

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 accurately modeling not only species composition means, but also higher moments of the data, such as species composition variances, it is necessary to recognize model limitations with respect to over-dispersed data. Among the simplest models for count data are the poisson and binomial models. Both models are typically specified 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 accurately represent higher moments in the data. McCullagh and Nelder (1989, pg. 124) commiserate about the

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 models 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 Monterey 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 predictive species composition distributions from each model 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).

	Poisson	Binomial	NB	BB
MSE	0.05286	0.05683	0.05188	0.05170
DIC	5675.25	6759.86	1301.51	1261.00
WAIC	5840.56	6939.74	1302.19	1261.30
$pr(M y)$	≈ 0	≈ 0	$\approx 10^{-16}$	≈ 1

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 disregard this overdispersion to prioritize fitting the data mean. 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 the negative binomial and beta-binomial models are able to produce more accurate estimates of both the mean and residual variance.

All of the measures in Table(1) consistently agree that the negative binomial



Figure 1: Interval Plot

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 selection 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 variability 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% probability under each predictive distribution). For the cases of Aurora and Bank rockfish, the empty upper regions seen in Figure(1) are understandable in terms of the relatively low density region of the posterior they represent, as seen in Figure(2).

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 . The $y_{ijklm\eta}$ are modeled as observations from a beta-binomial distribution conditional on parameters $\mu_{jklm\eta}$ and $\sigma_{jklm\eta}^2$,

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

Where $\mu_{jklm\eta}$ is the stratum level beta-binomial mean weight and $\sigma_{jklm\eta}^2$ 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, $\sigma_{jklm\eta}^2$ is related to $\mu_{jklm\eta}$ and the overdispersion parameter, ρ , via the following equation,

$$\sigma_{jklm\eta}^2 = \mu_{jklm\eta} \left(1 - \frac{\mu_{jklm\eta}}{n_{ijklm\eta}}\right) \left(1 + (n_{ijklm\eta} - 1) \rho\right).$$

Beta–Binomial Posterior Predictive Species Compositions

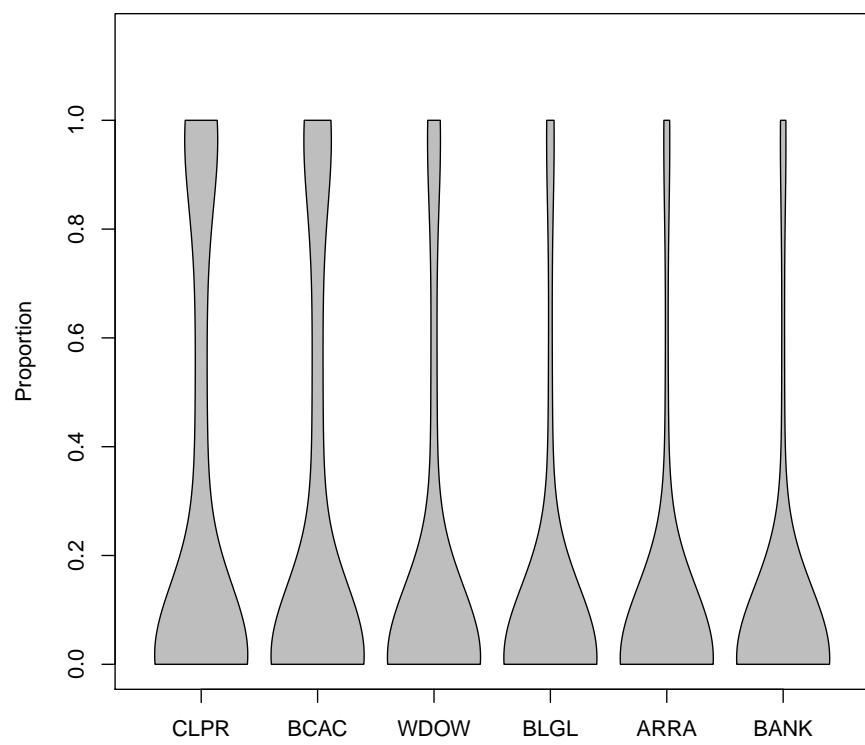


Figure 2: Violin Plot

ρ is the within cluster correlation. The situation where $\rho \rightarrow 1$ represents identical information content among replicates within a cluster, with maximal overdispersion relative to the binomial distribution. The situation where $\rho \rightarrow 0$ represents totally independent information content among 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 flexibility at the stratum level through its linear predictor,

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

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 strategy as described in the following section(XX).

Priors

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

$$\begin{aligned}\beta_0 &\propto 1 \\ \{\beta_j^{(s)}, \beta_k^{(p)}, \beta_l^{(g)}\} &\sim N(0, 32^2)\end{aligned}$$

Since the β_0 reference level is chosen arbitrarily, 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 ($\beta_l^{(g)}$) offsets 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. This principle is the basis for modern model selection criteria (cite/Janyes?). As a simple example consider the structure of the MSE metric for evaluating a predictor, $\hat{\theta}$, with respect to some true parameter θ ,

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

Including additional model parameters is a great way to decrease model bias, however it is apparent from the structure of the MSE that decreasing model bias alone is not a good way of arriving at a well performing model. Models with good MSEs jointly minimize the bias of their parameter estimates, in addition to estimation uncertainty of their parameters.

A model can minimize bias, without regard 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 useless since each estimated parameter is bound to their respective observations, and thus such a model has no information for which to base predictions on future data.

For modeling $\beta_{m\eta}^{(t)}$ we consider a spectrum of models which span a wide range of possible number of parameters and 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)

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

(M2) represents a fixed effects model for year parameters, but estimates a single hierarchical variance parameter, $v^{(q)}$, shared among the $\beta_\eta^{(q)}$. $v^{(q)}$ has the effect of partially pooling information among all quarters. The actual degree of pooling is determined from the data, through the way the data shapes the $v^{(q)}$ posterior.

M3

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

(M3) represents a fixed effects model for quarter parameters, but estimates a single hierarchical variance parameter, $v^{(y)}$, shared among the $\beta_m^{(y)}$. $v^{(y)}$ has the effect of partially pooling information among years but not quarters.

(M4)

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

(M4) estimates two hierarchical variance parameters, $v^{(y)}$ and $v^{(q)}$. $v^{(y)}$ partially pools information among the $\beta_m^{(y)}$, and separately $v^{(q)}$ partially pools information among the $\beta_\eta^{(q)}$.

(M5)

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

(M5) functions similarly as (M4), in that it has hierarchical partial pooling among both the $\beta_m^{(y)}$ and $\beta_\eta^{(q)}$ parameters, except that it introduces 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 $\beta_{m\eta}^{(y:q)}$ is also modeled with a single hierarchical variance parameter, v , shared among all of the $m\eta$ categories. Although the interaction term adds many parameters to the model, the shared v parameter functions to shrink extraneous $\beta_{m\eta}^{(y:q)}$ estimates back toward the common stratum mean.

M6

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

(M6) is largely the same as (M5), 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.

M7

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

(M7) follows the same idea as (M6), however here interaction terms are allowed to partially pool interactions within a common year, across the quarters of that year, but not between years. (M7) involves fitting slightly more parameters than (M6) because in this setting we model more than 4 years of data at once.

Heirarchical variance parameters are estimated from the data. As the above models learn the posteriors of the heirarchical 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 and heavy tailed $v \sim IG(1, 2 \times 10^3)$ 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% interval for a $N(0, 2^2)$ (i.e. 0 ± 3.91), includes almost the entirety of the domain when back transformed to exist in the unit interval (i.e. 0.5 ± 0.48).

	M1	M2	M3	M4	M5	M6	M7
MSE	NA						

	M1	M2	M3	M4	M5	M6	M7
DIC	NA						
WAIC	NA						
$pr(M y)$	NA						

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 which decrease in the possible number of implied parameters. Models M5, M6, and M7 represent models which build in complexity, from M4, via the inclusion of heirarchical interaction terms.

From M1 to M4 all model selection criterion consistently support the inclusion of heirarchical main effects on both year and quarter parmeters. As various model forms for interaction term are added in M5, M6, and M7 the data seem to support prior structures which involve the potential for a higher degree of partial pooling while still supporting the inclusion of year-quarter interactions. All moddel selection criterion agree that M5 shows the highest degree of support. Presumably M5 strikes a balance between maximal potential for heirarchical pooling, 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 is intrinsic to the model structure itself, in the absense of data. The prior predictive of modeled weight is considered over the nominal 100 pound cluster as described in the the sampling protocol (cite).

As seen in Figure(XX) the prior predictive of M5 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.

Species Composition Prediction

Estimating model M5 in a fully Bayesain way gives access to the full posterior distribution of all of the parameters of the model. It is useful to emphasise that in the Bayesain setting these parameters are themselves 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 obtained by computing the desired moments from the posterior samples. Additionally the fact that the parameters are full distributions, means that any

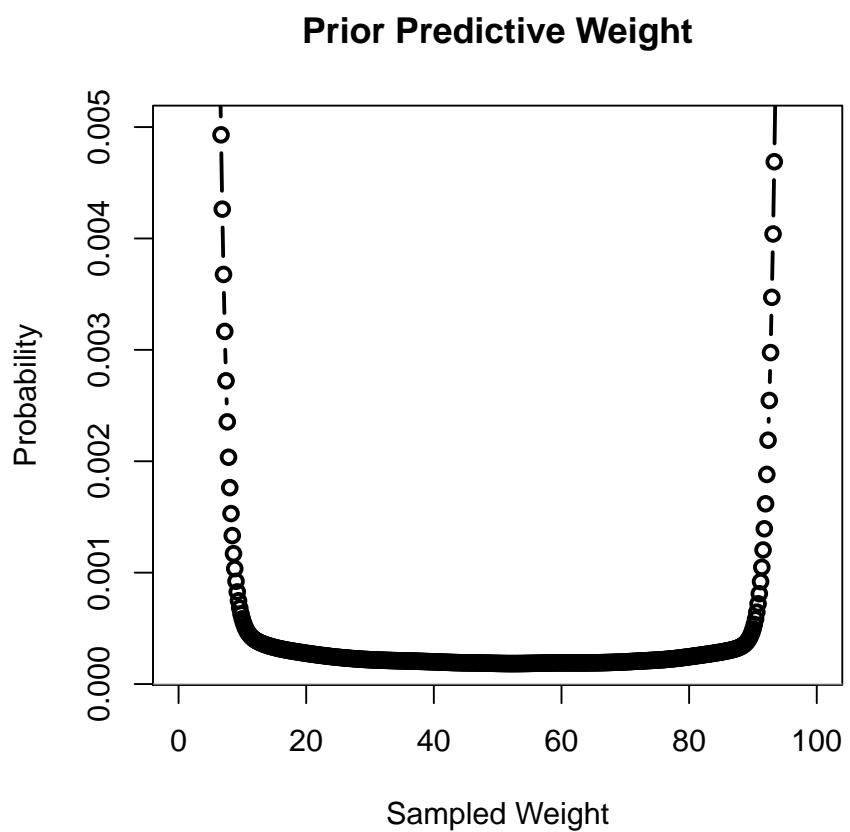


Figure 3: Prior Prediction

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 distribution of sampled weight for a particular stratum.

$$p(y_{jklm\eta}^*|y) = \iint \text{BB}\left(y_{jklm\eta}^*|\mu_{jklm\eta}, \sigma_{jklm\eta}^2\right) P\left(\mu_{jklm\eta}, \sigma_{jklm\eta}^2|y\right) d\mu_{jklm\eta} d\sigma_{jklm\eta}^2.$$

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

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

$$\pi_{jklm\eta}^* = \frac{y_{jklm\eta}^*}{\sum_j y_{jklm\eta}^*} \quad y_{jklm\eta}^* \neq 0.$$

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

Despite efforts to pool data through time via the heirarchical priors described in section (XX), port sampling data still remains sparse.

Presently, stratum with deminishingly small sample sizes are managed by an ad-hoc “borrowing” protocol, outlined in Pearson and Erwin (1997). Their protocol for pooling data across port complexes only calls for spatial pooling to fill holes brought about by unsampled strata. Naturally such a protocol introduces a biase which depends on the availabilty 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.

Given the degree of sparcity in these data it is certainly possible that models pooling data between port complexes may offer predictive benefits, however port complex pooling schemes should minimize any inconsistency brought about by the pooling scheme itself.

To formalize the idea of port pooling we only consider pooling strategies which remain consistent within the modeled period and explicitly outline state-wide

pooling strategies, rather than protocols that change from stratum to stratum. That is to say, for a particular market category and modeling time period, port pooling is defined as some portion of the set of port complexes in California. This formulation does not make it clear which ports should be pooled together in the partitioning of the state, however this formulation does achieve consistency among port pooling within market categories and modeled time periods, while providing a mathematical structure for considering the space of possible pooled models.

The space of possible pooled models is well defined in terms of the size of the set of items to be partitioned, K , as described 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 this setting the set of items to be partitioned is the set of port complexes in California, of which there are $K = 10$, implying a grand total of XXXXX ways of partitioning the port complex in California in each market category and modeled time period. The brute force model selection strategy of computing all XXXXX of these partitioning strategies is computationally infeasible. However, maybe not all pooling schemes represent biologically relevant models. For example, perhaps it is reasonable to pool only among adjacent ports, or to assert that biologically similar regions can only extend across a small number of ports (if so, how many?).

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

Results

- General Products
- Degree of smoothing (hierarchical parameters)
- Posterior v. Current
 - Report degree of similarity
- Prediction v. Data
 - Report predictive accuracy

Discussion

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

Draft 2

- Orthodox scientific method restructure
 - methods
 - results
 - discussion
- Add MSE biase 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] McCullagh P. & Nelder, J.A. (1989). Generalized Linear Models, 2nd ed. London: Chapman and Hall.
- [4] 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.
- [5] 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/>.
- [6] 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/>.

- [7] 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.
- [8] Sen, A.R. (1984). Sampling commercial rockfish landings in California. NOAA Tech Memo. NOAA-TM-NMFS-SWFSC-45.
- [9] Sen AR. (1986). Methodological problems in sampling commercial rockfish landings. *Fish Bull.* 84: 409-421 .
- [10] 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.