Improving Catch Estimation Methods in Sparsely Sampled Mixed-Stock Fisheries.

Nick Grunloh, Edward Dick, Don Pearson, John Field, Marc Mangel

Abstract

Introduction

Context

Data

- Collection issues
  - funding => nature of sparcity
- Lay down goal modeling goal
  - mean
  - uncertainty

Methods

Data Generating Model

Something something heirarchical poisson model. Something something (Shelton, 2012).

For the purposes of accuratly modeling not only species composition means, but also higher moments of the data, such as species composition variances, it is neccisary to recognize model limitations with respect to over-dispersed data. Amoung the simplest models for count data are the poisson and binomial models. Both models are typically specificed with a single degree of freedom for modeling all of the moments of the data, and thus they rely heavily on their respective data generating processes to accuratly represent higher moments in the data. McCullagh and Nelder (1989, pg. 124) commiserate about the

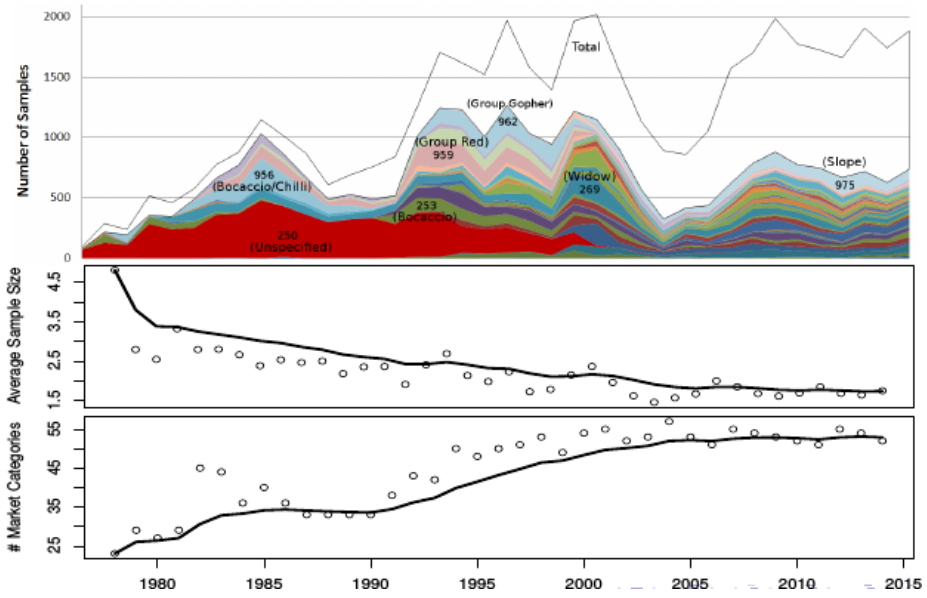


Figure 1: Spase Data

prevalence of over-dispersed data in cluster sampling, and explain ways in which cluster sampling itself may result in over-dispersion.

Extending the Poisson and binomial models to deal with over-dispersion, typically involves adding additional parameters for the purpose of modeling higher moments of the data. The negative binomial (NB) distribution is often used as an over-dispersed extension of the poisson model, since it can be expressly written as an infite mixture of poisson distributions. While the beta-binomial model is typically used to as an over-dispersed extension of the binomial model.

An Example

To discern between these discrete modeling options we consider a small scale example of the Poisson, binomial, negative binomial, and beta-binomial models fit to the port sampling integer weight data from market category 250, in the Montery port complex trawl fishery in 1990. ($anywillwork$) This stratum was visited 38 times by port samplers, collecting a total of 67 cluster samples, resulting in 344 model observations across 21 ($atleast; URCK$) unique species. Each of the above models are fit to these data.

The Poisson and binomial models attempt to model both the mean and residual variance of the data with a single parameter for each species. By definition these models have residual varinces which are tied to the species means. Simply estimating the mean parameters in these cases may not be sufficient to produce

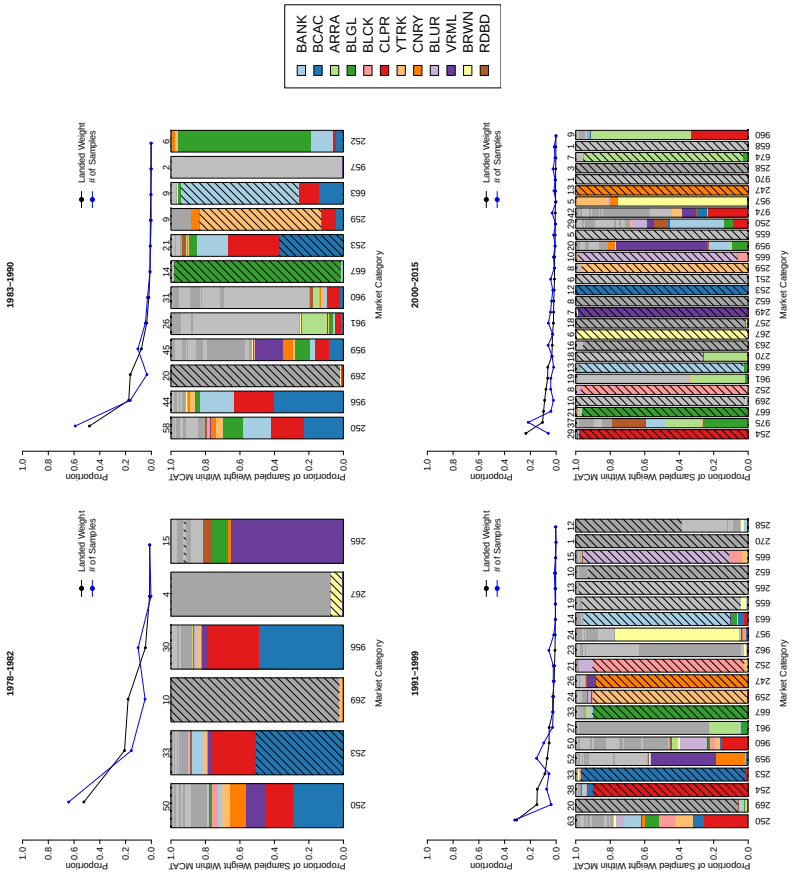


Figure 2: panel

models which predict well.

In contrast, the negative binomial and beta-binomial models estimate an additional parameter which is intended to disentangle the mean and residual variance estimates. Thus it is possible that the negative binomial and beta-binomial models may produce more accurate estimates of both the mean and residual variance.

For each of the above mentioned models The predictive species composition distributions are visualized in Figure(1) as 95% Highest Density Intervals (HDI) (*citations*), plotted on top of the predictive means for each model and the observed species compositions from the data in Figure(1). For brevity we only consider the most prevalent six species in this example (CLPR, BCAC, WDOW, BLGL, ARRA, BANK). Additionally, the MSE, DIC, WAIC, and Bayesian marginal likelyhood model probabilities are computed for each model as measures of model fit as seen in Table(1).

Table(1) show a clear prefernce for the overdispersed models, with the most overall support for the beta-binomial model. This initial result guides the use of the beta-binomial data generating model.

Operationalized Model

For a particular market category, $y_{ijklm\eta}$ is the i^{th} sample of the j^{th} species' weight, in the k^{th} port, caught with the l^{th} gear, in the η^{th} quarter, of year m. As supported by the results in Figure(1) and Table(1), the $y_{ijklm\eta}$ are modeled as observations from a beta-binomial distribution conditional on parameters $\mu_{jklm\eta}$ and $\sigma^2_{jklm\eta}$,

$$y_{ijklm\eta} \sim BB(\mu_{jklm\eta},\ \sigma^2_{jklm\eta}).$$

Where $\mu_{jklm\eta}$ is the stratum level beta-binomial mean weight and $\sigma^2_{jklm\eta}$ is the stratum level residual variance. $\mu_{jklm\eta}$ is related to a linear predictor, $\theta_{jklm\eta}$, via the mean function,

$$\mu_{jklm\eta} = n_{ijklm\eta}\frac{\exp(\theta_{jklm\eta})}{1+\exp(\theta_{jklm\eta})} = n\ \text{expit}(\theta_{jklm\eta}) = n\ \text{logit}^{-1}(\theta_{jklm\eta}).$$

Here $n_{ijklm\eta}$ is the known cluster size for each sample. Additionally, $\sigma^2_{jklm\eta}$ is related to $\mu_{jklm\eta}$ and the overdispersion parameter, ρ, via the following equation,

$$\sigma^2_{jklm\eta} = \mu_{jklm\eta}\Big(1-\frac{\mu_{jklm\eta}}{n_{ijklm\eta}}\Big)\Big(1+(n_{ijklm\eta}-1)\ \rho\Big).$$

ρ is the within cluster correlation. The situation where $\rho \to 1$ represents identical information content amoung replicates within a cluster, with maximal overdispersion relative to the binomial distribution. The situation where $\rho \to 0$ represents totally independent information content amoug replicates within a cluster, and the beta-binomial model approaches the binomial model. ρ explicitly models average overdispersion across all stratum, while $\mu_{jklm\eta}$ gives the model flexiblity at the stratum level through the θ linear predictors,

$$\theta_{jklm\eta} = \beta_0 + \beta_j^{(s)} + \beta_k^{(p)} + \beta_l^{(g)} + \beta_{m\eta}^{(t)}.$$

Firstly, θ includes a reference level intercept (β_0). Secondly, θ is factored among the many strata by additive offsets from β_0 for each of the species ($\beta_j^{(s)}$), port-complex ($\beta_k^{(p)}$), and gear-group ($\beta_l^{(g)}$) categories. Finally year and quarter parameters are indicated generally here inside the $\beta_{m\eta}^{(t)}$ term. Several forms for $\beta_{m\eta}^{(t)}$ are explored each implying a different prior and partial pooling strategies as described in the following section(XX).

Priors

To complete the bayesian formulation of this model priors are expressed in a largly diffuse manner.

$$\beta_0 \propto 1$$
$$\left\{\beta_j^{(s)}, \beta_k^{(p)}, \beta_l^{(g)}\right\} \sim N(0, 32^2)$$

Since the β_0 reference level is choosen arbitarily, with no conception of which values it may take, no restrictions are placed on the value of the intercept. The species ($\beta_j^{(s)}$), port-complex ($\beta_k^{(p)}$), and gear-group ($\beta_l^{(g)}$) offests are assigned diffuse normal priors. The large fixed values of the prior variance hyperparameters produce behavior similar to classical fixed effect models for species, port-complex, and gear- group parameters.

In returning to the time parameter model, $\beta_{m\eta}^{(t)}$, it is useful to consider how the inclusion of predictively superfluous parameters may cause overfitting and weaken model performance through the bias-variance dilemma (Ramasubramanian, K., & Singh, A., 2016). Simply put, the bias-variance dilemma means that model formulation is not simply a bias reduction task, but rather the goal is to formulate models which reduce bias, while jointly minimizing uncertainty. Janyes (2003, pg. 511) describes how the inclusion of estimation bias via the Bayesian methodolgy may produce better performing estimates, more quickly, than unbiased counterparts.
Amoung the simlest ways to see the principle is in the structure of the MSE

performance metric, and how it can be explicitly written to value both estimator bias and variance, as follows.

$$\text{MSE}(\hat{\theta}) = \mathbb{E}\left[\ (\hat{\theta} - \theta)^2\ \right] = \overbrace{\mathbb{E}\Big[\ \Big(\hat{\theta} - \mathbb{E}(\hat{\theta})\Big)^2\ \Big]}^{\text{Var}(\hat{\theta})} + \overbrace{\Big(\ \mathbb{E}(\hat{\theta}) - \theta\ \Big)^2}^{\text{Bias}(\hat{\theta},\theta)^2}$$

Furthermore a model can mimimize bias, without reguard for estimation uncertainty, by including one model parameter to be fit to each observation. These parameter estimates are totally unbiased, however such a model is also predictively usless since each estimated parameter is specifically bound to a particular observation, and thus such a model does not generalize.

For modeling $\beta_{m\eta}^{(t)}$ we consider a spectrum of models which span a wide range of partially pooled models with several different predictive structures as seen below.

(M1)

$$\begin{aligned}
\beta_{m\eta}^{(t)} &= \beta_m^{(y)} + \beta_\eta^{(q)} \\
\beta_m^{(y)} &\sim N(0, 32^2) \\
\beta_\eta^{(q)} &\sim N(0, 32^2)
\end{aligned}$$

(M1) represents a fixed effects model for additive year and quarter parameters. Here each year and quarter receive totally independent and diffuse priors.

(M2)

$$\begin{aligned}
\beta_{m\eta}^{(t)} &= \beta_m^{(y)} + \beta_\eta^{(q)} \\
\beta_m^{(y)} &\sim N(0, v^{(y)}) \\
\beta_\eta^{(q)} &\sim N(0, v^{(q)})
\end{aligned}$$

(M2) estimates two heirarchical variance parameters, $v^{(y)}$ and $v^{(q)}$. $v^{(y)}$ has the effect of partially pooling information among year parameters, while $v^{(q)}$ partially pools information among quarter parameters. The actual degree of pooling among each of the years and quarters is determined by the data. Depending on the posterior distributions of $v^{(y)}$ and $v^{(q)}$, $\beta^{(y)}$ and $\beta^{(q)}$ may be shrunk back toward the common mean for small v or allowed to take largely distinct values in the case of large estimates of the v.

(M3)

$$\beta_{m\eta}^{(t)} = \beta_m^{(y)} + \beta_\eta^{(q)} + \beta_{m\eta}^{(y:q)}$$
$$\beta_m^{(y)} \sim N(0, v^{(y)})$$
$$\beta_\eta^{(q)} \sim N(0, v^{(q)})$$
$$\beta_{m\eta}^{(y:q)} \sim N(0, v)$$

(M3) functions similarly as (M2), in that it has heirarchical partial pooling among both the $\beta_m^{(y)}$ and $\beta_\eta^{(q)}$ parameters, except that it intoduces a two-way interaction term between year and quarter. This interaction term allows estimates for particular quarters to differ from year to year, as opposed to the previous models in which quarters within a year are assumed to be identical from year to year.

Furthermore the $\beta_{m\eta}^{(y:q)}$ are also modeled heirarchically to introduce a single variance parameter, v, shared among all of the $m\eta$ time chunks. Although this interaction term adds many parameters to the model, the shared v parameter functions to shrink extranious $\beta_{m\eta}^{(y:q)}$ estimates back toward the common stratum mean.

(M4)

$$\beta_{m\eta}^{(t)} = \beta_{m\eta}^{(y:q)}$$
$$\beta_{m\eta}^{(y:q)} \sim N(0, v)$$

(M4) simplifies (M3) by excluding year and quarter main effects. This leaves all temporal information in the data to be modeled solely by the quarterly $\beta_{m\eta}^{(y:q)}$ interaction terms. This model reqresents more opportunity for partial pooling through time than (M3), as fewer time parameters are introduced. Furthermore all of the $\beta_{m\eta}^{(y:q)}$ are heirarchically pooled back toward a single common stratum mean via the single shared variance parameter, v.

(M5)

$$\beta_{m\eta}^{(t)} = \beta_{m\eta}^{(y:q)}$$
$$\beta_{m\eta}^{(y:q)} \sim N(0, v_\eta)$$

(M5) is largely the same as (M4), but it represents slightly less potential partial pooling through its heirarchical prior variances, v_η, on $\beta_{m\eta}^{(y:q)}$. Here interaction terms are allowed to partially pool interactions across years, within a common quarter, but since each quarter is assigned a separate variance parameter no pooling is possible between quarters.

(M6)

$$\beta_{m\eta}^{(t)} = \beta_{m\eta}^{(y:q)}$$
$$\beta_{m\eta}^{(y:q)} \sim N(0, v_m)$$

(M6) follows the same idea as (M5), however here interaction terms are allowed to partially pool interactions within a common year, across the quarters of that year, but not between years. (M6) often involves fitting slightly more parameters than (M5) because, at least in this setting, it is typical to model more than four years of data at once.

Heirarchical variance parameters are estimated from the data. As the above models learn the posteriors of the hierarchical variance parameters, it effects the degree of shrinkage as well as the effective number of parameters held within the respective heirarchies (Gelman, 2014). To achieve this, each variance parameter must itself be assigned a prior to be estimated. For all of the heirarchical variance parameters included in the above models v is assigned a diffuse $v \sim IG(1,\ 2\times10^3)$ prior.

Finally the overdispersion parameter, ρ, is first logit transformed to have values spanning the entire real number line and assigned the prior $\text{logit}(\rho) \sim N(0, 2^2)$. The $N(0, 2^2)$ prior is indeed a symmetric, and far reaching, prior when back transformed to the unit interval. To notice this, it is helpful to realize that the central 95% inverval for a $N(0, 2^2)$ (i.e. 0 ± 3.91), includes almost the entirety of the back transformed unit interval (i.e. 0.5 ± 0.48).

Species Composition Prediction

Estimating model (M4) in a Bayesain way gives access to the full posterior distribution of all of the parameters of the model. It is useful to emphasize that in the Bayesian setting, these parameters have full distributions, and they are typically handeled as a large number of samples from the joint posterior distribution of the parameters. Once the posterior sampling is complete, this simplifies parameter mean and variance estimation since the required moments are simply obtainted by computing the desired moments from the posterior samples. Additionally the fact that the parameters are full distributions, means that any functions which contain, or are derived from, parameters are themselves random variables with the function representing a random variable transformation.

To obtain predicted species compositions from this model, first consider the posterior predictive distribition of sampled weight for a particular stratum.

$$p(y^*_{jklm\eta}|y) = \iint \text{BB}\Big(y^*_{jklm\eta}|\mu_{jklm\eta}, \sigma^2_{jklm\eta}\Big) P\Big(\mu_{jklm\eta}, \sigma^2_{jklm\eta}|y\Big) d\mu_{jklm\eta} d\sigma^2_{jklm\eta}.$$

Here BB is the data generating beta-binomial distribution for a predictive observation and $P(\mu_{jklm\eta}, \sigma^2_{jklm\eta}|y)$ is the posterior distribution of the parameters given the observed data. Integration of the parameters, $\mu_{jklm\eta}$ and $\sigma^2_{jklm\eta}$, is done by monte carlo integration to obtain samples from the predictive distribution, $p(y^*_{jklm\eta}|y)$, for sampled weights in the $jklm\eta^t h$ stratum.

Obtaining predictive species compositions from predictive weights amounts to computing the following transformation,

$$\pi^*_{jklm\eta} = \frac{y^*_{jklm\eta}}{\sum_j y^*_{jklm\eta}} \quad y^*_{klm\eta} \neq 0.$$

Here $\pi^*_{jklm\eta}$ is the models representation of the observation level species composition for species j in the k^{th} port, caught with the l^{th} gear, in the η^{th} quarter, of year m.

Model Exploration & Averaging

Presently, stratum with deminishingly small sample sizes are managed by an ad-hoc "borrowing" protocol, outlined in Pearson and Erwin (1997). The protocol for pooling data across port complexes calls for spatial pooling only when forced to fill holes brought about by unsampled strata. Naturally, such a protocol introduces a bias in speecies compositions which depends on the availiability of data in each stratum and thus makes comparisions between periods with pooled and unpooled data inconsistent with eah other. Furthermore, the current ad-hoc "borrowing" protocol makes it difficult to know exactly when "borrowing" has occured.

Handeling these data as heirarchical models allows the models described in section (XX) to avoid the ad-hoc "borrowing" protocol used in Pearson and Erwin (1997). The heirarchical structure of the model, in combination with the Bayesian predictive framework, allows holes in the data to be filled with posterior predictive distributions for any unobserved strata.

Despite the benefits of modeling these data as Bayesian heirarchical models, port sampling data still remains sparce. Given the degree of sparcity in these data it is certainly possible that models which consider an additional degree of data pooling between port complexes may offer predictive benefits. In exploring strategies for pooling data across space it is necessary to formalize the port complex pooling scheme in a way which provides a mathematically understandable and scalable structure to build upon while minimizes issues of inconsistent comparisions between statum.

Given the categorical nature of port complex variables, the typical heirarchical regularization priors amoung port complexes are not appropriate. Rather, we frame port complex pooling as a model uncertainty problem, in which it is

assumed that some degree of port complex pooling is appropriate, but the exact degree of pooling, and particular partitioning of the pooled port complexes are not known.

Port complex pooling is achieved by repeatedly fitting model M5 with different partitionings of the port complex variables within a particular market category and modeling time period. This model exploration exercise explores the possible ways to produce groupings of the existing port complexes so as to discover predictively useful partionings of the port complexes. Insisting that the port complex groupings be partitions of the availiable port complexes provides a well defined mathematical structure for exploring the space of pooled port complexes, additionally due to the fact that partitions are stationary in time provides consistency within the modeled period.

The size of the space of possible pooled models is in the setting is well defined in terms of the size of the set of items to be partitioned, K, as given by the Bell numbers (B_K),

$$B_K = \sum_{\hat{k}=0}^{K} \frac{1}{\hat{k}!} \left(\sum_{j=0}^{\hat{k}} (-1)^{\hat{k}-j} \binom{\hat{k}}{j} j^K \right).$$

In the case of California the set of items to be partitioned is the set of port complexes in California, of which there are $K = 10$, implying a grad total of $B_{10} = 115975$ ways of partitioning the port complexs in California in each market category and modeled time period. The brute force model selection strategy of computing all 115975 of these partitionings strategies is computationally infeasible. However, not all pooling schemes represent biologically relevant models. For example, perhaps it is reasonable to pool only among adjacent ports (i.e. no disontinuities between port complex in space), or perhaps it is reasonable to assert that biologically similar regions can only extend across a small number of ports.

Here only adjacent port poolings are considered, such that the maximum size of a port complex grouping is three port complexes. These are the only constraints that are enforced on port complex partitions here, although many other constraints may in theory be chosen. These constraints were chosen so as to mirror the currently accepted protocols in Pearson and Erwin (1997) within the context of this framework. When these two simple constraints are applied, the number of models to explore in each modeled period is reduced to a much more manageable 274 models.

An exhaustive search of the models in the biologically constrained subspace of B_{10}, allows for a concrete comparison of the relative predictive accuracy of each partitioning. Additionally the partitioned models provide a set of candidate models for use in Bayesian Model Averaging (BMA) (Hoeting et al., 1999). BMA averaging, as applied here, essetially allows the model exploration strategy to

average across all realavent port complex partionings and adds robustness to the end species composition estimates.

For the μ^{th} model in a set of candidate models M, then the BMA weight for M_μ follows directly from Bayes Theorem as follows,

$$\omega_\mu = Pr(M_\mu|y) = \frac{p(y|M_\mu)p(M_\mu)}{\sum_\mu p(y|M_\mu)p(M_\mu)}$$

??*modelpriors*?? Where ω_μ is the posterior probability that model μ is the true data generating model of the data, conditional on the subspace of candidate models and the observed data. ω_μ is then straightforwardly used to average together whichever posterior quantites desired, as follows

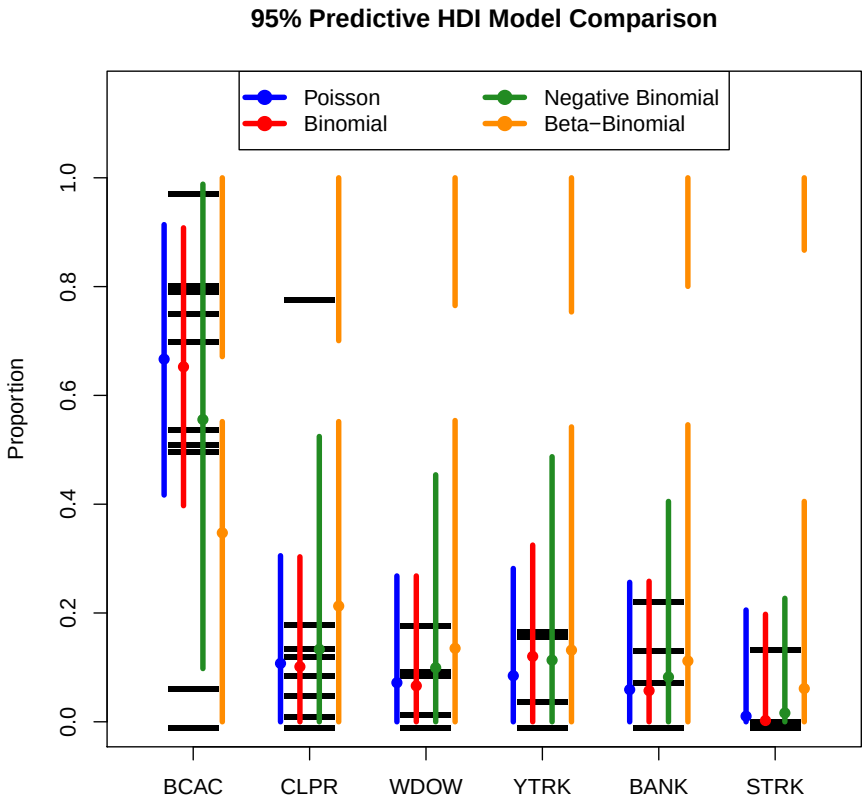
$$\bar{p}(\theta|y) = \sum_\mu \omega_\mu p(\theta|y, M_\mu).$$

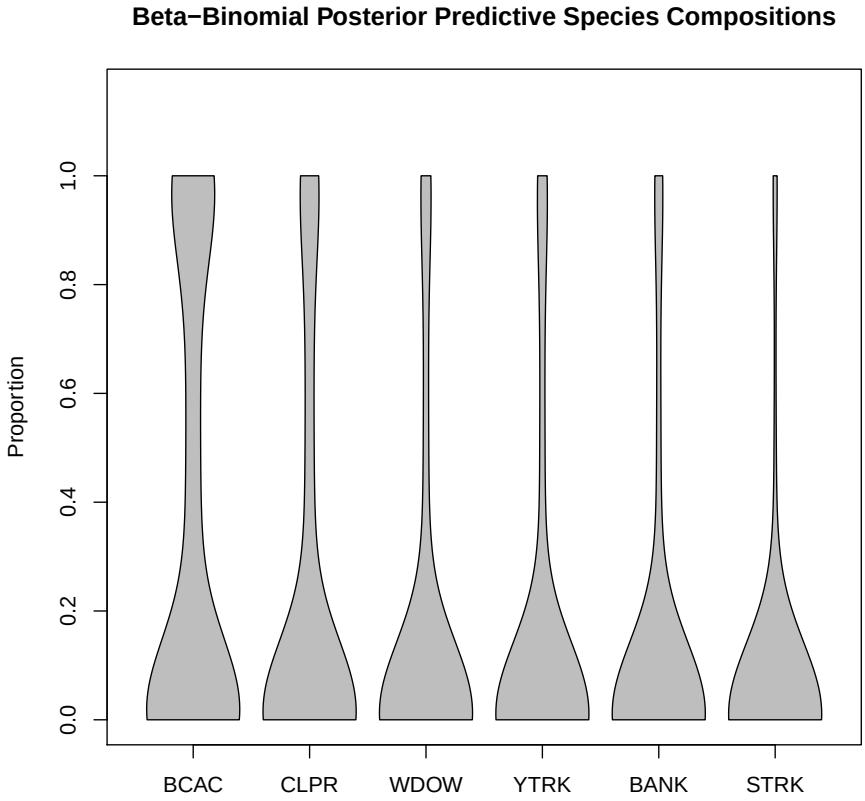
- how to deal with ports
  - model uncertainty around port
  - bell number for exploration
  - constrained exploration
  - bayesian model averaging

Results

Data Generating Model

	Poisson	Binomial	NB	BB
MSE	0.06412	0.06264	0.05171	0.04479
DIC	1342.27	1571.46	345.89	340.86
WAIC	1421.61	1665.41	345.09	341.66
$pr(M\|y)$	≈ 0	≈ 0	$\approx 10^{-7}$	≈ 1


95% Predictive HDI Model Comparison
Poisson
Binomial
Negative Binomial
Beta−Binomial
Proportion
0.0
0.2
0.4
0.6
0.8
1.0
BCAC
CLPR
WDOW
YTRK
BANK
STRK



The large spread of the observed species compositions seen in Figure(1) visually demonstrate the degree of overdispersion present in port sampling data. The Poisson and binomial models disregaurd this overdispersion to prioritize fitting the data mean.

All of the measures in Table(1) consistently agree that the negative binomial and beta-binomial models out perform the overdispersed Poisson and binomial models. Furthermore, all of the metrics in Table(1) indicate that the beta-binomial model outperforms the negative binomial model. Depending on the users value system toward model selection (e.g. predictive or inferential), the support for the beta-binomial model over the negative binomial model may vary, but it is worth noting that the more robust model selction tools show stronger support for the beta-binomial model, with Bayesian model probabilities indicating practically conclusive support for the beta-binomial model.

The split beta-binomial intervals seen in Figure(1) are the consequence of confining a large amount of residual variability to the unit interval. The beta-binomial

is the only model considered here, which estimates such a large degree of variablility and thus it is the only model that produces predictive species composition distributions of the sort. Figure(2) shows the beta-binomial predictive distributions as a violin plot demonstrating how the beta-binomial model arranges predictive density over the unit interval. The predictive intervals in Figure(1) are the smallest possible regions on each density so that the intervals contain 95% of the predictive density (i.e. these regions represent the densest packing of 95% probbaility under each predictive distribution).

Predictor and Prior Selection

	M1	M2	M3	M4	M5	M6
MSE	0.127245	0.127042	0.126801	0.122373	0.127236	0.126573
DIC	39790.26	39491.64	39244.91	37231.70	39407.02	39406.41
WAIC	39745.58	39446.51	39192.25	37182.93	39354.11	39353.49
$pr(M\|y)$	≈ 0	≈ 0	≈ 0	≈ 1	≈ 0	≈ 0

Table(XX) displays the relative support for the model structure on the $\beta_{m\eta}^{(t)}$ time parameters. From M1 to M4 the models represent a spectrum of models with an increasing potential of shrinkage among time parameters. Models M5 and M6 represent models which build in complexity, from M4, via the inclusion of multiple heirarchical variance parameters amoung the interaction terms.

Across all of the time models, model M4 displays consistent support over all other candidate models considered here. Model M4 represents a model with maximal potential for pooling through time, while still maintaing the ability to model differences in seasonality from year to year.

As a final check of the model structure and the implied prior information the prior predictive is considered. The prior predictive distribution summarizes the information that is intrinsic to the model structure itself, in the absense of data. The prior predictive of modeled weight is considered over a 100 pound cluster size, which is consistent with aggregating the two nominal 50 pound cluster samples described by Sen (1984) in the original sampling protocol.

As seen in Figure(XX) the prior predictive of (M4) is both symmetric and quite diffuse over the 100 pound domain. The U shape of the distribution is a side effect of the diffusion of the selected prior. As data are added to the model the indecisive U shape of this distribution collapses toward the data in the posterior.

- Degree of smoothing (heirarchical parameters)

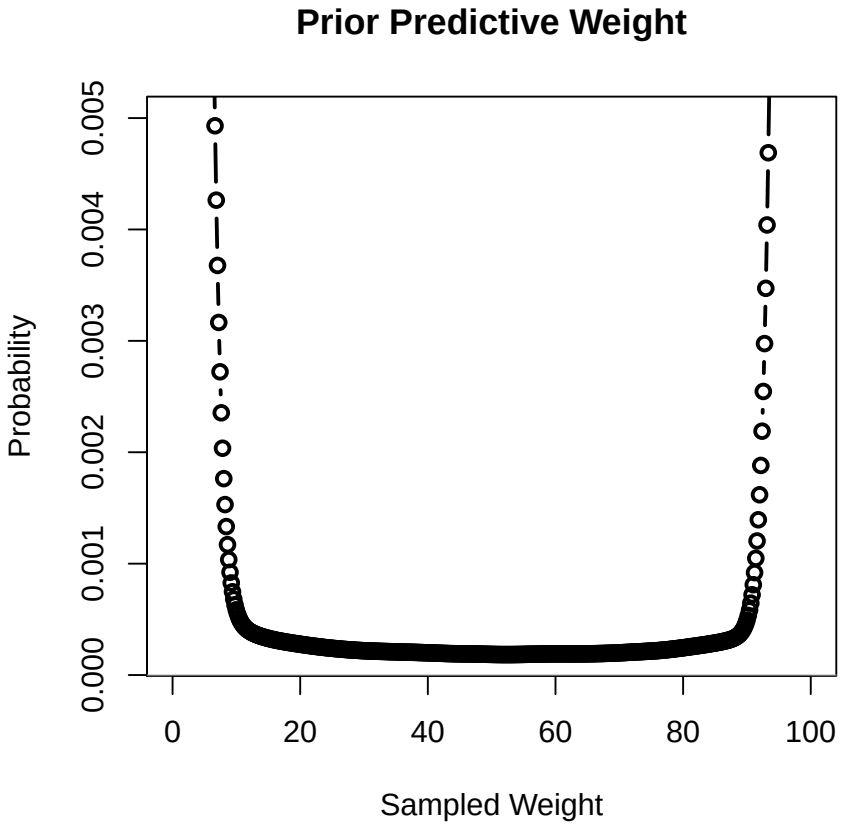


Figure 3: Prior Prediction

Model Exploration & Averaging

- Partitioning results
  - WAIC => Biogeography
  - Pr(M) => model averaging
- Prediction v. Data
  - Report predictive accuracy

78-82

	68%	95%	99%
195	NA	NA	NA
250	67.1%	96.1%	98.7%
253	67.3%	96.3%	98.9%
262	NA	NA	NA
265	69.6%	96.0%	97.8%
269	68.2%	88.8%	90.2%
270	NA	NA	NA
956	68.3%	96.7%	99.2%
959	NA	NA	NA
961	NA	NA	NA

83-90

	68%	95%	99%
245	NA	NA	NA
250	68.1%	96.0%	99.0%
253	69.3%	97.1%	98.9%
259	83.8%	91.9%	92.9%
262	NA	NA	NA
269	68.6%	94.2%	94.7%
270	NA	NA	NA
663	NA	NA	NA
667	69.4%	92.5%	93.5%
956	67.5%	96.2%	99.0%
959	67.4%	96.4%	99.0%
960	68.0%	96.1%	98.6%
961	68.6%	94.6%	97.8%

Compare to Current (*Discussion*)

Discussion

- General Math/Science
- Database Stuff
- Looking Forward
  - forcasting/hindcasting
    - simple
    - timeseries models
  - more computation faster
    - broader model exploration
    - broader spatial expansion

Draft 2

- Orthodox scientific method restructure
  - methods
  - results
  - discussion
- Add MSE bias variance premonition/forshadowing

References

[1] Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014). Bayesian data analysis (Vol. 2). Boca Raton, FL, USA: Chapman & Hall/CRC.

[2] Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: a tutorial. Statistical science, 382-401.

[3] Jaynes, E. T. (2003). Probability theory: The logic of science. Cambridge university press.

[4] Ramasubramanian, K., & Singh, A. (2016). Machine Learning Using R. Apress.

[5] McCullagh P. & Nelder, J.A. (1989). Generalized Linear Models, 2nd ed. London: Chapman and Hall.

[6] Pearson, D.E., and Erwin, B. (1997). Documentation of California's commercial market sampling data entry and expansion programs. NOAA Tech Memo. NOAA-TM-NMFS-SWFSC-240.

[7] R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[8] Rue H., Martino S., Lindgren F., Simpson D., Riebler A. (2013). R-INLA: Approximate Bayesian Inference using Integrated Nested Laplace Approximations. Trondheim, Norway. URL http://www.r-inla.org/.

[9] Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of the royal statistical society: Series b (statistical methodology), 71(2), 319-392.

[10] Sen, A.R. (1984). Sampling commercial rockfish landings in California. NOAA Tech Memo. NOAA-TM-NMFS-SWFSC-45.

[11] Sen AR. (1986). Methodological problems in sampling commercial rockfish landings. Fish Bull. 84: 409-421 .

[12] Shelton, A. O., Dick, E. J., Pearson, D. E., Ralston, S., & Mangel, M. (2012). Estimating species composition and quantifying uncertainty in multispecies fisheries: hierarchical Bayesian models for stratified sampling protocols with missing data. Canadian Journal of Fisheries and Aquatic Sciences, 69(2), 231-246.