North American Journal of Fisheries Management 28:1044–1058, 2008

DOI: 10.1577/M06-250.1

[Article]

Zero-Inflated Modeling of Fish Catch per Unit Area Resulting from Multiple Gears: Application to Channel Catfish and Shovelnose Sturgeon in the Missouri River

ALI ARAB*[1]
Department of Statistics, University of Missouri, Columbia, Missouri 65211, USA

MARK L. WILDHABER
U.S. Geological Survey, Columbia Environmental Research Center, Columbia, Missouri 65201, USA

CHRISTOPHER K. WIKLE AND CASEY N. GENTRY[2]
Department of Statistics, University of Missouri, Columbia, Missouri 65211, USA

Abstract.—Fisheries studies often employ multiple gears that result in large percentages of zero values. We considered a zero-inflated Poisson (ZIP) model with random effects to address these excessive zeros. By employing a Bayesian ZIP model that simultaneously incorporates data from multiple gears to analyze data from the Missouri River, we were able to compare gears and make more year, segment, and macrohabitat comparisons than did the original data analysis. For channel catfish *Ictalurus punctatus*, our results rank (highest to lowest) the mean catch per unit area (CPUA) for gears (beach seine, benthic trawl, electrofishing, and drifting trammel net); years (1998 and 1997); macrohabitats (tributary mouth, connected secondary channel, nonconnected secondary channel, and bend); and river segment zones (channelized, inter-reservoir, and least-altered). For shovelnose sturgeon *Scaphirhynchus platorynchus*, the mean CPUA was significantly higher for benthic trawls and drifting trammel nets; 1998 and 1997; tributary mouths, bends, and connected secondary channels; and some channelized or least-altered inter-reservoir segments. One important advantage of our approach is the ability to reliably infer patterns of relative abundance by means of multiple gears without using gear efficiencies.

The problem of a large proportion of zero values is common with data obtained from ecological studies involving counts of abundance, presence–absence, or occupancy rates (Clarke and Green 1988; Welsh et al. 1996; Berry et al. 2005; Martin et al. 2005). Ignoring and excluding zero values from the analysis of data obtained from field studies can result in loss of important information. For example, when studying abundance or presence–absence of species, a large proportion of zero values might indicate that the species is rare or hard to detect. Rare species and those with low detection probability are common in ecological studies, and standard distributions, such as Poisson, binomial, and negative binomial, do not provide a good fit to zero-heavy data. Zero-inflated modeling, which allows the model to account for a large proportion of zero values, is an appropriate approach to modeling zero-heavy data (Lambert 1992; Hall 2000). Formal testing procedures are available to determine when zero-inflated models (e.g., zero-inflated Poisson) are preferable over standard models (e.g., Poisson) for models without random effects, such as the score test of Van Den Broek (1995) and its extended version (Jansakul and Hinde 2002). For random effect models, general model selection tools, such as Akaike's information criterion and Bayes factors, can be used, depending on the modeling approach. However, when a very high percentage (i.e., more than half) of zero values are present in the data, a clear departure from the Poisson assumption is indicated, and conducting such formal tests is not appropriate.

Two popular models that account for data with excess zeros are the zero-inflated Poisson (ZIP) and the zero-inflated negative binomial (ZINB). The ZIP model is especially useful in analyzing count data with a large number of zero observations, and the ZINB model is more appropriate for cases where an upper bound exists for the response. The ZIP model has been applied to horticulture (Hall 2000), manufacturing (Lambert 1992), and other fields of study, including health operations (Wang et al. 2002),

* Corresponding author: aa577@georgetown.edu
[1] Present address: Department of Mathematics, Georgetown University, Washington, D.C. 20057, USA.
[2] Present address: Hallmark Cards, Kansas City, Missouri 64141, USA.

Received October 26, 2006; accepted January 11, 2008
Published online August 18, 2008

meteorology (Wikle and Anderson 2003), and ecology (Welsh et al. 1996; Martin et al. 2005). Given the large number of zeros that occur in fisheries data, we propose that a ZIP model could be used to determine which factors are related to identifying fish species occurrence and which are related to catch rates of fish.

In 1995, the U.S. Geological Survey (USGS) and the Montana Department of Fish, Wildlife and Parks studied benthic fishes in the warmwater portion of the Missouri River system (Berry and Young 2001; Berry et al. 2005). The Missouri River extends 2,339 mi from southwest Montana to the Mississippi River. Benthic fishes live or feed on the river bottom and are of particular interest because of their sensitivity to changes in habitat. The main goal of the study was to obtain the data needed to improve river management for benthic fishes by evaluating their status, distribution, and habitat associations in the Missouri River.

As is common in fisheries field studies, the data were obtained from multiple gears and include a large proportion of zeros, which makes analysis of the data complicated. Using standard parametric statistical methods on data from each gear separately, Berry et al. (2005) excluded several segments and macrohabitats from the analysis owing to high numbers of zero observations and probable violation of normality and homogeneity of variance assumptions. Berry et al. (2005) were limited by the large percentage of zero observations in the data set, which caused a loss of power due to combining data at larger spatial and temporal scales and analyzing each gear separately. These problems created issues and constraints on usage of the standard classic parametric statistical methods, such as analysis of variance (ANOVA), employed by Berry et al. (2005) to analyze these data.

Our goal was to develop and implement a modeling framework that would allow meaningful ecological interpretations based on the model results and increase the predictive precision of the model. The type of gear, macrohabitat, segment, and year could help identify characteristics explaining where certain species are most likely to populate. Also of interest is the large number of zeros in the data that cannot justifiably be deleted from the analysis. Therefore, these zeros must somehow be accounted for in the modeling process. The chosen model for these data is the ZIP model, which will explain the mean fish count and the zero-inflation probability (i.e., excess zero observations). We consider a Bayesian approach, which provides a flexible modeling framework and is easy to implement for the ZIP model with random effects. Use of random effects in the model is essential, in that it helps account for uncertainties and obtain more valid and comprehensive inference compared to models with fixed effects only.

To demonstrate this approach, the channel catfish *Ictalurus punctatus* and shovelnose sturgeon *Scaphirhynchus platorynchus* were considered. The channel catfish serves as a good starting point for the analysis because it is more of a habitat generalist and is quite common in the Missouri River. The shovelnose sturgeon is of increasing interest to researchers because it serves as a surrogate for the pallid sturgeon *S. albus*, an endangered species (Ruelle and Keenlyne 1994; Bramblett and White 2001). The shovelnose sturgeon is more of a habitat specialist and is much less populous than the channel catfish. The results for these two species provide insight into analyses of similar fisheries data.

Methods

Data Collection

Twenty-six different species of benthic fish (Berry et al. 2005) were included in the Missouri River Benthic Fishes Study (MRBFS). To analyze the data, researchers divided the Missouri River into three zones: the upper or least-altered zone, the middle or inter-reservoir zone, and the lower or channelized zone (Figure 1). The least-altered zone included the lower Yellowstone River. The inter-reservoir zone was characterized by short riverine segments between the six large mainstream reservoirs. The channelized zone was channelized for navigation, and flows were controlled by discharges from upstream dams and by inputs from tributaries. Each zone was then divided into segments, creating a total of 27 segments for the entire river. The river was partitioned into 10–100-km-long segments based on geomorphic (e.g., tributaries, geology) and constructed (e.g., impoundments, channelization, urban areas) features. Of the 27 segments included in the study design, only 15 were sampled during the 3 years of the study (Figure 1). The least-altered zone included segments 3, 5, and 9; the inter-reservoir zone included segments 7, 8, 10, 12, 14, and 15; and the channelized zone included segments 17, 19, 22, 23, 25, and 27 (Figure 1).

The six primary macrohabitats found in the river were identified. These macrohabitats were defined within segments to be “distinctive, repeatable natural and man-made physical features” (Berry and Young 2001). Those six macrohabitats were inside bend, outside bend, channel crossover, tributary mouth, connected secondary channel, and nonconnected secondary channel, as shown in Figure 2. The three macrohabitats associated with bends (inside bend, outside bend, and channel crossover) were placed in the general category bend, resulting in a total of four

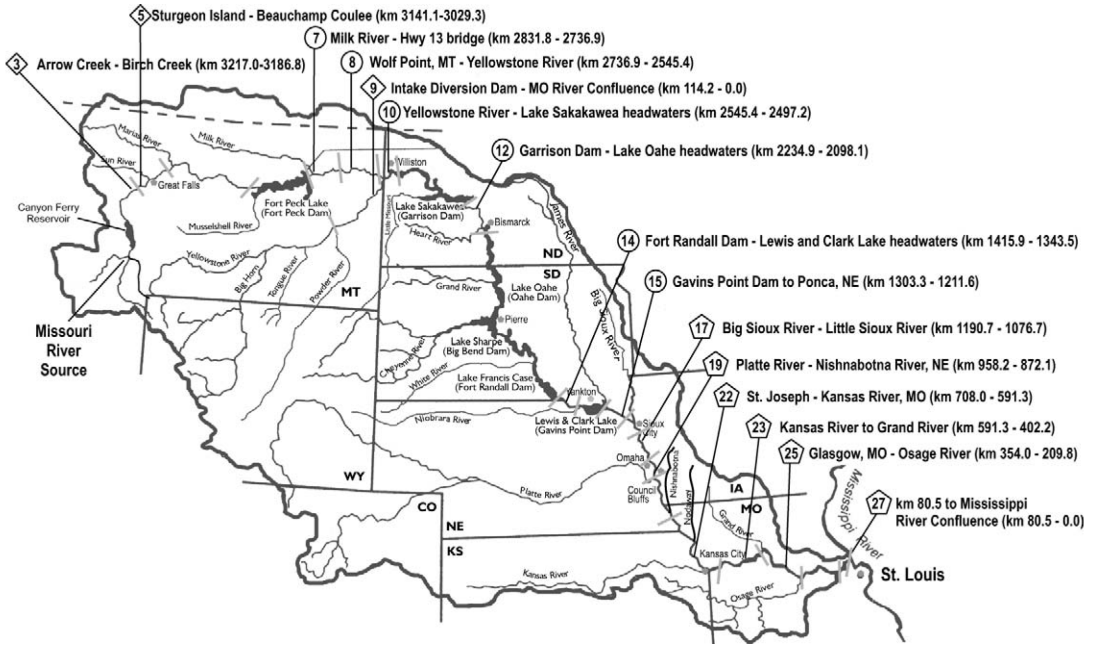

FIGURE 1.—Missouri River Benthic Fishes Study area from Montana to its confluence with the Mississippi River in Missouri (numbers within diamonds indicate least-altered segments, numbers within circles indicate inter-reservoir segments, and numbers within pentagons indicate channelized segments).

different macrohabitats for statistical purposes. The averaging of the three bend macrohabitats was necessary because they were not selected independently (i.e., all three were sampled at each bend).

For each gear deployment, a simple collect-and-count method was used for catching the fish, in which the number of fish, the particular fish species, and the size of area sampled (estimated by the width of the gear and the distance covered as described in Berry et al. 2005) were recorded. Each year, for each segment, researchers could have used as many as four of the five different gears chosen for this study to collect fish within as many as five different randomly chosen occurrences of each macrohabitat type (Table 1). Not all gears were used in all macrohabitats because no gear was considered effective at sampling them all. Of these five gears, four (the benthic trawl, beach seine, drifting trammel net, and electrofishing) are active gears and one (the stationary gill net) is a passive gear. Active gears are the gears that are nonstationary, in that the gear is moved over the sampling area to collect fish. Passive gears are stationary, in that they are located within the sampling area for a specific amount of time to collect fish. In our analysis, we consider only the active gears owing to uncertainty about transformation of sampled area for the passive gear into a scale comparable to the active gears. Finally, not only do the different gears cover different areas, but each is designed differently, making each more prone to catching different species and sizes of fish. All of these factors combined affect the number of fish caught in any particular sample by any particular gear. The process of collecting and counting the fish was

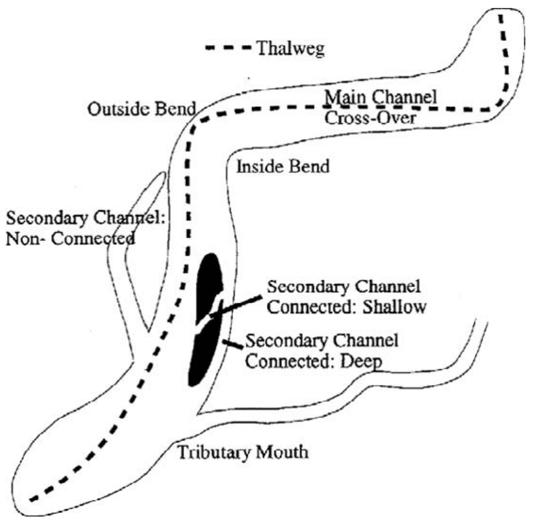

FIGURE 2.—Schematic showing the macrohabitats sampled during the Missouri River Benthic Fishes Study.

TABLE 1.—Habitats and fish collection gears for Missouri River Benthic Fishes Study.

Habitat	Fish collection gear				
	Seine[a]	Gill net[b]	Electrofishing[c]	Trawl[d]	Trammel net[e]
Nonconnected secondary channel	X	X	X		
Connected secondary channel					
Shallow	X				
Deep	X		X	X	X
Channel crossover				X	X
Channel outside bend			X	X	X
Channel inside bend					
Channel border				X	X
Bars	X				
Pools		X			
Steep shoreline			X		
Tributary mouth					
Small		X	X		
Large			X	X	X

[a] Beach seine was 10.7 m long and 1.8 m high with 5-mm mesh.
[b] Gill net was 30.5 m long and 1.8 m high (four 7.6-m-long panels of 19-, 38-, 51-, and 76-mm mesh).
[c] Boat electrofishing was conducted with 5,000-W generator and pulsed DC (two netters with 5-mm mesh dip nets).
[d] Bottom trawl mouth was 2 m wide and 0.5 m high (net was 5.5 m long with inner net of 3.2-mm mesh).
[e] Drifted trammel net was 22.9 m long and 1.8 m deep (25-mm-mesh inner wall and 2,203-mm-mesh outer wall).

repeated for 3 years from 1996 to 1998 (see Berry et al. 2005 for details).

The process of data collection included obtaining multiple subsamples of fish with different gears within each segment over randomly chosen macrohabitats. The data used in this analysis and the analysis done by Berry et al. (2005) were obtained after combining subsample data in the mesohabitat (smaller-scale habitat within a macrohabitat) and then the mesohabitat level, resulting in data at the macrohabitat level. This was necessitated by the varying numbers of subsamples collected at the mesohabitat level within the macrohabitat level. This was done so that each subsample had an equal level of influence on the resulting means. The resulting data contain 1,477 observations, of which a large portion were zeros (see Figure 3). Owing to no or very few nonzero observations, segments 7 and 12 were dropped for channel catfish, and electrofishing, beach seine, and nonconnected secondary channel were dropped for shovelnose sturgeon. This resulted in 1,278 observations for channel catfish and 657 for shovelnose sturgeon. The resulting data included 58% zeros for channel catfish and 61% zeros for shovelnose sturgeon.

Modeling Approach

We consider two different types of zero values in the data: structural zeros and sampling zeros. Structural zeros are the zero values that correspond to species not occurring at that particular site, and sampling zeros correspond to sites where the species occurred but was not detected (Royle 2006). Sampling zeros are unavoidable in habitat analysis owing to the sensitivity of observations to habitat conditions and gear detectability issues. Sampling zeros can generate serious implications in the analysis that can influence the ability of accurate inference from the data (Moilanen 2002; MacKenzie et al. 2003); for example, the recording of false absences can result in serious biases

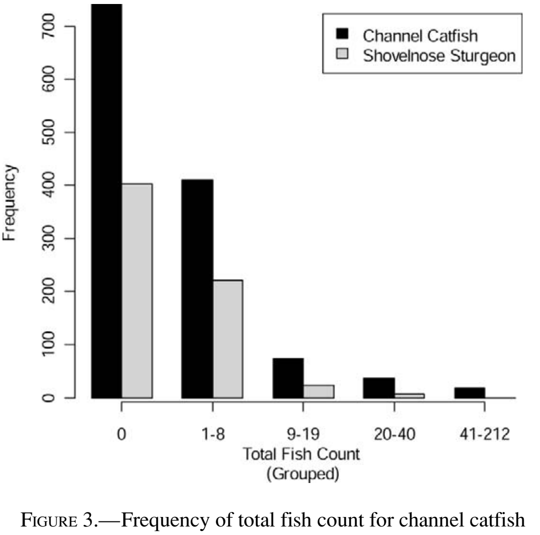

FIGURE 3.—Frequency of total fish count for channel catfish and shovelnose sturgeon. The highest number of channel catfish (212) was caught in a bend with a beach seine, and the highest number of shovelnose sturgeon (40) was caught in a bend with a drifting trammel net.

in the model parameter estimates. Martin et al. (2005) make recommendations on the choice of appropriate modeling approaches to model the source of zeros. In reality, often there exists a mixture of sampling zeros and structural zeros that can be addressed employing a zero-inflated model if we have information on the detection probabilities. A hurdle model or a two-stage modeling approach is common for modeling heavy-zero data when structural zeros are believed to be the only source of zeros (Lambert 1992; Greene 2003). In general, a zero-inflated Poisson model can be used when one is not certain about the nature of the source of zeros.

In this problem, since the data are count data, we consider a Poisson process to account for nonzero observations. Moreover, we consider that the zero inflation is due to both sampling zeros and structural zeros and we consider gears, macrohabitats, and segments as the covariates in estimation of the zero-inflation probability. Gears are probably the primary source of the sampling zeros, although the impact of different gears on generation of sampling zeros might be species specific. The efficiency of gears in detecting species is a function of various habitat-specific factors; in addition, certain gears are more efficient than others for certain species. Macrohabitats and segments are probably the main contributors to the structural zeros. Although the sampling and structural zeros are confounded and hard to separate without information on detection probabilities of gears, it may be possible to infer different probable causes for zeros based on a combination of the ecological prior knowledge and the results of the analysis. We consider a ZIP model that allows us to model both sources of zeros simultaneously by using the indicator variables corresponding to gears, macrohabitats, and segments as covariates for modeling the zero-inflation probability.

ZIP model.—Let y_{ijkl} be the fish count from segment i, macrohabitat j, gear k, and year l. Then we can say that $y_{ijkl} \sim 0$ with probability p_{ijk} and $y_{ijkl} \sim \text{Poisson}(\lambda_{ijkl}a_{ijkl})$ with probability $1 - p_{ijk}$, where $\lambda_{ijkl}a_{ijkl}$ is the Poisson intensity representing the mean number of fish caught, λ_{ijkl} represents the mean catch per unit area (CPUA), and a_{ijkl} accounts for the different areas (or level of effort) involved in each separate measurement. Therefore, we can write this model formally as

$$P(y_{ijkl} = 0) = p_{ijk} + (1 - p_{ijk})e^{-(\lambda_{ijkl}a_{ijkl})} \tag{1}$$

$$P(y_{ijkl} = x) = (1 - p_{ijk})e^{-(\lambda_{ijkl}a_{ijkl})}[(\lambda_{ijkl}a_{ijkl})^x/x!]. \tag{2}$$

The described ZIP model is a mixture distribution of a point mass at 0 (i.e., excess zeros) and a Poisson distribution. A nonzero fish count follows a Poisson distribution with intensity $\lambda_{ijkl}a_{ijkl}$. A zero fish count could either be produced by a zero-generating process (with zero-inflation probability p_{ijk}) or follow a Poisson distribution (with probability $1 - p_{ijk}$). Note that the term $(\lambda_{ijkl}a_{ijkl})^x/x!$ in equation (1) is equal to 1, since $x = 0$.

Both λ_{ijkl} and p_{ijk} can be modeled employing canonical link functions (McCullagh and Nelder 1989):

$$\log_e(\lambda_{ijkl}) = \beta_0 + \mathbf{x}'_{ijkl}\boldsymbol{\beta} \tag{3}$$

$$\text{logit}_e(p_{ijk}) = \gamma_0 + \mathbf{z}'_{ijk}\boldsymbol{\gamma}, \tag{4}$$

where β_0 and γ_0 are random intercepts, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are vectors of random effects, and $\mathbf{x}$ and $\mathbf{z}$ are vectors of covariates of interest with elements representing indicator variables corresponding to the variables gear, segment, macrohabitat, and year. The logit function, commonly used in generalized linear models, is defined as follows (McCullagh and Nelder 1989):

$$\text{logit}(p_{ijk}) = \log_e\left(\frac{p_{ijk}}{1 - p_{ijk}}\right).$$

Note that the covariates in the model correspond to the levels of several categorical variables, which makes interpretation of the intercepts impractical. However, the intercepts are considered random to help account for uncertainties such as sampling errors and possible categorical covariates that were excluded from the analysis. The coefficients of the model corresponding to a specific level of a categorical variable are interpreted as the mean fish CPUA in that level relative to the baseline level (the level of the category set to zero). The choice of baseline level for a categorical variable is arbitrary.

Lambert (1992) employs an expectation–maximization (EM) algorithm (Hartley 1958; Dempster et al. 1977) to obtain the maximum likelihood estimates for the ZIP parameters. Hall (2000) adapts Lambert's methodology to a ZIP model with random effects. Maximum likelihood estimation is also possible by using nonlinear mixed model estimation methods such as PROC NLMIXED in SAS (e.g., see Littell et al. 2006). Other recently developed tools for zero-inflated modeling of count data, which are available in statistical software packages such as Stata (Stata Corporation 2003) and SAS (PROC COUNTREG), provide simple implementation of such modeling approaches for the user. Although the optimization techniques used by most of these methods are some of the best available, convergence problems are often detected for complex models. Here, we consider a Bayesian implementation for a ZIP model with random effects (e.g., Wikle and Anderson 2003; Martin et al. 2005). The Bayesian implementation provides a more flexible and reliable estimation tool. A Bayesian approach provides an easy but concise way to deal with the different sources of uncertainty involved in the problem discussed in this paper. The advantage of the modeling option described in this paper over other available options is easy implementation of a sophisticated statistical model with more valid and comprehensive inference.

Bayesian estimation and Markov chain–Monte Carlo methods.—The Bayesian modeling framework for zero-inflated models is a flexible modeling approach that not only provides a tool for researchers to simultaneously model data from multiple gears with a high percentage of zeros but also enables them to include scientific knowledge or beliefs in the model by assigning prior probabilities to the unknown variables and using data to update these beliefs (Wikle and Anderson 2003). The coefficients in the model are random effects. Furthermore, in the Bayesian framework, inferential statements on model parameters (called credible intervals) and *P*-values on hypotheses are more in line with common sense interpretations (Congdon 2001).

In general, given a sampling distribution $f(x \mid \theta)$, where θ is the unknown parameter of interest, and prior probability $p(\theta)$, Bayesian inference is based on the posterior distribution $p(\theta \mid x)$. Considering that the joint distribution can be decomposed as

$$f(x, \theta) = f(x|\theta)p(\theta).$$

Bayes' rule can be applied to obtain the posterior distribution,

$$f(\theta|x) \propto f(x|\theta)p(\theta),$$

where the integral of $f(x, \theta)$ with respect to θ is the normalizing constant. The posterior distribution is often very complex and the normalizing constant integral cannot be analytically solved. Instead, one can simulate from the posterior distribution and the simulated values used in a Monte Carlo framework to make inferences. Markov chain–Monte Carlo (MCMC) methods are a popular approach to simulating from the posterior distributions. The MCMC methods are a class of algorithms for sampling from probability distributions based on construction of a Markov chain that has the desired distribution as its stationary distribution (Gelfand and Smith 1990) and includes such algorithms as Metropolis–Hastings and the Gibbs sampler (Casella and George 1992; Robert and Casella 2004). Gibbs sampling is the main basis of the freely distributed software WinBUGS (Lunn et al. 2000).

A key issue in implementing the Gibbs sampler (or any other MCMC sampler) is that the number of iterations of the algorithm should be large enough to guarantee that the chain approaches stationarity (i.e., convergence to the target density). Typically, the first 1,000–5,000 iterations are considered the burn-in period of the chain and are thrown out (Congdon 2001). The number of burn-in iterations required can be influenced by the choice of starting value.

In this problem, we consider normal priors for the random effects (including a random intercept), $\beta_i \sim N(0, \tau^{-1})$ and $\gamma_i \sim N(0, \tau^{-1})$, where τ is the precision (considered known in our case) for the normal density (note that $\tau = [1/\sigma^2]$, where σ^2 is the variance of the normal density). Choosing a very small value for the precision (or very large variance) results in a "vague" or "noninformative" prior distribution, a common choice in Bayesian modeling that allows data to guide the analysis (Congdon 2001, 2005). We used 10^{-6} for the value of τ to have a noninformative prior (for shovelnose sturgeon, 10^{-2} was used because of convergence problems in WinBUGS). To sample from the posterior densities, the Gibbs sampler in WinBUGS is used with 50,000 iterations, with 10,000 iterations considered as the burn-in period to guarantee convergence (based on visual inspection of the MCMC chain) to the target posterior distribution (i.e., chain achieves stationarity). To guarantee convergence, three chains are used with different starting values for 1 million iterations and "thinned" using every 20th realization (resulting in 50,000 realizations) and 10,000 iterations as the burn-in period. The thinning procedure helps reduce the autocorrelation between the iterations of the MCMC chain, resulting in improved inference, and also reduces the memory and storage requirements. The WinBUGS code is shown in the appendix.

A visual investigation of the plot of the random variables versus the number of iterations is a common but informal method to check convergence. However, in cases where visual assessment of the MCMC samples is not sufficient to draw conclusions about the convergence of the algorithm, formal tests of convergence, such as the Geweke (Geweke 1992) and Raftery–Lewis tests (Raftery and Lewis 1992), among others, can be used. In our case, a visual assessment of the MCMC samples and the autocorrelation plots of the chains was sufficient to conclude that the algorithm has converged.

We tested the robustness of the ZIP model to excessive zero values by conducting a series of simulations. The goal was to obtain an understanding of the model's limitation to provide reliable estimates under high percentages of zero values. In the Bayesian modeling context, we consider the failure to achieve convergence for the MCMC chain as indicating the model's inability to provide reliable estimates. The simulation was conducted by randomly removing nonzero values from available data on channel catfish and shovelnose sturgeon and by fitting the model to the data obtained by this procedure. We considered four different scenarios for each species: 85, 90, 95, and 99% zero values.

Note that since the independent variables in the model are categorical variables, the parameters correspond to the difference in mean responses for a certain level of the categorical variable relative to the baseline levels (the level of the category set to zero). The levels considered arbitrarily as baseline levels are as follows: segment 27 for segments, tributary mouth for macrohabitats, 1998 for years, and beach seine (channel catfish) and drifting trammel net (shovelnose sturgeon) for gears. The definitions of the indicator variables corresponding to the levels of the categorical variables are arbitrary and, for example, the choice of baseline categories does not affect the overall results of the analysis. The obtained MCMC samples correspond to realizations from the posterior density of the unknown parameters rather than just a point estimate as obtained in classic parametric methods (e.g., maximum likelihood estimation). The nonoverlapping densities were considered as indicating a significant difference among the coefficients of the model. In a Bayesian setting, when using noninformative prior densities, a 95% credible interval for the estimates approximately coincides with a 95% confidence interval for the maximum likelihood estimates. If desired, we can use point estimates (e.g., posterior mean, median, or mode) derived from the posterior density.

Results

For channel catfish, the drifting trammel net significantly (i.e., nonoverlapping 95% credible intervals) increases the zero-inflation probability compared with all other gears, including the baseline (beach seine), meaning that the drifting trammel net is more likely to correspond to excess zeros (Table 2); electrofishing and the benthic trawl significantly decrease the zero-inflation probability compared with the beach seine. Bends and connected secondary channels significantly decrease the zero-inflation probability relative to tributary mouths (baseline) and nonconnected secondary channels. Segments 3, 5, 8, 10, 14, 15, 17, and 19 significantly increase the zero-inflation probability relative to all other segments, including the baseline (segment 27).

The results from the Poisson mean model for channel catfish (Table 2) indicate that the mean CPUA of electrofishing is lower than that of the beach seine and benthic trawl; the drifting trammel net has the lowest mean CPUA (Figure 4). The mean CPUA of 1997 is significantly lower than the mean CPUA of 1998 (Figure 5). Nonconnected secondary channels have the lowest mean CPUA; the mean CPUA of connected secondary channels is significantly lower than that of the baseline (tributary mouths; Figure 6). For segments, mean CPUA in segment 27 is significantly higher than that in all other segments (except for segments 3, 22, and 23; Figure 7; Table 2). The mean CPUA for segments 22, 23, and 27 is significantly higher than that of all other segments, except segment 3.

For the shovelnose sturgeon, the effect of the benthic trawl and drifting trammel net on the zero-inflation probability is not significantly different (Table 3). Bends and connected secondary channels significantly decrease the zero-inflation probability relative to tributary mouths. Segments 8, 9, and 23 are the only segments in which the zero-inflation probability is significantly lower than in segment 27.

The results from the Poisson mean model for the shovelnose sturgeon (Table 3) indicate that mean CPUA is significantly higher for the benthic trawl than for the drifting trammel net (Figure 8). The mean CPUA of 1997 is significantly lower than that of 1998 (Figure 9). The mean CPUAs of bends and connected secondary channels are significantly lower than that of the tributary mouth (Figure 10). The mean CPUA of shovelnose sturgeon in segment 27 is significantly higher than that in segment 8 (Figure 11). The mean CPUAs for segment 8 are significantly lower than the CPUAs for segments 9, 12, 19, 22, 23, 25, and 27 (baseline).

The simulations conducted to test the robustness of the model to zero values indicated convergence problems for both channel catfish and shovelnose sturgeon when the data contained very high percentages of zeros. Convergence problems for channel catfish simulations were apparent only for the case with 99% zero values. However, for shovelnose sturgeon with 90% zero values or higher, the model showed severe lack of convergence and was unable to provide useful inference.

Discussion

The model discussed in this paper allows us to conduct analysis on fisheries data that is impossible or hard to analyze with classic methods such as ANOVA and regression. The approach we present allows the researcher to assess fish population abundance patterns in space and time without disregarding important issues, such as the effect of gear efficiency and detection probability on the presence of sampling and

TABLE 2.—Results for the posterior densities of the model coefficients for the Poisson and zero-inflated Poisson models for channel catfish. The results shown are the mean, SD, median, and lower and upper bounds for the 95% credible interval for the posterior densities of the model random effects. The levels of the categorical variables set to zero to define dummy variables are: tributary mouth (macrohabitat), segment 27 (segment), 1998 (year), and beach seine (gear).

Coefficient	Mean	SD	Lower bound	Median	Upper bound
		Poisson			
Intercept	−0.6873	0.08326	−0.8503	−0.6869	−0.5254
Gear					
Benthic trawl	−4.249	0.04377	−4.334	−4.249	−4.163
Trammel net	−7.66	0.08793	−7.835	−7.66	−7.49
Electrofishing	−4.877	0.04656	−4.969	−4.877	−4.787
Year					
1996	−0.1159	0.04553	−0.2059	−0.1157	−0.02714
1997	−0.2091	0.03524	−0.278	−0.2091	−0.14
Habitat					
Bend	−0.034	0.06699	−0.1641	−0.03439	0.09819
Connected secondary channel	−0.2068	0.07785	−0.359	−0.2072	−0.05426
Nonconnected secondary channel	−0.7207	0.1161	−0.9509	−0.72	−0.4953
Segment					
3	−0.3459	0.143	−0.6343	−0.3428	−0.07388
5	−0.5104	0.09311	−0.6953	−0.5095	−0.3304
8	−1.881	0.3234	−2.572	−1.86	−1.306
9	−0.7694	0.06703	−0.9018	−0.7691	−0.6389
10	−1.042	0.1286	−1.301	−1.04	−0.7969
14	−0.8074	0.1695	−1.152	−0.8033	−0.4865
15	−0.9855	0.09139	−1.167	−0.9845	−0.809
17	−0.4764	0.06325	−0.6004	−0.4765	−0.3524
19	−0.2696	0.07782	−0.4234	−0.2692	−0.1179
22	0.3558	0.05713	0.2439	0.3558	0.4679
23	0.1732	0.05581	0.06401	0.173	0.2828
25	−0.1446	0.05816	−0.2588	−0.1446	−0.03072
		Zero-inflated Poisson			
Intercept	0.3017	0.3373	−0.3612	0.3019	0.9607
Gear					
Benthic trawl	−0.7021	0.2003	−1.096	−0.7014	−0.3117
Trammel net	0.621	0.2208	0.1898	0.6207	1.055
Electrofishing	−1.067	0.205	−1.472	−1.066	−0.6676
Habitat					
Bend	−0.928	0.2396	−1.401	−0.9266	−0.4604
Connected secondary channel	−0.6422	0.2663	−1.166	−0.6409	−0.122
Nonconnected secondary channel	−0.5076	0.3897	−1.281	−0.5041	0.2473
Segment					
3	2.307	0.4165	1.516	2.299	3.149
5	1.664	0.3141	1.053	1.662	2.287
8	3.334	0.4848	2.418	3.321	4.326
9	0.4385	0.2945	−0.139	0.4388	1.015
10	1.521	0.3433	0.8526	1.52	2.2
14	1.964	0.3543	1.281	1.961	2.669
15	0.8561	0.3123	0.2448	0.8564	1.469
17	1.111	0.3332	0.4622	1.109	1.768
19	1.03	0.3362	0.3713	1.03	1.69
22	0.122	0.3284	−0.5248	0.1224	0.7635
23	0.0871	0.3069	−0.5157	0.08743	0.6863
25	0.4331	0.2851	−0.1266	0.4325	0.9916

structural zeros. Additionally, this approach allows for comparison of gear performance among fish species.

The Bayesian approach employed in fitting the ZIP model provides more powerful and comprehensive inference by allowing the use of random effects and a random intercept in the model owing to uncertainties in the data collection procedures (e.g., uncertainty about optimal sampling design, measurement, and sampling errors) and lack of confidence about the involvement of appropriate covariates (e.g., other physical and ecological variables for which data are not available), in addition to its ability to use prior knowledge about

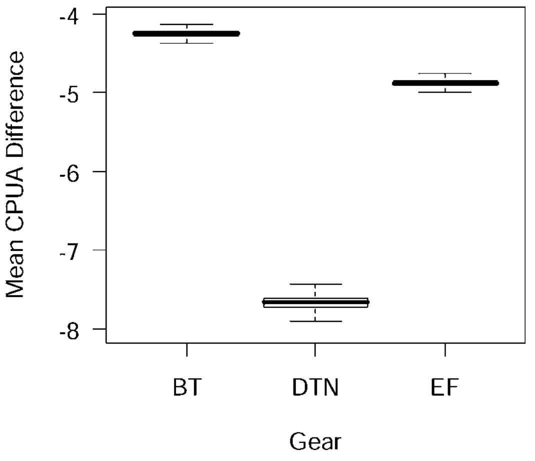

FIGURE 4.—Box plots for the posterior distributions of the coefficients corresponding to the difference in the mean catch per unit area (CPUA) of channel catfish between the beach seine (the baseline gear [set to zero]) and three other gears—the benthic trawl (BT), drifting trammel net (DTN), and electrofishing (EF). Boxplot horizontal lines: lower line = lower quartile (25th percentile), middle line = median (50th percentile), upper line = upper quartile (75th percentile). Boxplot whiskers: lower whisker = smallest observation, upper whisker = largest observation. The box dimensions correspond to the spread (and possible skewness) of the data.

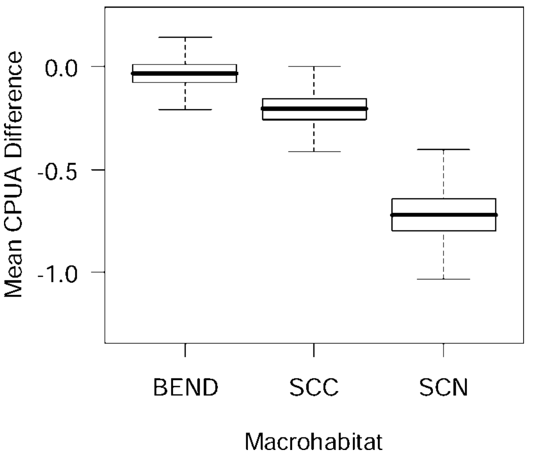

FIGURE 6.—Box plots for the posterior distributions of the coefficients corresponding to the difference in the mean catch per unit area (CPUA) of channel catfish between tributary mouths (the baseline macrohabitat [set to zero]) and three other macrohabitats—bends (BEND), nonconnected secondary channels (SCN), and connected secondary channels (SCC). Boxplot horizontal lines: lower line = lower quartile (25th percentile), middle line = median (50th percentile), upper line = upper quartile (75th percentile). Boxplot whiskers: lower whisker = smallest observation, upper whisker = largest observation. The box dimensions correspond to the spread (and possible skewness) of the data.

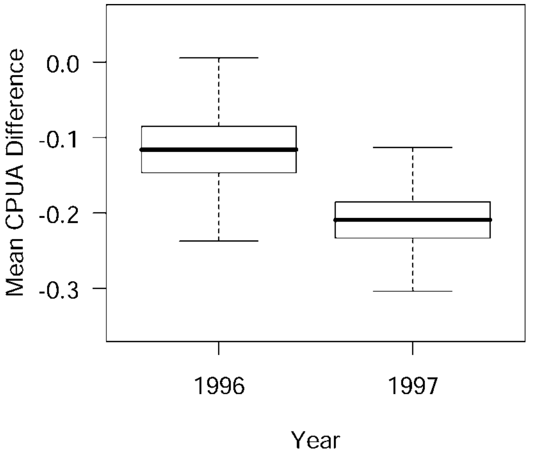

FIGURE 5.—Box plots for the posterior distributions of the coefficients corresponding to the difference in the mean catch per unit area (CPUA) of channel catfish between 1998 (the baseline year [set to zero]) and the years 1996 and 1997. Boxplot horizontal lines: lower line = lower quartile (25th percentile), middle line = median (50th percentile), upper line = upper quartile (75th percentile). Boxplot whiskers: lower whisker = smallest observation, upper whisker = largest observation. The box dimensions correspond to the spread (and possible skewness) of the data.

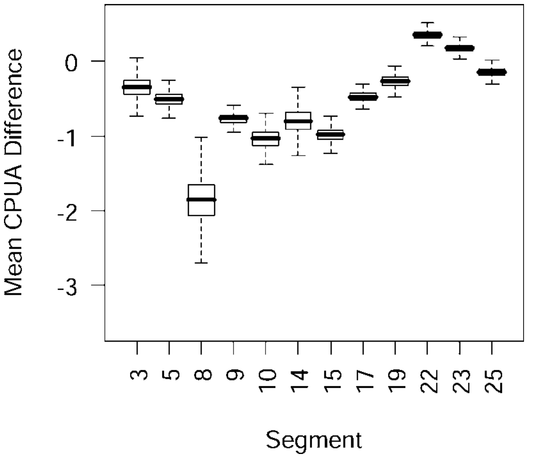

FIGURE 7.—Box plots for the posterior distributions of the coefficients corresponding to the mean catch per unit area (CPUA) of channel catfish between the baseline segment (segment 27) and other segments; segments 3, 5, and 9 are in the least-altered zone, segments 7, 8, 10, 12, 14, and 15 in the inter-reservoir zone, and segments 17, 19, 22, 23, 25, and 27 in the channelized zone. Boxplot horizontal lines: lower line = lower quartile (25th percentile), middle line = median (50th percentile), upper line = upper quartile (75th percentile). Boxplot whiskers: lower whisker = smallest observation, upper whisker = largest observation. The box dimensions correspond to the spread (and possible skewness) of the data.

TABLE 3.—Results for the posterior densities of the model coefficients for the Poisson and zero-inflated Poisson models for shovelnose sturgeon. The results shown are the mean, SD, median, and lower and upper bounds for the 95% credible interval for the posterior densities of the model random effects. The levels of the categorical variables set to zero to define dummy variables are tributary mouth (macrohabitat), segment 27 (segment), 1998 (year), and drifting trammel net (gear).

Coefficient	Mean	SD	Lower bound	Median	Upper bound
Poisson					
Intercept	−7.779	0.3281	−8.46	−7.768	−7.169
Benthic trawl	0.8286	0.1675	0.4498	0.8435	1.118
Year					
1996	−0.0796	0.09467	−0.2661	−0.0791	0.1044
1997	−0.2264	0.07631	−0.3771	−0.2261	−0.0774
Habitat					
Bend	−1.066	0.2152	−1.474	−1.072	−0.6308
Connected secondary channel	−0.7235	0.2272	−1.156	−0.7271	−0.2667
Segment					
3	−0.1597	0.3206	−0.7735	−0.1652	0.4867
5	0.3536	0.2702	−0.1426	0.3415	0.9183
7	−0.1016	0.2756	−0.6076	−0.1137	0.4741
8	−1.192	0.3266	−1.807	−1.2	−0.5269
9	0.6444	0.261	0.1701	0.6314	1.196
10	0.2735	0.2933	−0.2749	0.2651	0.8778
12	1.02	0.2955	0.4656	1.011	1.632
14	−1.129	0.4877	−2.095	−1.117	−0.2074
15	0.1701	0.2967	−0.3889	0.161	0.7771
17	−0.8584	0.4633	−1.728	−0.8712	0.0671
19	0.3978	0.4261	−0.4876	0.4132	1.189
22	0.7975	0.2695	0.3045	0.785	1.364
23	0.7946	0.2644	0.3107	0.782	1.354
25	0.4568	0.3197	−0.1572	0.4507	1.101
Zero-inflated Poisson					
Intercept	1.789	0.6674	0.4705	1.79	3.101
Benthic trawl	0.5424	0.4469	−0.5251	0.6014	1.224
Habitat					
Bend	−1.878	0.542	−2.982	−1.864	−0.8538
Connected secondary channel	−1.488	0.5958	−2.696	−1.475	−0.3515
Segment					
3	0.179	0.7078	−1.196	0.1744	1.593
5	−1.209	0.6524	−2.544	−1.19	0.01841
7	−0.7136	0.6119	−1.936	−0.708	0.4754
8	−2.443	1.765	−6.908	−1.998	−0.1228
9	−1.949	1.048	−4.797	−1.786	−0.5021
10	−1.329	0.7375	−2.87	−1.297	0.01742
12	1.2	0.6431	−0.01349	1.18	2.517
14	−0.6153	1.509	−4.826	−0.2802	1.324
15	−0.3749	0.6062	−1.561	−0.3761	0.8098
17	−1.888	1.899	−6.609	−1.43	0.6814
19	0.594	0.7105	−0.7267	0.5973	1.929
22	−1.133	0.6049	−2.302	−1.14	0.08346
23	−1.412	0.5752	−2.535	−1.415	−0.2682
25	0.0283	0.5636	−1.044	0.01475	1.18

model parameters (e.g., gear efficiency, species-preferred macrohabitats, previously observed population numbers, results from previous studies, and the opinions of other experts). This approach has a very easy and straightforward implementation. Also, this procedure allows us to conduct multiple comparisons to obtain information on gear performance.

Comparing the results obtained by our Bayesian approach for fitting the ZIP model with those from the analyses conducted by Berry et al. (2005), the applicability of our approach in addressing ecological questions is clear. Using this approach, we were able to combine data collected by multiple gears and analyze aggregated data, which the classic statistical methods used by Berry et al. (2005) could not accomplish. Many of our general results from the ZIP model agree with the conclusions of Berry et al. (2005) while providing a more complete, robust analysis (see below; also see Tables 4, 5). This analysis incorporates data from most active gears, years, segments, and macrohabitats and extends comparisons to most levels originally intended by Berry et al. (2005) that were

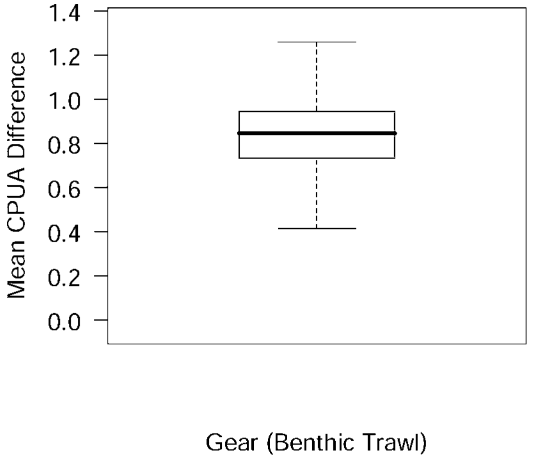

FIGURE 8.—Box plot for the posterior distribution of the coefficients corresponding to the difference in the mean catch per unit area (CPUA) of shovelnose sturgeon between the drifting trammel net (the baseline gear [set to zero]) and the benthic trawl (BT). Boxplot horizontal lines: lower line = lower quartile (25th percentile), middle line = median (50th percentile), upper line = upper quartile (75th percentile). Boxplot whiskers: lower whisker = smallest observation, upper whisker = largest observation. The box dimensions correspond to the spread (and possible skewness) of the data.

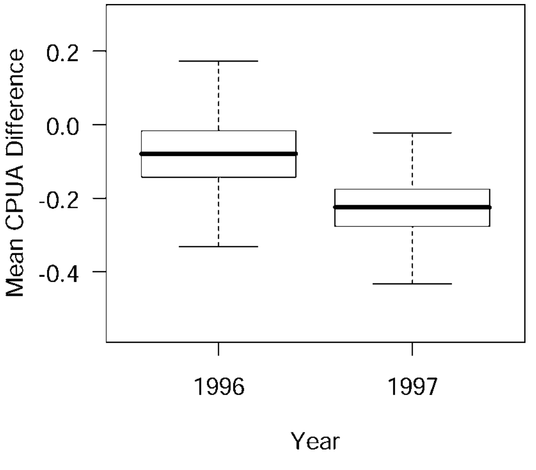

FIGURE 9.—Box plots for the posterior distributions of the coefficients corresponding to the difference in the mean catch per unit area (CPUA) of shovelnose sturgeon between 1998 (the baseline year [set to zero]) and the years 1996 and 1997. Boxplot horizontal lines: lower line = lower quartile (25th percentile), middle line = median (50th percentile), upper line = upper quartile (75th percentile). Boxplot whiskers: lower whisker = smallest observation, upper whisker = largest observation. The box dimensions correspond to the spread (and possible skewness) of the data.

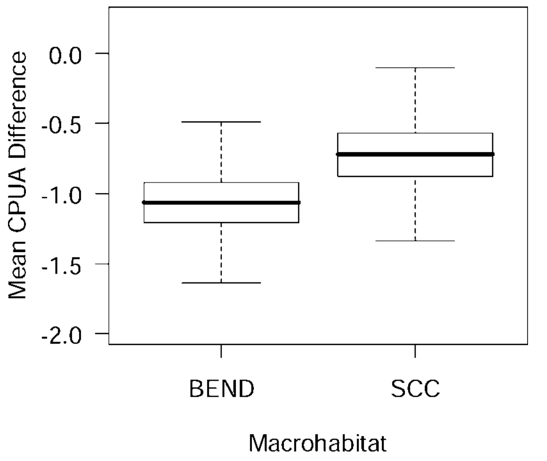

FIGURE 10.—Box plots for the posterior distributions of the coefficients corresponding to the difference in the mean catch per unit area (CPUA) of shovelnose sturgeon between tributary mouths (the baseline macrohabitat [set to zero]) and bends (BEND) and connected secondary channels (SCC). Boxplot horizontal lines: lower line = lower quartile (25th percentile), middle line = median (50th percentile), upper line = upper quartile (75th percentile). Boxplot whiskers: lower whisker = smallest observation, upper whisker = largest observation. The box dimensions correspond to the spread (and possible skewness) of the data.

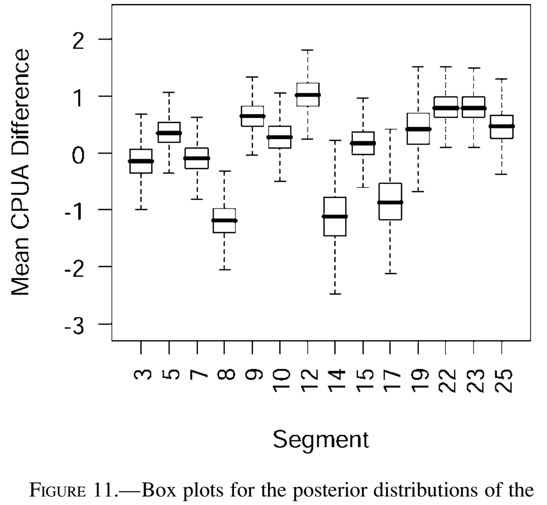

FIGURE 11.—Box plots for the posterior distributions of the coefficients corresponding to the difference in mean catch per unit area (CPUA) of shovelnose sturgeon between the baseline segment (segment 27) and other segments; segments 3, 5, and 9 are in the least-altered zone, segments 7, 8, 10, 12, 14, and 15 in the inter-reservoir zone, and segments 17, 19, 22, 23, 25, and 27 in the channelized zone. Boxplot horizontal lines: lower line = lower quartile (25th percentile), middle line = median (50th percentile), upper line = upper quartile (75th percentile). Boxplot whiskers: lower whisker = smallest observation, upper whisker = largest observation. The box dimensions correspond to the spread (and possible skewness) of the data.

TABLE 4.—Comparison of results from the zero-inflated Poisson (ZIP) model with a previous analysis of the data by Berry et al. (2005) for channel catfish. Abbreviations are as follows: CPUA = catch per unit area, TRM = tributary mouth, SCC = connected secondary channel, CH = channelized, IR = inter-reservoir, LA = least-altered, SCN = nonconnected secondary channel.

Variable	ANOVA (Berry et al. 2005)	Bayesian ZIP
Year	No significant differences for any of year contrasts for data collected with beach seines and benthic trawls. Significant difference for electrofishing data but no trend.	Significant differences (1998 > 1997 based on highest-to-lowest mean CPUA).
Macrohabitat	Substantially more channel catfish in inside bends (a subsection of "Bend"), TRM, and SCC in CH zone.	General agreement with Berry et al. (2005) but extends their results to all segments (Bend > SCN; TRM > SCC > SCN based on highest-to-lowest mean CPUA).
River segment and zone	CH segments have higher mean CPUA than most IR segments. Excluded LA segments 3, 5, 7, 8, 10, 12, and 14.	CH segments have higher mean CPUA than most IR segments. Most LA segments have higher mean CPUA than IR segments.
Gear	Benthic trawl and electrofishing perform best.	All gears can be ranked (highest to lowest) based on mean CPUA as follows: beach seine > benthic trawl > electrofishing > drifting trammel net.

not possible, partly because of the presence of many zeros (this was especially true for shovelnose sturgeon).

The difference between the robustness of the model for data on these two different species, indicated by the simulation studies, is probably related to the spatial distribution of the nonzero values. Notice that because the channel catfish is a habitat generalist, nonzero values are available for this species through the spatial extent of the data; however, because the shovelnose sturgeon is more of a habitat specialist, it is not as uniformly distributed spatially.

Before we discuss results of our model, it must be noted that presentation of biologically meaningful interpretations of the coefficients in the zero-inflation part of the model is difficult without having information to separate the sampling and structural zeros. Note that the zero-inflation probability is mainly a latent process that, although being important for the mechanism of the model, lacks meaningful interpretability. There have been previous attempts to simplify such interpretations (e.g., Lam et al. 2006), but we believe such approaches are appropriate only when useful information is available on detection probability and gear efficiency (also see Martin et al. 2005). Note that the effects of covariates on the zero-inflation probability are related to sufficiency or insufficiency (i.e., lack of fit) of the Poisson model to fit the data.

General Channel Catfish Patterns

Our results employing a Bayesian approach to fitting a ZIP model for channel catfish generally support the results presented by Berry et al. (2005) while extending them to include comparisons that they were unable to make because of statistical test limitations (see Table 4). Our results agree with Berry et al.'s analysis of catch per unit effort (CPUE), in that beach seine, benthic trawl, and electrofishing seem to perform best for the channel catfish with mean CPUA lower for drifting trammel net. Berry et al. (2005) indicate that 56% of total channel catfish catch was collected by electrofishing and beach seine combined but that channel catfish were caught in good numbers in all gear except drifting trammel net. Our results show that the mean CPUA of shovelnose sturgeon was highest in 1998, significantly decreased in 1997, and slightly

TABLE 5.—Comparison of results from the zero-inflated Poisson (ZIP) model with a previous analysis of the data by Berry et al. (2005) for shovelnose sturgeon. Abbreviations are defined in Table 4.

Variable	ANOVA (Berry et al. 2005)	Bayesian ZIP
Year	Significant differences between 1996 and 1997 for data collected with benthic trawl. Significant differences between 1997 and 1998 for data collected with drifting trammel net (no trend).	Significant difference between 1997 and 1998; 1998 had a higher mean CPUA for all gears.
Macrohabitat	Only able to compare "Bend" and SCC (no significant difference detected).	Significant differences for the following: TRM > Bend and SCC (BEND and SCC not significantly different).
River segment and zone	Some IR segments have significantly lower mean CPUA than most CH or LA segments. Included only three CH segments.	Some IR segments have significantly lower mean CPUA than most CH or LA segments. Included all CH segments.
Gear	Drifting trammel net is the only gear with significantly high catch per unit effort.	Benthic trawl performs significantly better than drifting trammel net.

decreased in 1996. Berry et al. (2005) found no significant differences for these contrasts when using beach seines and benthic trawls and significant differences when using electrofishing, but they gave no trends. Berry et al. (2005) concluded that there were substantially more channel catfish in inside bend (a component of bend), tributary mouth, and connected secondary channel macrohabitats than in nonconnected secondary channel macrohabitats in the channelized zone. Our model results extend theirs and show a higher CPUA in connected secondary channels, tributary mouths, and bends for all segments of the river. As did Berry et al. (2005) for CPUE, from our results we conclude that for channel catfish the CPUA for inter-reservoir segments is significantly lower than for channelized zone segments. In addition, we were able to make channel catfish comparisons with least-altered segments that Berry et al. (2005) were unable to make because they excluded segments 3, 5, 7, 8, 10, 12, and 14 owing to lack of nonzero observations. In our analysis, only segments 7 and 12 were excluded owing to no or few nonzero observations.

General Shovelnose Sturgeon Patterns

For shovelnose sturgeon, Berry et al. (2005) note that many planned contrasts were not possible owing to low total catch. Based on their results, 64% of shovelnose sturgeon were caught by drifting trammel net, the only gear that indicated significant differences. Table 5 provides a summary of comparison of our results to the results provided in Berry et al. (2005). Our results show that benthic trawl performs slightly better than drifting trammel net in collecting shovelnose sturgeon. Our results show significantly higher mean CPUA for 1998 compared with 1997. Berry et al. (2005) found significant differences between 1996 and 1997 for benthic trawls and drifting trammel nets and between 1997 and 1998 for drifting trammel nets, but they give no trends. Shovelnose sturgeon were caught in all macrohabitats except nonconnected secondary channels (Berry et al. 2005), which explains why nonconnected secondary channels were excluded from our analysis. Our results show significantly lower mean CPUAs for bends and connected secondary channels than for the baseline (tributary mouths). Berry et al. (2005) could only compare bends to connected secondary channels for the active gears and found no significant differences, while they did show a significantly higher mean CPUE for stationary gill net (the passive gear excluded from our analyses) for bend compared to tributary mouth. Just as Berry et al. (2005) conclude, our results show that shovelnose sturgeon mean CPUA in some inter-reservoir zone segments is significantly lower than in the channelized zone segments or least-altered zone segments; the Berry et al. (2005) analysis includes only three channelized zone segments.

Implications

The Bayesian zero-inflated model not only provides a tool for modeling zero-heavy data using random effects but also allows for paired comparisons among main effects. This feature transforms the present model into a versatile tool for scientists to use in addressing design issues, such as optimal gear selection, and also provides species-specific information on changes of fish abundance in space and time that are essential for monitoring programs.

A useful characteristic of a Bayesian approach to modeling abundance data with excess zeros is that it provides a framework in which to account for prior information wherever possible. Unfortunately, in our analysis we lacked reliable knowledge that could be incorporated into the model as prior information. However, we believe that access to a modeling framework (such as the one presented in this paper) that allows for the inclusion of prior information might motivate fisheries biologists to conduct experiments to obtain such information (such as estimates of gear performance).

Comparisons between standard models for count data (e.g., Poisson, negative binomial) and their zero-inflated versions (ZIP, ZINB) show that for heavy-zero data, the zero-inflated models are a more natural and powerful modeling choice. The reader is referred to Martin et al. (2005) for more details and a formal discussion.

The present model enables the researcher to combine data obtained from different macrohabitats, gears, segments, and years and conduct a single analysis to make useful conclusions for each species. This model can be extended and generalized to cases with multiple species and different layers of spatial and temporal data, such as covariates corresponding to the physical characteristics of different macrohabitats, by including spatial or temporal random effects (Wikle and Anderson 2003).

One important caution to note when using the present model is that it lets us make inferences about combinations of segments, macrohabitats, gears, and years that did not occur (e.g., gears not fishable in certain macrohabitats). In general, this type of inference is not reliable because of lack of data and thus requires extrapolation. However, since the present model includes a random intercept and random effects that account for the uncertainties regarding data collection and inclusion of appropriate covariates, the extrapolation results are more reliable than the classic cases (models with fixed effect only) where such uncertainties are not accounted for.

Finally, another important advantage of the present modeling approach for fisheries biology studies is the ability to reliably infer patterns of relative abundance based on data obtained by multiple gears without using gear efficiencies, which stems from the fact that the zero-inflation probability inference considered in the model adjusts for habitat and species-related differences in performance of gears.

Acknowledgments

The data presented here were collected as part of the MRBFS by the Missouri River Benthic Fishes Consortium (MRBFC). The MRBFC included L. C. Bergstedt, C. R. Berry, P. J. Braaten, D. J. Dieterman, D. L. Galat, C. S. Guy, M. A. Pegg, C. L. Pierce, M. P. Ruggles, L. C. Sappington, D. L. Scarnecchia, T. L. Welker, R. G. White, M. L. Wildhaber, and B. A. Young. Funding for this research was provided primarily by the U.S. Army Corps of Engineers, Missouri River Division, Omaha and Kansas City Districts, with additional financial support from the U.S. Environmental Protection Agency (Regions 7 and 8), National Park Service, U.S. Bureau of Reclamation, USGS, U.S. Fish and Wildlife Service, Iowa Department of Natural Resources, Kansas Department of Wildlife and Parks, Missouri Department of Conservation, and North Dakota Game and Fish Department.

References

Berry, C. R., Jr., M. L. Wildhaber, and D. L. Galat. 2005. Fish distribution and abundance. Population structure and habitat use of benthic fishes along the Missouri and lower Yellowstone rivers, volume 3. U.S. Geological Survey, Cooperative Research Units, South Dakota State University, Brookings.

Berry, C. R., Jr., and B. A. Young. 2001. Introduction to the benthic fishes studies: population structure and habitat use of benthic fishes along the Missouri and lower Yellowstone rivers, volume 1. U.S. Geological Survey, Cooperative Research Units, South Dakota State University, Brookings.

Bramblett, R. G., and R. G. White. 2001. Habitat use and movements of pallid and shovelnose sturgeon in the Yellowstone and Missouri rivers in Montana and South Dakota. Transactions of the American Fisheries Society 130:1006–1025.

Casella, G., and E. I. George. 1992. Explaining the Gibbs sampler. American Statistician 46:167–174.

Clarke, K. R., and R. H. Green. 1988. Statistical design and analysis for a "biological effects" study. Marine Ecology Progress Series 46:213–226.

Congdon, P. 2001. Bayesian statistical modeling. Wiley, Chichester, UK.

Congdon, P. 2005. Bayesian models for categorical data. Wiley, Chichester, UK.

Dempster, A., N. Laird, and D. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society B 39:1–38.

Gelfand, A. E., and A. F. M. Smith. 1990. Sampling-based approaches to calculating marginal densities. Journal of the American Statistical Association 85:398–409.

Geweke, J. 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. Pages 169–193 *in* J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors. Bayesian statistics 4. Oxford University Press, Oxford, UK.

Greene, W. 2003. Econometric analysis, 5th edition. Prentice-Hall, Englewood Cliffs, New Jersey.

Hall, D. B. 2000. Zero-inflated Poisson and binomial regression with random effects: a case study. Biometrics 56:1030–1039.

Hartley, H. 1958. Maximum likelihood estimation from incomplete data. Biometrics 14:174–194.

Jansakul, N., and J. P. Hinde. 2002. Score tests for zero-inflated Poisson models. Computational Statistics and Data Analysis 40:75–96.

Lam, K. F., H. Xue, and Y. B. Cheung. 2006. Semiparametric analysis of zero-inflated count data. Biometrics, 62:996–1003.

Lambert, D. 1992. Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics 34:1–14.

Littell, R. C., G. A. Milliken, W. W. Stroup, R. D. Wolfinger, and O. Schabenberber. 2006. SAS for mixed models, 2nd edition. SAS Institute, Cary, North Carolina.

Lunn, D. J., A. Thomas, N. Best, and A. Spiegelhalter. 2000. WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing 10:325–337.

MacKenzie, D. I., J. D. Nichols, J. E. Hines, M. G. Knutson, and A. B. Franklin. 2003. Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. Ecology 84:2200–2207.

Martin, T. G., B. A. Wintle, J. R. Rhodes, P. M. Kuhnert, S. A. Field, S. J. Low-Choy, A. J. Tyre, and H. P. Possigham. 2005. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. Ecology Letters 8:1235–1246.

McCullagh, P., and J. A. Nelder. 1989. Generalized linear models, 2nd edition. Chapman and Hall/CRC Press, London.

Moilanen, A. 2002. Implications of empirical data quality for metapopulation model parameter estimation and application. Oikos 96:516–530.

Raftery, A. E., and S. Lewis. 1992. How many iterations in the Gibbs sampler? Pages 763–773 *in* J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors. Bayesian statistics 4. Oxford University Press, Oxford, UK.

Robert, C. P., and G. Casella. 2004. Monte Carlo statistical methods, 2nd edition. Springer-Verlag, New York.

Royle, J. A. 2006. Site occupancy models with heterogeneous detection probabilities. Biometrics 62:97–102.

Ruelle, R., and K. Keenlyne. 1994. The suitability of shovelnose sturgeon as a pallid sturgeon surrogate. Proceedings of the South Dakota Academy of Science 73:67–82.

Spiegelhalter, D. J., A. Thomas, N. G. Best, and D. Lunn. 2003. WinBUGS, version 1.4. Imperial College and Medical Research Council Biostatistics Unit, London.

Stata Corporation. 2003. Stata statistical software, release 8.0. Stata Corporation, College Station, Texas.

Van Den Broek, J. 1995. A score test for zero inflation in a Poisson distribution. Biometrics 51:738–743.

Wang, K., K. K. Yau, and A. H. Lee. 2002. A zero-inflated Poisson mixed model to analyze diagnosis-related groups with majority of same-day hospital stays. Computer Methods and Programs in Biomedicine 68:195–203.

Welsh, A., R. Cunningham, C. Donnelly, and D. Lindenmayer. 1996. Modelling the abundance of rare species: statistical models for counts with extra zeros. Ecological Modelling 88:297–308.

Wikle, C. K., and C. J. Anderson. 2003. Climatological analysis of tornado report counts using a hierarchical Bayesian spatiotemporal model. Journal of Geophysical Research 108:STS 12-1–STS 12-15.

Appendix: WinBUGS code for ZIP model

```
model{for (i in 1:n){
tfshcnt[i] ~ dpois(mu[i])
mu[i]<-lambda[i,T[i]]*tarea[i]
# Log-linear model Poisson means
lambda[i,1] <- 0
log(lambda[i,2]) <-
b[1]+b[2]*bend[i]+b[3]*scc[i]+b[4]*scn[i]+b[5]*seg3[i]+b[6]*seg5[i]+b[7]*seg7[i]+b[8]*seg8[i]+b[9]*s
   eg9[i]+b[10]*seg10[i]+b[11]*seg12[i]+b[12]*seg14[i]+b[13]*seg15[i]+b[14]*seg17[i]+b[15]*seg19[i]
   +b[16]*seg22[i]+b[17]*seg23[i]+b[18]*seg25[i]+b[19]*y96[i]+b[20]*y97[i]+b[21]*bt[i]+b[22]*dtn[i]+b[
   23]*ef[i]
# Logistic regression for the zero-inflation probability
logit(P[i,1]) <-
a[1]+a[2]*bend[i]+a[3]*scc[i]+a[4]*scn[i]+a[5]*seg3[i]+a[6]*seg5[i]+a[7]*seg7[i]+a[8]*seg8[i]+a[9]*s
   eg9[i]+a[10]*seg10[i]+a[11]*seg12[i]+a[12]*seg14[i]+a[13]*seg15[i]+a[14]*seg17[i]+a[15]*seg19[i]
   +a[16]*seg22[i]+a[17]*seg23[i]+a[18]*seg25[i]+a[19]*bt[i]+a[20]*dtn[i]+a[21]*ef[i]
P[i,2] <- 1-P[i,1]
T[i] ~ dcat(P[i,1:2]);
# Flat Priors on parameters
for (j in 1:21){a[j] ~ dnorm(0,1.0E-6)}
for (j in 1:23){b[j] ~ dnorm(0,1.0E-6)}
}
```