elements of ϕ within a class share the same value. Therefore, a random draw from the Dirichlet process $DP(\tau, N(\alpha, \sigma^2))$ may be computed as follows:

$$\begin{aligned} \phi_i \mid \phi_1, \dots, \phi_{i-1} \\ = \begin{cases} \phi_c^* & \text{with probability } \frac{n_{i,c}}{i-1+\tau} \\ \phi \sim N(\alpha, \sigma^2) & \text{with probability } \frac{\tau}{i-1+\tau}, \end{cases} \end{aligned} \tag{3}$$

where ϕ_c^* ($c = 1, 2, \dots, k$) denotes the c th of k distinct values in $\phi_{1:(i-1)} = (\phi_1, \dots, \phi_{i-1})$ and $n_{i,c} = \sum_{j=1}^{i-1} I(\phi_j = \phi_c^*)$ denotes the number of elements in $\phi_{1:(i-1)}$ that equal ϕ_c^* . In other words, a random draw from $DP(\tau, N(\alpha, \sigma^2))$ either equals one of the previous draws or is drawn independently from the base measure $N(\alpha, \sigma^2)$.

The precision parameter τ corresponding to the Dirichlet process prior plays an especially important role in the distribution of ϕ . In (3), note that the probability that ϕ_i differs from all previously drawn parameter values is proportional to τ ; therefore, higher values of τ lead to a higher probability of many unique values (i.e., a higher number of classes k relative to the sample size n). In fact, given our assumption of a normal base measure $N(\alpha, \sigma^2)$, the hierarchical model described by Royle and Dorazio (2006) corresponds to a limiting form of our Dirichlet process mixture (i.e., as $\tau \rightarrow \infty$). Therefore, the Dirichlet process provides a natural extension of the parametric model developed by Royle and Dorazio (2006). We will have more to say about the importance of τ in this context in Section 3.

2.3 Detection Model

Following Royle and Dorazio (2006), we assume that the probability of detection at site i , θ_i , is fixed and that the variation in detectability among sites may be modeled using an exchangeable prior. For example, here we use a beta(a, b) distribution to specify stochastic variation in θ_i . More sophisticated models of detection are possible if site-level covariates are available and are thought to be informative of the variation in detection probabilities among sample locations (Royle and Dorazio, 2006); however, in the situation we address here such covariates are assumed to be unavailable.

3. Model Fitting and Prediction

For many common sampling protocols, the conditional posterior distributions of the model parameters will have nonstandard forms. In the Web Appendix, we provide the conditional distributions needed for our analysis of removal counts in Section 4. In general, we adopt Markov chain Monte Carlo techniques (Robert and Casella, 2004) to generate random observations from these distributions. We chose prior distributions of our model’s hyperparameters ($\tau, \alpha, \sigma^2, a, b$) to specify prior indifference in their magnitudes. Based on preliminary calculations, posterior inferences appeared to be insensitive to our choice of priors with the exception of one parameter, τ , the Dirichlet precision. In particular, we fitted the model using a variety of gamma distributions as priors for τ and discovered that the choice of prior exerts a strong influence on the posterior of τ and other model parameters. In retrospect, this is not really surprising. The conditional posterior density function of τ depends only on the realized number of classes k , the sample size n , and the prior density $\pi(\tau)$ as follows:

$$\begin{aligned} p(\tau \mid \cdot) &\propto p(k \mid \tau, n) \pi(\tau) \\ &\propto \tau^k \frac{\Gamma(\tau)}{\Gamma(\tau+n)} \pi(\tau) \end{aligned} \tag{4}$$

(Escobar and West, 1995). Therefore, Gibbs samples of τ depend only on n and k , allowing $\pi(\tau)$ to exert considerable influence on the posterior. In these circumstances, the desirable feature of robustness that provided our motivation for using a Dirichlet process cannot be attained because inferences are sensitive to the choice of prior on τ . Consequently, we developed an alternative to a fully hierarchical Bayesian specification of the model on τ , which we describe in the next section.

3.1 Empirical Bayes Estimation of τ

Prior sensitivity in estimating τ has been noted in other applications of the Dirichlet process, leading Liu (1996) and McAuliffe, Blei, and Jordan (2006) to suggest an empirical Bayes approach wherein posterior inferences are computed while conditioning on a maximum likelihood estimate (MLE) of τ . Calculating the MLE of τ is computationally intensive, requiring repeated applications of a Gibbs sampler, which we now describe.

The fundamental result to be exploited is Theorem 4 of Liu (1996), which establishes that the MLE of τ satisfies an equivalence between the conditional prior and conditional posterior

means of K (= a random variable for the number of distinct classes). More specifically, Liu proved

$$E(K | n, \hat{\tau}, \text{data}) = E(K | n, \hat{\tau}) = \sum_{i=1}^n \frac{\hat{\tau}}{\hat{\tau} + i - 1}$$

where $\hat{\tau}$ denotes the MLE of τ . This result suggests that $\hat{\tau}$ may be computed as follows:

1. Use Gibbs sampling to generate a sample of the posterior conditioned on a fixed value of τ . Initially this value is just a guess within the range $1/\log n$ to $n/\log n$ (McAuliffe et al., 2006).
2. Approximate the posterior mean $E(K | n, \tau, \text{data})$ by averaging the values of k in the posterior sample: $\bar{k} = \frac{1}{R} \sum_{r=1}^R k^{(r)}$, where $k^{(r)}$ denotes the simulated value of k in the r th draw of a posterior sample of size R .
3. Compute the value of τ that satisfies $\bar{k} = \sum_{i=1}^n \tau / (\tau + i - 1)$.
4. Repeat Steps 1–2 using the value of τ obtained in Step 3 until convergence.

Thus, the MLE of τ is computed by alternating between inference and estimation steps. Inferences for all other parameters (that is, other than τ) are computed by conditioning on $\hat{\tau}$, the value of τ at convergence. It is natural to consider the potential effects of uncertainty in $\hat{\tau}$ on inference of other model parameters. Though some analysts would adopt a prior for τ centered at $\hat{\tau}$, we chose to avoid this double use of the data. Sensitivity analyses (not shown) indicate that this practice does not allow much learning to occur (i.e., prior \approx posterior).

3.2 Prediction of Abundance

Predictions of abundance are often required at one or more unsampled locations. The Dirichlet process mixture provides predictions of abundance that are robust to errors in model specification. Moreover, calculating a draw from the posterior–predictive distribution of abundance is no more difficult than calculating a draw from the Dirichlet process prior (in (3)). For example, let $\tilde{\phi}$ denote the mean abundance (on a logarithmic scale) of animals at an unsampled location. A random draw from the posterior predictive distribution of $\tilde{\phi}$ is computed as follows:

$$\tilde{\phi}^{(r)} | \phi^{(r)} = \begin{cases} \phi_c^{*(r)} & \text{with probability } \frac{n_c^{(r)}}{n + \hat{\tau}} \\ \phi \sim N(\alpha^{(r)}, \sigma^{2(r)}) & \text{with probability } \frac{\hat{\tau}}{n + \hat{\tau}}, \end{cases} \quad (5)$$

where the parenthesized superscript r denotes a random draw from the posterior distribution of model parameters, $\phi_c^{*(r)}$ ($c = 1, 2, \dots, k^{(r)}$) denotes the c th of $k^{(r)}$ distinct values in $\phi^{(r)}$, and $n_c^{(r)} = \sum_{j=1}^n I(\phi_j^{(r)} = \phi_c^{*(r)})$ denotes the number of elements in $\phi^{(r)}$ that equal $\phi_c^{*(r)}$. A random draw from the posterior predictive distribution of animal abundance \tilde{N} at an unsampled location is then computed as follows: $\tilde{N}^{(r)} \sim \text{Poisson}(\exp(\tilde{\phi}^{(r)}))$.

4. Analysis of Removal Counts

4.1 Okaloosa Darters

In July and August of 2005, the entire geographic range of Okaloosa darters was surveyed by collecting removal samples at each of 53 representative locations (Figure 1). At 25 of these locations, darters were detected and removed (using snorkel and dip net) in three successive passes of a 20-m stream reach, as described by Dorazio et al. (2005). At the remaining locations, darters were simply counted in a single pass of each stream reach. Two experienced biologists collected all of the samples, dividing the 53 sites about evenly between them. On average, these biologists required only a few hours to collect darters from each location. Given the short duration of sampling and the strong affinity of this species for its habitat, the local population of darters in each stream reach may safely be assumed to have been demographically closed during the survey.

The number of darters observed in the first pass of each sample location ranged from 0 to 202 fishes, suggesting substantial heterogeneity in local abundances (Table 1). At locations where multiple passes were taken, removal counts declined in successive passes indicating that sampling was successful in depleting the local populations of darters. We fit the model described in Section 2 to the removal counts using Gibbs sampling (see Web Appendix A). The joint posterior distribution of the model’s parameters revealed heterogeneity in both abundance and detection (Table 2). The posterior

Table 1
Number of darters removed at each of 53 locations

Sample location	Removal pass			Sample location	Removal pass		
	1st	2nd	3rd		1st	2nd	3rd
1	118	48	19	28	0	–	–
2	23	5	2	29	38	–	–
3	134	51	21	30	61	18	10
4	50	24	11	31	18	–	–
5	53	15	9	32	1	–	–
6	88	40	17	33	48	–	–
7	21	16	6	34	8	–	–
8	11	6	2	35	26	–	–
9	0	0	0	36	28	–	–
10	40	22	8	37	32	–	–
11	0	1	0	38	21	–	–
12	78	55	33	39	21	–	–
13	108	–	–	40	24	12	8
14	42	27	17	41	17	–	–
15	89	26	11	42	82	–	–
16	50	34	17	43	16	–	–
17	76	49	29	44	30	15	4
18	55	–	–	45	0	–	–
19	26	14	4	46	13	–	–
20	41	21	8	47	21	–	–
21	75	36	29	48	0	–	–
22	34	10	3	49	0	–	–
23	28	8	1	50	67	–	–
24	8	–	–	51	6	–	–
25	6	–	–	52	8	2	2
26	35	–	–	53	202	–	–
27	61	–	–				

Table 2

Posterior summary statistics obtained by fitting the model with Dirichlet process prior to the removal counts in Table 1

Parameter	Median	Mean	95% credible interval
$a/(a + b)$	0.52	0.52	0.47–0.57
$a + b$	34.7	41.8	13.0–114.6
α	3.64	3.58	2.33–4.50
σ	1.73	1.82	1.09–3.08
K	21.0	21.3	16–27

mean probability of detection, estimated as $a/(a + b)$, was 0.52 (95% credible interval = 0.47–0.57), and the mean precision of detection probabilities, estimated as $a + b$, was 41.8 (95% credible interval = 13.0–114.6). Thus, variation in detection probabilities, while present, appears to have been a minor component of the intersite variation in removal counts. In contrast, the joint posterior revealed substantial variation in abundance among sites and also considerable clustering. For example, the MLE of the Dirichlet precision parameter was 12.9, and the posterior mean number of classes was 21.3, less than half the sample size ($n = 53$). It is a coincidence that the posterior mean number of classes is close to 25, the number of “informative” sites where three removal counts were observed. If the analysis had included only the data from these sites, we would have estimated the MLE of the Dirichlet precision parameter to be 10.4 and the posterior mean number of classes to be 13.1, again roughly half the sample size.

A histogram approximation of the posterior–predictive distribution of mean abundance $\tilde{\phi}$ was computed using random draws of (5) and an optimal bin width as described by Wand (1997). The distribution of $\tilde{\phi}$ contains several modes that are associated with different levels of darter abundance (Figure 2). However, the magnitude of these differences appears understated owing to the logarithmic scaling of $\tilde{\phi}$.

Figure 3 illustrates the posterior–predictive distribution of local abundance \tilde{N} , which is defined on an arithmetic scale.

The predictive probabilities associated with $\tilde{N} = 0, 1, \dots, 300$ were computed using Monte Carlo averaging, $\Pr(\tilde{N} = N) = R^{-1} \sum_{r=1}^R g(N | \exp(\tilde{\phi}^{(r)}))$, where $g(\cdot | \lambda)$ denotes the PMF of a Poisson distribution with mean λ . Thus, Figure 3 was produced without any binning of abundances owing to the discrete support of the predictive distribution. The distribution of \tilde{N} is multimodal, containing at least five discernable peaks of abundance that span two orders of magnitude. For example, the probability that an unsampled location contains fewer than five darters is relatively high (0.13); at the opposite end of the distribution, relatively high abundances of darters (100 and 200 fishes per stream reach) also appear to have nonnegligible probability.

As a comparison, we fit a model to the darter counts that is identical to that described in Section 2 except that a normal $N(\alpha, \sigma^2)$ prior was assumed for ϕ . This “reduced” model is included in the class of models described by Royle and Dorazio (2006). The posterior predictive distributions of $\tilde{\phi}$ and \tilde{N} under this model are unimodal (Figures 2 and 3) and do not contain the clusters of abundance that are evident if one assumes a $DP(\tau, N(\alpha, \sigma^2))$ prior. The model based on the normal prior therefore appears to “oversmooth” the heterogeneity in abundance of darters.

An important consequence of this oversmoothing is that the normal prior does not provide satisfactory predictions of total population size. We computed a random draw from the predictive distribution of the total number of Okaloosa darters in the population by adding random draws from the posterior distributions of local abundance associated with the 53 sample locations (i.e., $\{N_i\}$) to 6606 random draws from the posterior–predictive distribution of abundance at an unsampled location (i.e., \tilde{N}), as described in Section 3.2. The number of unsampled stream reaches (6606) was obtained by dividing the total length of streams in the watershed by the length of a stream reach in our sample (20 m). These calculations were repeated 20,000 times to obtain a sample of 20,000 draws from the predictive distribution of total population abundance. There are conspicuous differences between

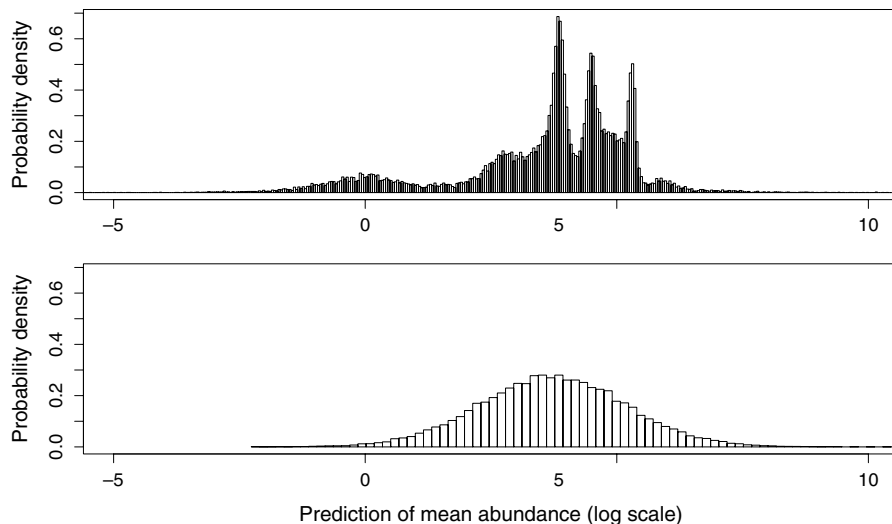


Figure 2. Comparison of posterior–predictive distributions of mean abundance $\tilde{\phi}$ of Okaloosa darters based on assuming a $DP(\tau, N(\alpha, \sigma^2))$ prior (upper panel) or a normal $N(\alpha, \sigma^2)$ prior (lower panel).

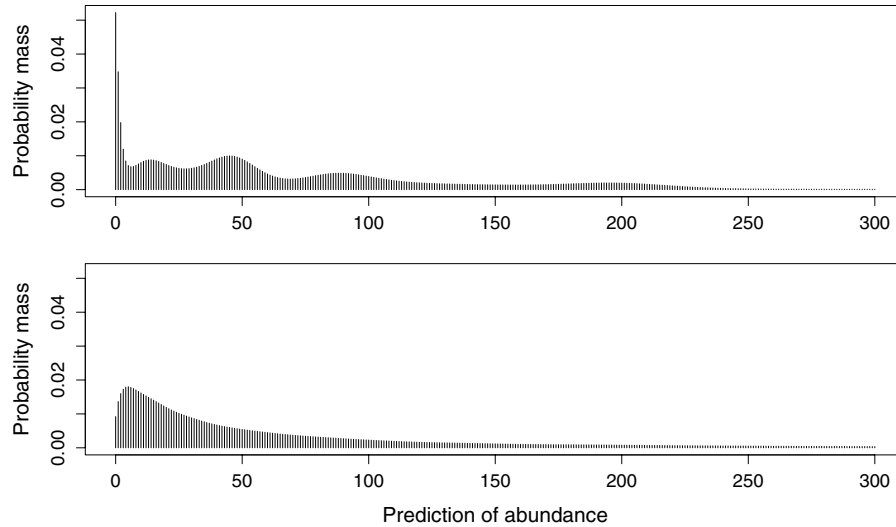


Figure 3. Comparison of posterior–predictive distributions of local abundance \tilde{N} of Okaloosa darters based on assuming a $DP(\tau, N(\alpha, \sigma^2))$ prior (upper panel) or a normal $N(\alpha, \sigma^2)$ prior (lower panel).

the predictive distributions of total abundance obtained by assuming a $DP(\tau, N(\alpha, \sigma^2))$ prior and a normal $N(\alpha, \sigma^2)$ prior (Figure 4). The former has a lower mean (550,000) and is less variable (SD = 8200). In contrast, when a normal prior is assumed, its predictions oversmooth the variation in local abundance, yielding predictions of total population abundance that are higher (mean = 623,000) and more variable (SD = 11,500).

It is worth noting that local abundance N_i and local detection probability θ_i are confounded at more than half of the locations in our sample where only one sample was taken. However, an important benefit of our hierarchical modeling framework is that these parameters can actually be estimated. The locations with three removals provide enough information

about the extent of variation in local abundance and detection to inform the inferences at sites with only 1 pass. For example, recall that we estimated that variation in detection among sample locations is relatively low; therefore, the posterior distributions of θ_i at sites with only one observed count tend to be centered near the posterior mean probability of detection (0.52), and estimates of N_i at these sites essentially amount to upward adjustments of the observed count, i.e., x_{i1}/θ_i .

4.2 Simulated Populations

We completed two simulation studies to examine the robustness of predictions of abundance based on the Dirichlet process prior to errors in model specification. In the first study

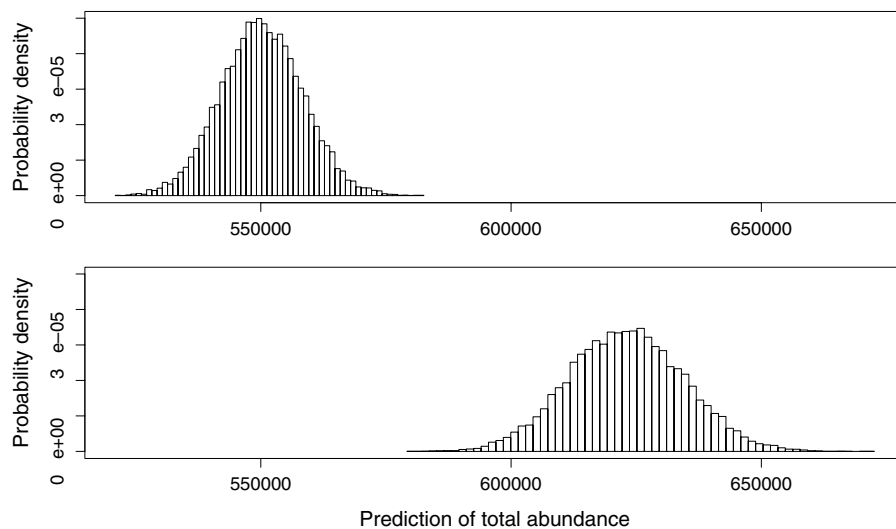


Figure 4. Comparison of posterior–predictive distributions of total population abundance of Okaloosa darters based on assuming a $DP(\tau, N(\alpha, \sigma^2))$ prior (upper panel) or a normal $N(\alpha, \sigma^2)$ prior (lower panel).

we fit the model described in Section 2 to simulated removal counts where the heterogeneity in mean abundance for $n = 10, 25, 53,$ or 100 locations was induced by drawing each element of ϕ from a finite mixture of three normal distributions, $N(\log(4), 0.25^2)$, $N(\log(40), 0.25^2)$, and $N(\log(200), 0.1^2)$, with mixing proportions, $0.6, 0.25,$ and 0.15 , respectively. The purpose of this study was to specify a multimodal distribution of abundances using a model not based on the Dirichlet process and to examine whether the Dirichlet process prior provided a satisfactory approximation of this distribution using samples of various sizes. Given a simulated value of ϕ_i , local abundance N_i was computed as a random draw from a Poisson distribution with mean $\exp(\phi_i)$. We assumed that three successive removals were collected at each site, except in samples of $n = 53$ sites where only one removal count was observed in 28 of the sample locations. This sampling protocol corresponds to that used in the survey of Okaloosa darters. The simulated vector of removal counts was computed as a multinomial draw (using (1)). The site-specific probability of detection and removal in a single pass was computed as a random draw from a beta distribution with mean $a/(a + b) = 0.5$ and precision $(a + b) = 25$. These values are approximately equal to the posterior modes of these parameters estimated in the analysis of darter counts.

We fit the model to the simulated removal counts using Gibbs sampling, as described earlier. Computing the MLE of the Dirichlet precision parameter τ is computationally intensive; therefore, we fit the model to only three samples of counts for each value of n . In samples of $n = 10$ sites, we were unable to estimate any clustering in the predictive distribution of mean abundances. In larger samples the predictive distributions of mean abundance $\tilde{\phi}$ seemed to approximate the number and location of abundance peaks in the three-component normal mixture (Figure 1 in Web Appendix B); however, the level of agreement was certainly not perfect, nor did it appear to vary with sample size. One would expect the level of agreement between the predictive distributions of abundance and the “true” data-generating distribution to improve in larger samples (i.e., $n \gg 100$), but the amount of computation required to fit such samples made this impossible for us to investigate. We note that the predictive density of $\tilde{\phi}$ is also the posterior expectation $E(G | \text{data})$, where $\phi_i \stackrel{\text{iid}}{\sim} G$ and $G \sim DP(\tau, G_0)$ (cf. Section 2.2). When the true underlying distribution being modeled is a location-scale mixture of normals, as in our simulation, consistency of the posterior using a Dirichlet mixture of normals as a prior on the unknown density has been established by Ghosal, Ghosh, and Ramamoorthi (1999). Under certain mild conditions, the convergence rate for estimating mixtures of normal densities using a Dirichlet mixture prior is known to be $\log(n)^\kappa / (n)^{1/2}$, where $\kappa > 1$ is a constant that depends on the type of the true mixture density and on the prior distributions of the parameters of the base measure G_0 (Ghosal and van der Vaart, 2001). Given our choice of priors on α and σ , a product of normal and uniform densities (see Web Appendix A), Theorem 6.2 of Ghosal and van der Vaart ensures that the posterior density is closer to the true mixture of normals (in terms of Hellinger distance metric) with increases in n . That we do not observe it numerically in Figure 1 of Web Appendix B is a consequence of lack of replication.

A second simulation study was completed to examine whether the Dirichlet process prior is prone to overfitting, i.e., does the model detect clusters of abundance when no clustering is present? We fit the model described in Section 2 to simulated removal counts where the heterogeneity in mean abundance for $n = 25$ locations was induced by drawing each element of ϕ from a single normal distribution $N(\log(40), 1.4^2)$. We assumed that three successive removals were collected at each site and calculated the removal counts using random multinomial draws, as described in the first simulation study. We fit the model to three simulated samples of removal counts using Gibbs sampling. The predictive distributions of mean abundance $\tilde{\phi}$ computed for these samples all appeared to be unimodal and to be reasonable approximations of the normal, data-generating distribution (Figure 2 in Web Appendix B); therefore, analyses based on the Dirichlet process prior do not appear to infer clustering when none is present.

5. Discussion

Spatial heterogeneity in abundance is commonly observed in natural populations of animals. In many species the sources of this heterogeneity are either unknown or only incompletely understood. Spatial covariates that are thought to be good predictors of an animal’s habitat before a survey is undertaken often fail to account for much of the observed variation in abundance after the survey is completed. In these circumstances the calculation of accurate predictions of abundance becomes problematic.

Our results suggest that the Dirichlet process prior may provide a satisfactory model in cases where the sources of heterogeneity in animal abundance are unobserved. The Dirichlet process prior is robust to errors in model specification and allows heterogeneity in abundance to be specified in a data-adaptive way. For example, the predictive distribution of abundance based on our Dirichlet process prior can be multimodal owing to its natural potential for identifying clusters (Blackwell and MacQueen, 1973; Escobar and West, 1995). Furthermore, neither the number nor location of clusters needs to be specified explicitly. In contrast, standard parametric prior distributions of abundance, such as the lognormal, cannot capture multimodality because of their shape restrictions. At the opposite end of the heterogeneity spectrum, if the actual distribution of abundances can be approximated by a unimodal distribution, the Dirichlet process prior also adapts to this situation. In fact, the parametric model described by Royle and Dorazio (2006) is a limiting case of our Dirichlet process prior with $\tau \rightarrow \infty$, as described in Section 2.2. Therefore, the Dirichlet process prior is a natural extension of this fully parametric model.

Our implementation of the Dirichlet process prior is easily extended to incorporate site-level covariates that are thought to account for some, but not all, of the heterogeneity in abundance. Assume that the mean abundance parameter ϕ_i at site i equals the sum of a systematic component $w_i' \alpha$ and a stochastic component a_i , where w_i denotes a (possibly vector valued) covariate of abundance observed at site i . Stochastic sources of heterogeneity in abundance may be specified with the Dirichlet process prior by assuming a zero-centered,

normal base measure $N(0, \sigma^2)$; this leads to the following hierarchical model of abundance:

$$\begin{aligned} N_i \mid \alpha, a_i &\sim \text{Poisson}(\exp(\mathbf{w}'_i \alpha + a_i)); \\ a_i \mid G &\sim G; \\ G \mid \tau, N(0, \sigma^2) &\sim DP(\tau, N(0, \sigma^2)), \end{aligned}$$

which is a natural extension of (2). The Gibbs sampler needed to fit this extension obviously requires a different set of full conditionals than those provided in Web Appendix A, but in principle, the model can be fitted. In the analysis of data, one may be tempted to build a model including several covariates (i.e., \mathbf{w}_i and α of several components); however, the dimensionality of the model may quickly get out of hand, and estimation may become unstable, especially if the sample size n is relatively low. Fortunately, the Dirichlet process prior in (2) is able to capture the essential features of heterogeneity without increasing the dimensionality of the problem. This illustrates yet another attractive feature of the Dirichlet process prior.

We anticipate that the Dirichlet process prior will be useful in other problems of ecological inference and prediction. For example, parametric mixing distributions are commonly used to specify latent sources of heterogeneity in detectability of animals in models of site occupancy (Dorazio and Royle, 2005; Royle, 2006) and in models of closed populations without spatial referencing (Dorazio and Royle, 2003). In both cases the assumed mixing distribution can exert considerable influence on the ecologically important estimands (Link, 2003); therefore, the robustness properties of the Dirichlet process prior may prove to be useful in these settings.

6. Supplementary Materials

Web Appendix A referenced in Sections 2 and 4 and Web Appendix B referenced in Section 4.2 are available under the Paper Information link at the *Biometrics* website <http://www.biometrics.tibs.org>.

ACKNOWLEDGEMENTS

Logistical support and funding was provided by Eglin Air Force Base, Department of Defense. Daniel Farrae, Amy Hester, Daniel Holt, Myra Hughey, Jane Jimeian, Melissa Kaintz, Christina Ralph, and Susan Vincent provided assistance in the field. The authors thank the associate editor and two anonymous referees for helpful suggestions in revising an earlier version of this article.

REFERENCES

- Antoniak, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Annals of Statistics* **2**, 1152–1174.
- Blackwell, D. and MacQueen, J. B. (1973). Ferguson distributions via Polya urn schemes. *Annals of Statistics* **1**, 353–355.
- Cifarelli, D. M. and Melilli, E. (2000). Some new results for Dirichlet priors. *Annals of Statistics* **28**, 1390–1413.
- Diaconis, P. and Kemperman, J. (1996). Some new tools for Dirichlet priors. In *Bayesian Statistics 5*, J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (eds), 97–106. Oxford: Oxford University Press.
- Dodd, C. K. and Dorazio, R. M. (2004). Using counts to simultaneously estimate abundance and detection probabilities in a salamander community. *Herpetologica* **60**, 468–478.
- Dorazio, R. M. and Royle, J. A. (2003). Mixture models for estimating the size of a closed population when capture rates vary among individuals. *Biometrics* **59**, 351–364.
- Dorazio, R. M. and Royle, J. A. (2005). Estimating size and composition of biological communities by modeling the occurrence of species. *Journal of the American Statistical Association* **100**, 389–398.
- Dorazio, R. M., Jelks, H. L., and Jordan, F. (2005). Improving removal-based estimates of abundance by sampling a population of spatially distinct subpopulations. *Biometrics* **61**, 1093–1101.
- Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* **90**, 577–588.
- Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics* **1**, 209–230.
- Gelfand, A. E., Kottas, A., and MacEachern, S. N. (2005). Bayesian nonparametric spatial modeling with Dirichlet process mixing. *Journal of the American Statistical Association* **100**, 1021–1035.
- Ghosal, S., Ghosh, J. K., and Ramamoorthi, R. V. (1999). Posterior consistency of Dirichlet mixtures in density estimation. *Annals of Statistics* **27**, 143–158.
- Ghosal, S. and van der Vaart, A. W. (2001). Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities. *Annals of Statistics* **29**, 1233–1263.
- Ghosh, J. K. and Ramamoorthi, R. V. (2003). *Bayesian Nonparametrics*. New York: Springer-Verlag.
- Hanski, I. and Simberloff, D. (1997). The metapopulation approach, its history, conceptual domain, and application to conservation. In *Metapopulation Biology: Ecology, Genetics, and Evolution*, I. Hanski and M. E. Gilpin (eds), 1–11. New York: Academic Press.
- Hyde, E. J. and Simons, T. R. (2001). Sampling plethodontid salamanders: Sources of variability. *Journal of Wildlife Management* **65**, 624–632.
- Ishwaran, H. and James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association* **96**, 161–173.
- Jung, R. E., Droege, S., Sauer, J. R., and Landy, R. B. (2000). Evaluation of terrestrial and streamside salamander monitoring techniques at Shenandoah National Park. *Environmental Monitoring and Assessment* **63**, 65–79.
- Korwar, R. M. and Hollander, M. (1973). Contributions to the theory of Dirichlet processes. *Annals of Probability* **1**, 705–711.
- Link, W. A. (2003). Nonidentifiability of population size from capture-recapture data with heterogeneous detection probabilities. *Biometrics* **59**, 1123–1130.
- Liu, J. S. (1996). Nonparametric hierarchical Bayes via sequential imputations. *Annals of Statistics* **24**, 911–930.
- MacEachern, S. N. and Müller, P. (1998). Estimating mixture of Dirichlet process models. *Journal of Computational and Graphical Statistics* **7**, 223–238.

- McAuliffe, J. D., Blei, D. M., and Jordan, M. I. (2006). Nonparametric empirical Bayes for the Dirichlet process mixture model. *Statistics and Computing* **16**, 5–14.
- Mukhopadhyay, S. and Gelfand, A. E. (1997). Dirichlet process mixed generalized linear models. *Journal of the American Statistical Association* **92**, 633–639.
- Müller, P. and Quintana, F. A. (2004). Nonparametric Bayesian data analysis. *Statistical Science* **19**, 95–110.
- Müller, P., Erkanli, A., and West, M. (1996). Bayesian curve fitting using multivariate normal mixtures. *Biometrika* **83**, 67–79.
- Neal, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics* **9**, 249–265.
- Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*, 2nd edition. New York: Springer-Verlag.
- Royle, J. A. (2006). Site occupancy models with heterogeneous detection probabilities. *Biometrics* **62**, 97–102.
- Royle, J. A. and Dorazio, R. M. (2006). Hierarchical models of animal abundance and occurrence. *Journal of Agricultural, Biological, and Environmental Statistics* **11**, 249–263.
- Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica* **4**, 639–650.
- Sinha, S., Mukherjee, B., Ghosh, M., Mallick, B. K., and Carroll, R. J. (2005). Semiparametric Bayesian analysis of matched case-control studies with missing exposure. *Journal of the American Statistical Association* **100**, 591–601.
- Teh, Y. W., Jordan, M. I., Beal, M. J., and Blei, D. M. (2006). Hierarchical Dirichlet processes. *Journal of the American Statistical Association* **101**, 1566–1581.
- United States Fish and Wildlife Service. (1998). Okaloosa darter (*Etheostoma okaloosae*) Recovery Plan (revised). 42 pp. Atlanta, Georgia.
- Wand, M. P. (1997). Data-based choice of histogram bin width. *American Statistician* **51**, 59–64.

Received May 2006. Revised April 2007.

Accepted May 2007.