Spatial Modeling of Red Snapper Shrimp Fleet Bycatch in the Gulf of Mexico

Beatrix Jones

March 12, 2004

1 Introduction

Estimation of bycatch is an important problem in fisheries management. Nichols (2004) and references therein outline the various observer programs that have been implemented to measure the bycatch taken by the Gulf of Mexico shrimping fleet. In an effort to use all data sources relevant to this problem, we would like to augment direct observations of bycatch on shrimping vessels with data from abundance surveys taken by the research vessel Oregon II. We would also like to exploit the spatial structure of the data to improve our bycatch estimates. This report outlines a model designed to accomodate these two goals. However, an MCMC approach to fitting the model has not been successful. We will outline the model, the fitting approach, and some possible remedies here.

In order to properly combine the different data types, we need to characterize the relationship between the abundance and bycatch measurements in a way that properly accounts for:

1. Microscale variation/measurement error for both bycatch and abundance.
2. The zero-enriched character of both types of measurements.

These attributes need to be combined with a a model that accommodates spatial structure.

Our exploratory data analysis suggests the proportion of zero measurements in a locale and the average of the log non-zero measurements there are at least loosely related: Figure 1 shows the relationship between these two quantities for both bycatch and abundance, for bins of size 1-degree latitude by 1-degree longitude with at least 10 observations (all had at least one non-zero observation). Least squares lines are also shown; the proportion of non-zero measurements is a statistically significant predictor of average catch in both cases (p=0.001 for abundance and p=0.013 for bycatch.)

In order to reflect this relationship, we model both types of measurements as functions of an underlying Gaussian process. The probability of a non-zero measurement is produced by a logistic model, and the value of each log non-zero measurement is given by a linear model. (The log catch is used throughout, to stabilize variance and give a symmetric error distribution.)

This approach is in contrast to the "delta distribution" (described in, e.g., Smith 1988) where the proportion of zero observations and the quantity of fish in the non-zero observations are assumed to be unrelated. Our model is also much more flexible than the Poisson or negative binomial model, where the relationship between the proportion of zeroes and the other values observed is tightly defined by the model. Also in contrast to the Poisson and negative binomial, our model easily accommodates the continuous nature of some of our data (resulting from extrapolation from subsampled catches).

2 Model

2.1 Mean process

Let quantities subscripted by i vary across space, by s across season, by j across season/year, and by k across measurements at a given season, year and place. Our Gaussian process Z has the following mean:

$$\mu_{is} + \mu_j$$

In other words, the mean consists of a spatially varying pattern for each season, constant across years, and a mean for each year/season combination, constant across space. The values for μ_{is} are taken from a loess smooth of the non-zero abundance data. These loess smooths are pictured in Figures 2-4. In areas where there are not non-zero abundance measurements, we set μ_{is} to zero, reflecting that our best guess for those areas is the yearly mean. The μ_j are fit in the MCMC procedure described in Section 3.

Figure 1: Proportion of zero observations vs average log(CPUE5) for nonzero observations, for both abundance and bycatch measurements.

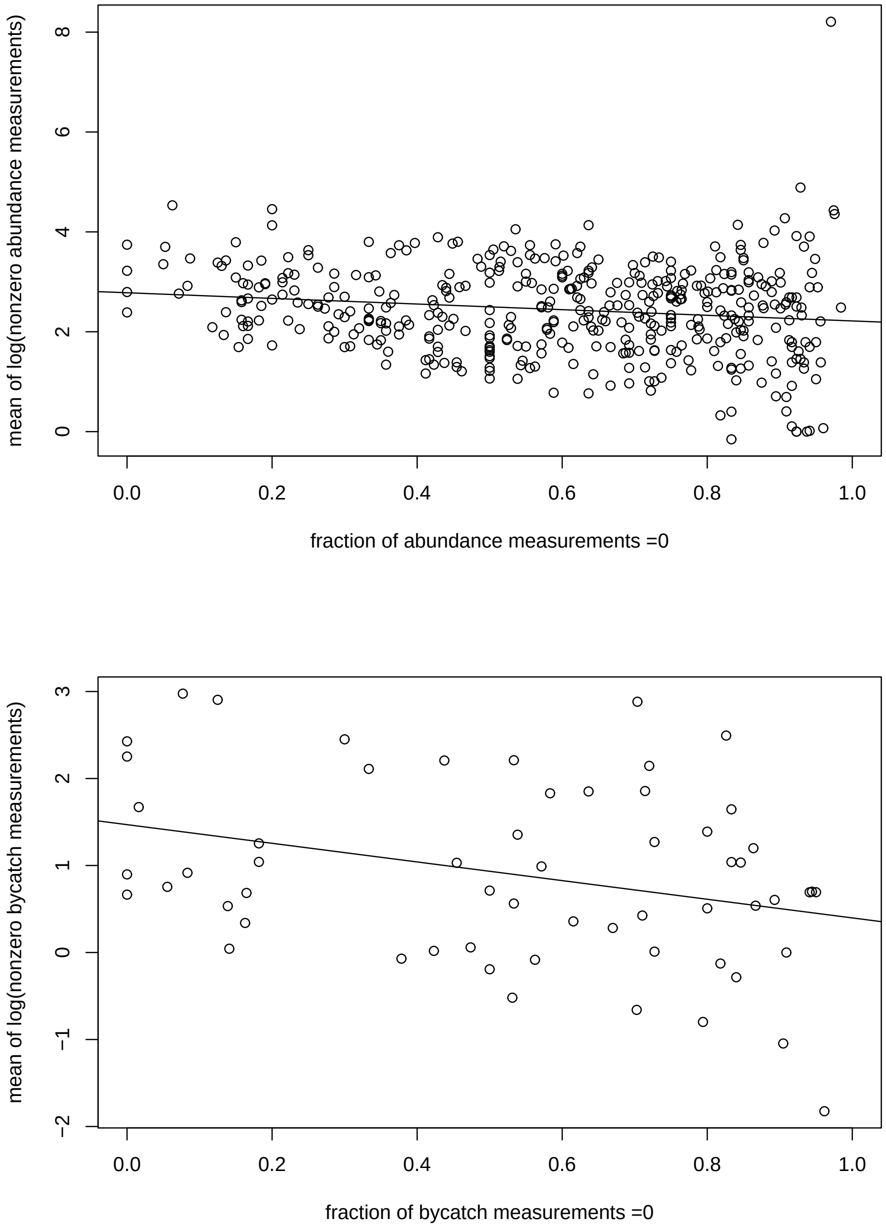
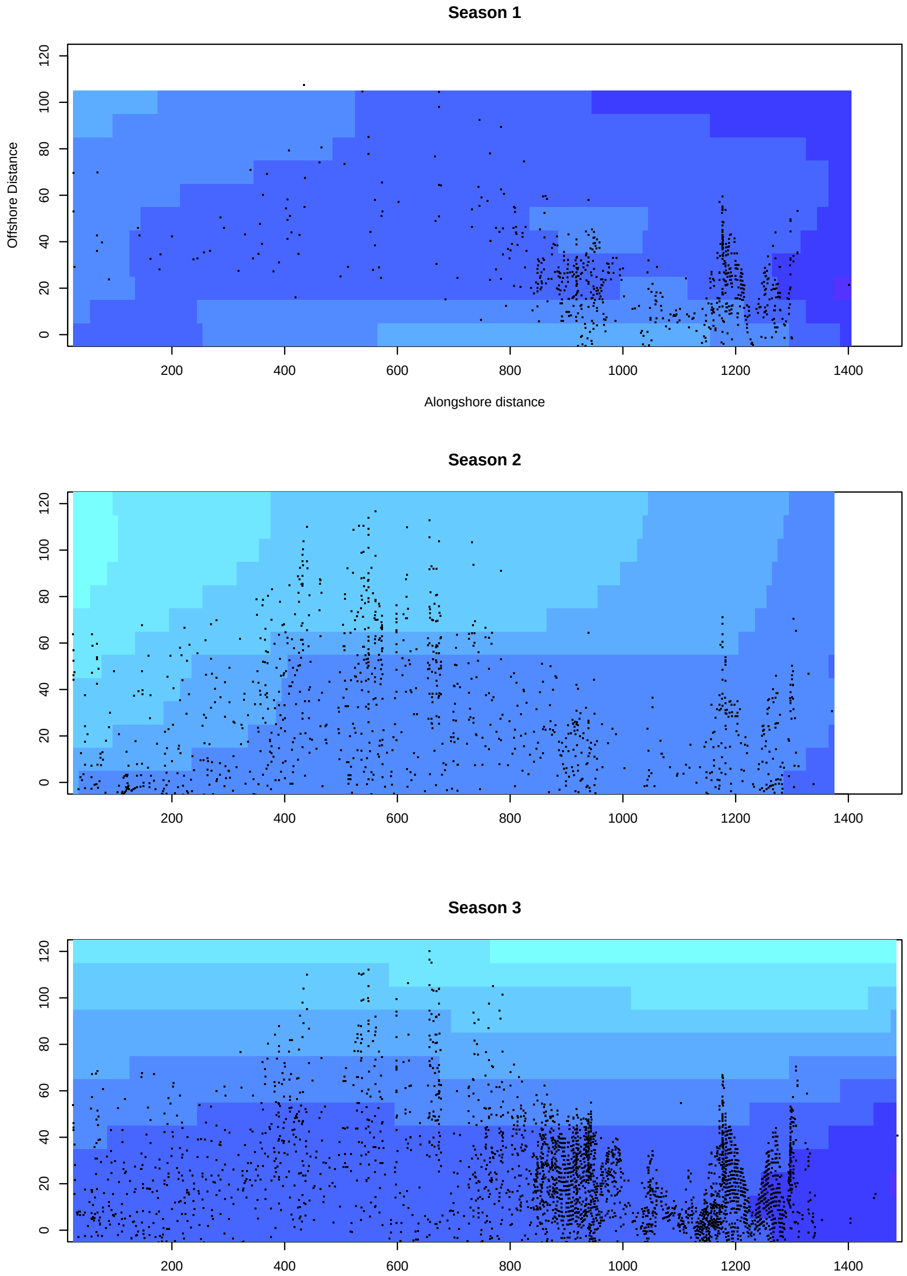


Figure 2: Loess smooth of log(CPUE) for positive abundance measurements. Darker areas had higher measurements. The points represent the measurement locations; values for areas with few points are more uncertain.



Season 1

Season 2

Season 3

2.2 Covariance process

The differences (Z_{ij}'s) of the abundance index from the mean surface are correlated in space. Two points i and i' are likely to have similar deviations from the mean if they are close together. The covariance of these two differences is described by the expression:

$$\sigma^2 \exp(-\sqrt{h^t B h})$$

where h is the vector separating the 2 points, and B is a symmetric, positive definite 2-by-2 matrix that describes a scaling and rotation of the space between the points. This can be thought of as transforming to a new distance measuring scheme where distance corresponds directly to correlation. For instance, if we had a single diagonal shoreline and measured our points on a latitude/longitude scale, we would expect B to rotate the space so that the coordinates represented the offshore and along shore distances, and scale the space so that the separation between points with different offshore distances were "stretched" reflecting their lower correlation compared with points with a similar differential in along shore distance on the original scale. This type of correlation function is called "geometrically anisotropic." (See Ecker and Gelfand 1999.)

In our situation, we are dealing with a curving coastline. We will locate each point with along shore and offshore distances to start with. Thus we expect B to scale these distances, but not to rotate them. This corresponds to the off-diagonal elements of B being zero.

2.3 Relation to measurements

Z represents an underlying "abundance index" which influences the measurements in the following ways:

$$\begin{aligned}
\text{Abundance at location } i = Z_i + \epsilon_i & \qquad \epsilon_i \sim N(0, \tau_1^2) && (1)\\
\text{Bycatch at location } i = \beta_{01} + \beta_{11} Z_i + \epsilon_i & \qquad \epsilon_i \sim N(0, \tau_2^2) && (2)\\
\Pr(\text{Abundance measurement at location } i > 0) &= \text{logit}(\beta_{02} + \beta_{12} Z_i) && (3)\\
\Pr(\text{Bycatch measurement at location } i > 0) &= \text{logit}(\beta_{03} + \beta_{13} Z_i) && (4)
\end{aligned}$$

where $\text{logit}(x) = \exp(x)/(1 + \exp(x))$ These relationships are assumed to be constant across time and space. The fitted linear relationship between the positive bycatch and abundance is in the same spirit as the EBLUP estimation in Jones (2004).

2.4 Priors

Our priors are chosen using a mixture of mathematical convenience and scientific knowledge. Some are parameterized using the total variance of the abundance measurements V. The idea behind this is that the model parameters together explain this total variance, so that any one of them individually should account for only a fraction of it. The "data snooping" involved in this does not follow the separation of prior and data prescribed by a strict Bayesian construction. If this is of concern a large number can be plugged in for V and a diffuse prior will be obtained.

- μ_j (one for each year/season combination): N(0, V).
- $\sigma^2, \text{uniform}(0, V)$;
- $\tau_1^2 \propto 1/\tau_1^2$, a standard improper prior.
- $(\beta_{01}, \beta_{11}, \tau_2^2) \propto 1/\tau_2^2$. This is a standard improper prior for the parameters of a regression model. Its use simplifies the sampling of these parameters, as explained in Section 3.
- β_{02}, β_{12}, β_{03}, β_{13} N(0, 25) , a fairly diffuse normal prior. We do not expect any of these to be large in absolute value.
- B_{11}, B_{12} must be positive; we give them diffuse Gamma priors with mean 1: Gamma(0.1, 0.1). (Variance of 10).

2.5 Computation Time

It was found that it was not computationally feasible to work with the available spatial information at its full resolution. The root of this problem is the large covariance matrices that describe the Gaussian process. Each year has different locations sampled, necessitating a different covariance matrix be computed for every year at each iteration. In addition, several years have upwards of 500 observations. If our ultimate goal is to estimate total bycatch, and effort data is collected on a grid of 210 cells, understanding the bycatch distribution on a finer spatial scale than that defined by these cells is not necessary. Binning the data so that Z's are computed for the bin centers only (and variation within bins is described by the τ parameters) will mean the same covariance matrix can be used for every year, and that matrix will have a moderate size.

3 Fitting Algorithm

As well as fitting the parameters, we sample values for the underlying abundance index Z. We use a component-wise algorithm, where groups of parameters are updated conditional on the others.

- The values of Z for each time period are updated jointly by sampling from the distribution of Z conditional on the parameters and the observed positive data. The metropolis ratio for acceptance or rejection of this proposal then boils down to the likelihood of the 0/1 data between the new and old Z.
- σ^2 is updated with a metropolis step. A new value of the form $\sigma_{\text{current}} * \exp(u)$, with u uniform between -0.1 and 0.1 is proposed.
- τ_1^2 is updated with a Gibbs step sampling from its posterior conditional on the other parameters; this is a scaled inverse χ^2 distribution with the scale factor equal to the mean squared residual (average of $(Y_{ijs} - Z_{ijs} - \mu_j - \mu_{is})^2$) and degrees of freedom equal to the number of positive abundance measurements.
- $\beta_{02}, \beta_{12}, \beta_{03}, \beta_{13}$ are updated jointly with a random walk step (between -0.1 and 0.1) proposed for each.
- $\beta_{10}, \beta_{11}, \tau_2^2$ are updated by sampling from their posterior conditional on Z. They form the parameters of a simple linear regression model of bycatch on the imputed Z's. With the improper prior given in section 2.3 the posterior of τ_2^2 is scaled Inv-$\chi^2(n-k, s^2)$, with $s^2 = 1/(n-k) \times$ (the residual sum of squares). Conditional on τ_2^2, (β_{01}, β_{11}) have a bivariate normal distribution, with mean the ordinary least squares estimate of (β_{01}, β_{11}) and variance $\tau_2^2 V_\beta$. V_β is derived from the predictor matrix X as $V_\beta = (X^T X)^{-1}$. (Gelman *et al.* 1995).
- The μ_j for each year are updated separately by proposing from their posterior distribution given only the positive abundance measurements, the corresponding Z values, and τ_1^2, and then accepting or rejecting this proposal based on the likelihood ratio between new and old values for the other data.

- B_{11}, B_{22} are updated jointly with a metropolis step New values of the form $B_{\text{current}} * \exp(u)$, with u uniform between -0.1 and 0.1 are proposed.

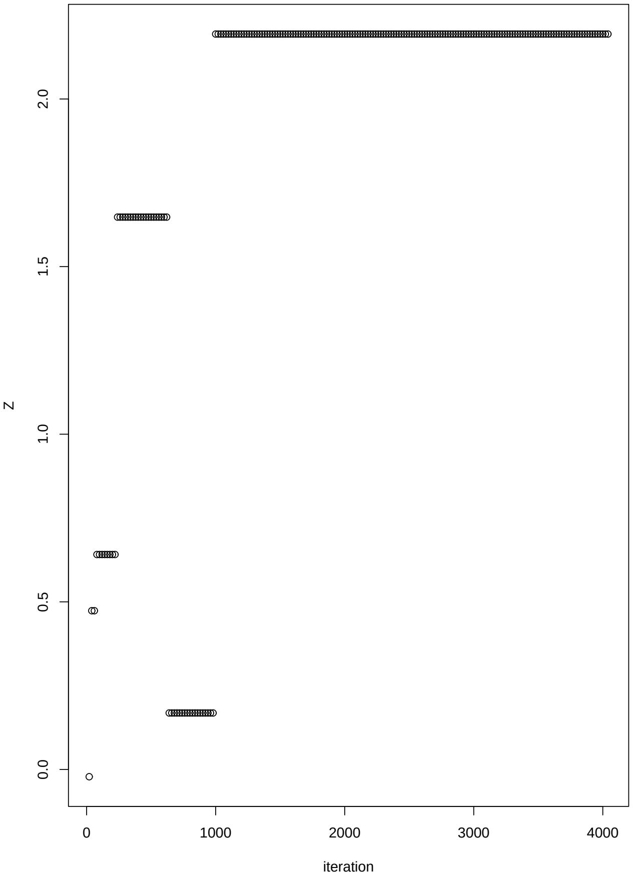
4 Results

The model was applied to all observer data available from 1972-1982, and the Characterization data and control net Evaluation data from both the Regional Research Program (1992-1997) and the summer 1998 BRD evaluations. The data are the same as those used in the GLM 'ALL' updates in the last red snapper assessment. The abundance data used were taken from survey measurements for the same time period (1972-1998).

The results of the algorithm described in the preceding section revealed severe mixing problems for the Z's in years that had a substantial number of positive measurements. A typical sample path for Z at a particular grid point is shown in Figure 3. This picture was produced after approximately 24 hours of computation on a Macintosh G4 Powerbook; while some speedup could be gained using a faster machine, it would not allow enough additional iterations to compensate for the very poor mixing. After some initial movement, almost all the proposed Z's are rejected. Poor mixing for the Z's is a known property of spatial models for binary data (Diggle *et al.* 1998); however it was hoped that the information from the continuous data would allow us to generate from good proposal distributions. In fact, in years where there is substantial continuous data the distribution generated is narrow enough that even small amounts of discordance between the processes implied by the binary and continuous data results in very low acceptance probabilities for the proposed Z's.

An initial attempt at solving this problem was to generate the proposal conditional not just on the continuous data, but also on an artificial data point obtained from applying an inverse logit transformation (and the inverse of the linear relationship between the logit value and Z) to the observed proportion of non-zero observations. (If there were no positive observations, the proportion 1/(2*number of samples) was used.) This was designed to make the proposal distribution more concordant with all the data. The acceptance/rejection probability then depends on the continuous positive observations as well as the binary status of the observation. This procedure only succeeded in pushing back the “sticking point” a few dozen iterations.

Figure 3: The MCMC time series for a particular element of Z.



There are many alterations of the MCMC algorithm that could potentially solve the mixing problem, but none are guaranteed to work. The simplest would be to modify the conditioning scheme that incorporates the binary data by giving the data point representing the binary data higher weight (essentially making it worth several data points.) This weighting may need to be tuned based on the proportion of observations that have positive measurements. Other “ad hoc” manipulations of the mean and or variance of the proposal distribution are also possible. The method involving the biggest departure from the current algorithm (which perhaps also has the greatest chance of success) would be to modify the approach for spatial generalized linear models in Christensen *et al.* to accomodate our hybrid model.

5 Conclusions

While I feel the model developed has attractive properties, I was not able to develop an adequate Markov chain Monte Carlo fitting algorithm. There are various possibilities for modifying the MCMC algorithm, but they are beyond the scope of the current project. We have been able to establish that any spatial treatment of this data should use a binned version of the data for computational efficiency.

References

Christensen, O. F., G. O. Roberts, and M. Sköld (2003) "Robust MCMC methods for spatial GLMM's." Technical Report #23, Centre for Mathematical Sciences, Lund University.

Diggle, P. J., J. A. Tawn and R. A. Moyeed (1998). "Model-based geostatistics (with discussion)," *Applied Statistics* **47:** 299-350.

Ecker, Mark and Alan Gelfand (1999) "Bayesian Modeling and Inference for Geometrically Anisotropic Spatial Data," *Mathematical Geology* **31:** 67-83.

Gelman, A., B. Carlin, H. S. Stern and D. B. Rubin (1995). *Bayesian Data Analysis* Chapman and Hall, London.

Jones, B. (2004). "EBLUP Small Area Estimation for Red Snapper Bycatch from the Gulf of Mexico Shrimp Fleet." Report to the Gulf of Mexico Fishery Management Council.

Nichols, S. (2004). "Some Bayesian approaches to the measurement of shrimp fleet bycatch." Report to the Gulf of Mexico Fishery Management Council.

Smith, Stephen (1988) "Evaluating the Efficiency of the Δ-Distribution Mean Estimator," *Biometrics* **44:** 485-493.