# SEDAR-17-AW-08

# A Bayesian approach to stochastic stock reduction analysis, with application to south Atlantic Spanish mackerel

September 11, 2008

Prepared by Paul B. Conn
National Marine Fisheries Service
NOAA Center for Coastal Fisheries & Habitat Research
Beaufort, NC

# 1 Introduction

Researchers have recently identified stochastic stock reduction analysis (SSRA) as an important tool in the suite of stock assessment methods available to fisheries analysts (Walters et al., 2006). This approach works by starting the population off at a range of values for initial biomass and productivity, and asking how probable each of the initial conditions are given the history of exploitation of the stock and how well simulated stock dynamics fit observed data. In this manner, one can get a better handle on plausible values for virgin recruitment (typically denoted $R_0$) and steepness of the stock-recruit curve ($h$; Mace and Doonan 1988), two of the more troublesome parameters to estimate in stock assessments.

Although attractive, SSRA analyses are numerically intensive, and computational ease is facilitated when there are few parameters to fit. For instance, when indices are being fit within a SSRA, one must also estimate catchability parameters, increasing the dimensionality of the problem and adding computing time. Further, results seem to be sensitive to the numerical method chosen for model fitting. For instance, there may be difficulty estimating the scale of population dynamics when fitting to landings and relative abundance indices alone.

Given the difficulties inherent in implementing SSRAs, I wanted to develop and an algorithm that would work well with the data sources available for our focal stock, south Altlantic Spanish mackerel. In particular, landings data were available for a relatively long period of time ($\sim$ 60 years) with indices of relative abundance only available for the last 23 years of the time series. In this working paper, I start by introducing an approach for fitting SSRA's using a bootstrap particle filter (Gordon et al., 1993), as articulated by Robert and Casella (2004, Algorithm A.59). This procedure is an adaptation of sequential importance sampling (e.g., McAllister et al., 1994), where filtering occurs after every iteration (i.e., year). I also introduce a procedure for gradient fitting that removes the need to estimate catchability parameters, and detail an approach for scaling the problem by including prior distributions for mortality rates. This approach is then applied to simulated data in order to ensure that the model produces coherent results. Finally, I apply this approach to Spanish mackerel in the south Atlantic (U.S.), providing estimates of productivity parameters and posterior probabilities of relative benchmarks.

# 2 Stochastic stock reduction analysis

## 2.1 Model structure

I describe a SSRA where uncertainty in population dynamics is described by the parameter vector

$$\boldsymbol{\theta} = \{R_0, h, \sigma_R, \epsilon_1, \epsilon_2, \ldots, \epsilon_Y\},$$

where $R_0$ gives average recruitment for an unexploited population, $h$ gives steepness, $\sigma_R$ gives the standard deviation of recruitment deviations around the stock recruit function, and $\epsilon_t$ gives the annual recruitment deviation in year $t$. The inclusion of uncertainty in the $\epsilon_t$ parameters is the fundamental difference between deterministic and stochastic stock reduction analysis;

inclusion of uncertainty in the latter is essential for adequately assessing population viability of a stock over the history of exploitation (McAllister et al., 1994).

In addition to the parameters of the SSRA, I assume that there are a number of inputs to the model, each of which is assumed to be known without error. These are given by

$$\boldsymbol{\phi} = \{\mathbf{M}, \mathbf{s}, \mathbf{l}, \mathbf{w}, \mathbf{m}\},$$

which represent age-specific vectors of natural mortality, selectivity of the gear(s) used in the fishery, length, weight, and sexual maturity, respectively. Several of these quantities (e.g., $\mathbf{l}$, $\mathbf{w}$, and $\mathbf{m}$), are often easy to obtain from published studies on life history. Quantities such as $\mathbf{M}$ and $\mathbf{s}$ are a bit more problematic; estimates of $\mathbf{s}$ may be obtained from other assessment methods (e.g., statistical catch-at-age models), while tagging studies or an appeal to the theory of life history invariants (e.g., Charnov, 1993) may be needed to come up with a plausible range of values for natural mortality.

Conditional on the values of $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$, I specify a deterministic population model for stock dynamics as follows:

- For simplicity, the population is started at equilibrium. This is accomplished by letting the age-specific abundance vector in year 1 be $\mathbf{N_1} = R_0 \times \Big[1, \exp(-M_1), \exp(-(M_1 + M_2)), \ldots, \exp(\sum_{a=1}^{A-1} -M_a), p_A\Big]$, where $p_A = \exp(\sum_{a=1}^{A} -M_a)/(1 - \exp(-M_A))$.
- Assuming constant, continuous hazard rates for natural and fishing mortality over the course of the year, I solve for the fishing mortality

rate $F_t$ such that

$$\sum_a N_{ta} \frac{\exp(-F_t s_a)}{\exp(-(F_t s_a) * M_a)} (1 - \exp(-(F_t s_a))) w_a = L_t,$$

where $L_t$ gives landings in weight in year $t$. The number of individuals in a given age class after the year is complete is then given by

$$N_{t+1,a} = N_{ta} \exp(-(F_t s_a + M_a)).$$

- The number of individuals alive at the middle of year $t$ in age class $a$ is calculated as $N_{ta}^{mid} = N_{t,a} \exp(-(F_t s_a + M_a)/2)$. This summary is used to predict expected relative abundance in year $t$.

- The number of new recruits in year $t+1$ is then generated as

  $$N_{t+1,1} = \epsilon_t \frac{0.8 R_0 h \mathrm{SSB}_{t+1}}{0.2 \Phi_0 R_0 (1-h) + (h - .2) \mathrm{SSB}_{t+1}}$$

  where $\mathrm{SSB_t}$ gives spawning stock biomass at the beginning of year $t$:

  $$\mathrm{SSB_t} = \sum_a N_{ta} m_a w_a.$$

## 2.2 Fitting the model to data

Conditional on an abundance vector, landings provide the information necessary to calculate annual fishing mortality rates. However, landings alone only provide a minimum bound on abundance; additional information is needed for the population model to be scaled correctly. One potential source of information that can help with this scaling is an index of relative abundance. Under the assumption of constant catchability, it is often assumed that total abundance is related to an index of abundance through an equation such as

$$I_t = q \sum_a N_{t,a} s_a w_a,$$

where $q$ denotes a catchability parameter. To fit the population model to such an index one would typically need to conduct simultaneous inference on $q$ (e.g., McAllister et al., 1994).

An alternative approach, which also assumes constant catchability, is to conduct inference on $\lambda_t = I_{t+1}/I_t$, which is related to the finite rate of population increase. This quantity is dimensionless, removing the need to estimate $q$. This approach, sometimes termed gradient matching, has recently been used to fit models to time series of dynamical systems (e.g., Ellner et al., 2002). One possibility for for fitting the population model to time series estimates of $\lambda_t$ is to impose a probability model such as $\lambda_t^{\text{pred}} \sim \text{Normal}(\lambda_t^{\text{obs}}, \sigma_\lambda)$. Here, one could use a hierarchical Bayesian analysis or simple calculation using the delta method (Seber, 1982) to get an approximate estimate of $\sigma_\lambda$ from index time series and accompanying measures of precision.

Owing to the large number of latent recruitment deviations in SSRAs, classical maximum likelihood inference is problematic. An alternative in this case is to use Bayesian inference. Posterior inference may be based on Markov chain Monte Carlo (cf., Gelman et al., 2004), or alternatively on a population Monte Carlo method such as sequential importance sampling (SIS). The latter approach has been used in a number of fisheries applications (e.g. McAllister et al., 1994; McAllister and Ianelli, 1997; Newman and Lindley, 2006), and involves sampling the initial state vector $\boldsymbol{\theta}$ a large number of times (say $n_p$), based on assumed prior distributions for parameters. Each sample, termed a "particle," is passed through the population model. The probability of retaining a particle then depends on the fit to data; those particles surviving this process then contribute to inference about the initial parameters.

There are many possible algorithms for performing sequential importance sampling. However, simpler algorithms (e.g., McAllister et al., 1994) can lead to problems with "threading the needle" (Walters et al., 2006) in the context of SSRAs, whereby an extremely large sample of particles would be needed to reasonably emulate observed changes in the population. In contrast, relatively sophisticated algorithms may be needed when trying to fit models with a large number of correlated parameters (e.g., Buckland et al., 2004; Newman et al., 2006; Newman and Lindley, 2006). I settled on an algorithm with moderate complexity known as the "bootstrap particle filter" (Gordon et al., 1993), as described by Robert and Casella (2004, Algorithm A.59). This algorithm, adapted for purposes of SSRA's proceeds as follows:

- Step 1: Randomly sample $n_p$ values from prior distributions for $R_0$, $h$, and $\sigma_R$. The $i$th draw from each distribution are associated with particle $i$.

- Step 2: Initialize year, $t = 1$, and for each particle, generate initial population vector

- Step 3: For each particle, generate a recruitment deviation $\epsilon_t \sim \text{lognormal}(0, \sigma_R)$. Propagate the population forward one time step.

- Step 4: Assign a weight $w_p$ to each particle, which equals 0 if landings exceed abundance, and equals $\mathrm{L}(\lambda_t|\boldsymbol{\theta})$ otherwise; $\mathrm{L}(\lambda_t|\boldsymbol{\theta})$ gives the likelihood (e.g., normal) for the observed value of population change.

- Step 5: Resample the particles with replacement, where the probability of selecting particle $p$ is given by $w_p / \sum_p w_p$. Increment year, $t = t + 1$

- Step 6: Repeat steps 3 to 5 until the end of the study.

The collection of parameters in the final sample is then an approximation to the posterior distribution of model parameters.

## 2.3 Scaling using independent estimates of mortality

Initial results using this approach indicated substantial positive bias in estimates of $R_0$ when there were a large number of years of landings preceding the first observation of $\lambda$. Briefly, the problem appeared to be that particles with lower values of $R_0$ would often be removed from analysis because the only criterion for filtering at the beginning of the time series was whether or not landings exceeded abundance. To survive a period of high landings, a low $R_0$ population would need to have a sequence of high recruitment years, which was increasingly unlikely the longer the period preceding $\lambda$ estimates.

As a possible way to achieve the right scaling in estimates of $R_0$, we considered including auxiliary estimates of $Z_t$, the cumulative instantaneous mortality rate, and their standard errors, $\sigma_t^z$ in the estimation procedure. Such estimates are often available from catch curve analyses when the age structure of harvests are also sampled. The basic idea was to impose extra structure on the $F_t^{pred}$ that were needed to produce landings in year $t$. This was done by specifying a probability model for $Z_t^{pred} = F_t^{pred} + M^*$. Specifically, we assumed that $Z_t^{pred} \sim \text{Normal}(Z_t^{obs}, \sigma_t^z)$, so that the likelihood in Step 4 of the algorithm in section 2.2 was replaced by $\text{L}(\lambda_t|\boldsymbol{\theta}) \times L(Z_t^{pred}|Z_t^{obs}, \sigma_t^z)$. Catch curve analysis often uses individuals greater or equal to a threshold age for analysis (so that equal selectivity may be assumed); in this case, $M^*$ gives the values of natural mortality most appropriate for this age grouping.

## 3 Simulation testing

For simulation testing, we assumed that there were a total of $Y = 40$ years of data; landings in weight were available each year, but an index of abundance was only available for the last 20 years of the study. In each case, we set $R_0 = 30,000,000$, $h = 0.4$, $\sigma_R = 0.5$, and $F = -log(0.8)$ to generate data. Recruitment deviations were simulated stochastically, but all other calculations were deterministic. Values for selectivity, growth, maturity, and weight at age were chosen to be similar to Spanish mackerel. Since Spanish mackerel are sexually dimorphic, males and females were modeled differently, with sex-specific vectors in each case. For simplicity a sex ratio of 0.5 was assumed at the time of recruitment. With the plus group chosen to occur at age $A = 11$, age-specific vectors were set to

$$\mathbf{s_M} = [0.1, 0.4, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0],$$

$$\mathbf{s_F} = [0.1, 0.7, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0],$$

$$\mathbf{m_F} = [0.0, 0.9, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0],$$

$$\mathbf{l_M} = [310, 394, 454, 497, 528, 550, 566, 578, 586, 592, 596],$$

$$\mathbf{l_F} = [310, 394, 454, 497, 528, 550, 566, 578, 586, 592, 596],$$

$$\mathbf{w_M} = [0.60, 1.19, 1.80, 2.34, 2.79, 3.15, 3.42, 3.63, 3.78, 3.89, 3.98], \text{and}$$

$$\mathbf{w_F} = [0.60, 1.19, 1.80, 2.34, 2.79, 3.15, 3.42, 3.63, 3.78, 3.89, 3.98].$$

Here, $M$ denotes males and $F$ denotes females. Males did not contribute to spawning stock biomass.

For estimation, we specified the following priors for model parameters:

$$[R_0] : \text{Uniform}(10000000, 60000000)$$

$$[h] : \text{Uniform}(0.25, 0.95)$$

$$[\sigma_R] : \text{Uniform}(0.35, 0.80).$$

Observed estimates of $\lambda_t$ were set to their true values, and a 0.05 standard error was assumed for model fitting. The observed value for $Z_t^{obs}$ was set to it's true value, and a 0.05 standard error was assumed. A total of 200,000 particles was used in the estimation procedure.

Using this approach, estimated posterior samples of population trajectories were able to track changes in $\lambda_t$ over time (Figure 1). Plots of posterior distributions of model parameters were centered around the values used to simulate data (Figure 2), and posterior plots for stock status in relationship to management benchmarks were centered on true stock status (Figure 3). Posterior summaries for stock status included uncertainty in both the true value of maximum sustainable yield, as well as spawning stock biomass and fishing mortality rate at the end of the time period. Thus, computing the posterior probabilities of overfishing or of the stock being overfished are relatively straightforward calculations.

# 4 Spanish mackerel analysis

I obtained Spanish mackerel landings data in weight from (SEDAR, 2008a) for the period 1950-2007 off the Atlantic seaboard. These landings were adjusted somewhat to account for unreported shrimp bycatch and intermittent recreational landings at the beginning of the time series (SEDAR, 2008b, Chapter 2). Although some commercial landings records existed for years preceding 1950, they had a number of breaks; further, no recreational or bycatch estimates were available for these years. We thus decided to employ an initialization period of 50 years, whereby landings were linearly interpolated back to zero in 1900 (Figure 4).

Estimates of $\lambda_t$ were obtained from a Bayesian hierarchical analysis of seven different gear types, and ran from 1982-2007 (SEDAR, 2008d). We assumed a normal distribution for $\lambda_t$, using empirical estimates of standard error from SEDAR (2008d) to characterize uncertainty (Figure 5). However, standard errors of $\lambda$ estimated below 0.05 were replaced with 0.05 in an effort to prevent particle depletion. Similarly, we used estimates of total mortality form catch-curve analysis (SEDAR, 2008c) to characterize prior distributions for fishing mortality rate, $F_t$. We chose to base priors on $F_t$ rather than $Z_t$ because estimates of natural mortality varied by age (SEDAR, 2008a). Estimates of $Z_t$ based on age 2 and older cohorts almost always fell between 0.5 and 1.4 for the years that age composition samples were available (1981-2007; Figure 6). However, subtracting out M=0.35 (the Hoenig estimate of natural mortality) led to a range for $F_t$ that was approximately 0.15 to 1.05 for the period 1982-2003. We thus specified a Uniform(0.15, 1.05) prior on $F_t$ for the period 1982 to 2007. For the period 1950-1980, we knew less about $F_t$ because no age samples were available; for this period we admitted more uncertainty by assuming that $F_t \sim Uniform(0.05, 2.0)$ was a reasonable prior distribution. For the initialization period (1901-1949), we imposed an even less informative prior of Uniform(0.00,2.0). However, we set recruitment deviations for the initialization period to zero to prevent lower values of $R_0$ from being removed prematurely from the population of particles.

Model inputs ($\boldsymbol{\phi}$) were the same as for simulated data, with the following exceptions:

$$\mathbf{l_M} = [291, 377, 431, 465, 486, 499, 507, 512, 515, 517, 518],$$

$$\mathbf{l_F} = [306, 404, 472, 519, 552, 575, 591, 602, 610, 616, 620],$$

$$\mathbf{w_M} = [0.49, 1.05, 1.550, 1.93, 2.19, 2.37, 2.48, 2.56, 2.60, 2.63, 2.65], \text{and}$$

$$\mathbf{w_F} = [0.57, 1.28, 2.01, 2.66, 3.19, 3.59, 3.89, 4.11, 4.26, 4.38, 4.46].$$

$$\mathbf{M} = [0.50, 0.41, 0.36, 0.33, 0.31, 0.30, 0.29, 0.29, 0.29, 0.28, 0.28]$$

These values were recommended by the SEDAR 17 data workshop (SEDAR, 2008a). Similarly, prior distributions for $R_0$, $h$, and $\sigma_R$ were the same as for simulations, reflecting a considerable degree of uncertainty about stock dynamics.

A SSRA was fit to these data using the bootstrap particle filter (see section 2.2) with 50,000 particles. Posterior estimates of stock-recruit parameters indicated that $R_0$ was likely quite high (on the order of 60 million fish), that steepness was close to 0.6, and that recruitment variability is pronounced (Figure 7)

I also calculated posterior probabilities of population status in relation to management benchmarks. This was accomplished by estimating $F_{msy}$ and $SSB_{msy}$ for each particle after the final filtering round, and by keeping track of the terminal fishing mortality and spawning stock biomass for these particles. Posterior distributions for $F_T/F_{msy}$ and $SSB_T/SSB_{msy}$ thus incorporated uncertainty in parameter estimates (Figure 8). Results from the SSRA suggest that the stock is overfished and that overfishing is occuring. Relevant probabilities can be obtained by (1) integrating the posterior for $F_T/F_{msy}$ with limits taken from one to infinity, and (2) integrating the posterior for $SSB_T/SSB_{msy}$ with limits taken from zero to one.

We urge caution in interpreting these results. It was evident that particle depletion was a problem in this analysis, in that only a few combinations of initial values survived the filtering process. Also, this analysis does not incorporate data on the age structure of the population; thus statistical catch-at-age analysis should be the preferred analysis. My(limited) experience with fitting the SSRA to long time series indicates a tendency for populations with low $R_0$ to go extinct prematurely. This may be the case with the present analysis.

# References

Buckland, S. T., K. B. Newman, L. Thomas, and N. Koester. 2004. State-space models for dynamics of wild animal populations. Ecological Modelling **171**:157–175.

Charnov, E. L. 1993. Life history invariants: some explorations of symmetry in evolutionary ecology. Oxford University Press, Exford.

Ellner, S. P., Y. Seifu, and R. H. Smith. 2002. Fitting population dynamic models to time-series data by gradient matching. Ecology **83**:2256–2270.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 2004. Bayesian Data Analysis, 2nd Edition. Chapman and Hall, Boca Ration.

Gordon, N., J. Salmond, and A. Smith. 1993. A novel approach to non-linear/non-Gaussian Bayesian state estimation. IEEE Proceedings on Radar and Signal Processing **140**:107–113.

Mace, P. M., and I. J. Doonan, 1988. A generalized bioeconomic simulation model for fish population dynamics. Technical Report 88/4, N.Z. Fisheries Assessment Resource Document.

McAllister, M. K., and J. N. Ianelli. 1997. Bayesian stock assessment using catch-age data and the sampling-resampling importance algorithm. Canadian Journal of Fisheries and Aquatic Sciences **54**:284–300.

McAllister, M. K., E. K. Pikitch, A. I. Punt, and R. Hilborn. 1994. A Bayesian approach to stock assessment and harvest decisions using the sampling/importance resampling algorithm. Canadian Journal of Aquatic Sciences **51**:2673–2687.

Newman, K. B., S. T. Buckland, S. T. Lindley, L. Thomas, and C. Fernàndez. 2006. Hidden process models for animal population dynamics. Ecological Applications **16**:74–86.

Newman, K. B., and S. T. Lindley. 2006. Accounting for demographic and environmental stochasticity, observation error, and parameter uncertainty in fish population dynamics models. North American Journal of Fisheries Management **26**:685–701.

Robert, C. P., and G. Casella. 2004. Monte Carlo Statistical Methods, 2nd Edition. Springer, New York.

Seber, G. A. F. 1982. The Estimation of Animal Abundance and Related Parameters, 2nd Edition. Blackburn Press, Caldwell, NJ, USA.

SEDAR, 2008a. SEDAR 17 Data Workshop Report.

SEDAR, 2008b. SEDAR 17 Assessment Workshop Report.

SEDAR, 2008c. SEDAR-17-AW-04: Catch curve analysis of age composition data for Spanish mackerel.

SEDAR, 2008d. SEDAR-17-AW-06: Methods for combining multiple indices into one, with application to south Atlantic (U.S.) Spanish mackerel.

Walters, C. J., S. J. D. Martell, and J. Korman. 2006. A stochastic approach to stock reduction analysis. Canadian Journal of Fisheries and Aquatic Sciences **63**:212–223.

Figure 1: A plot of the predicted annual changes in relative abundance for 30 randomly selected particles from the posterior distribution (solid lines) plotted against true (simulated) $\lambda_t$ values. Filtered particles do a reasonable job at tracking observed population changes.

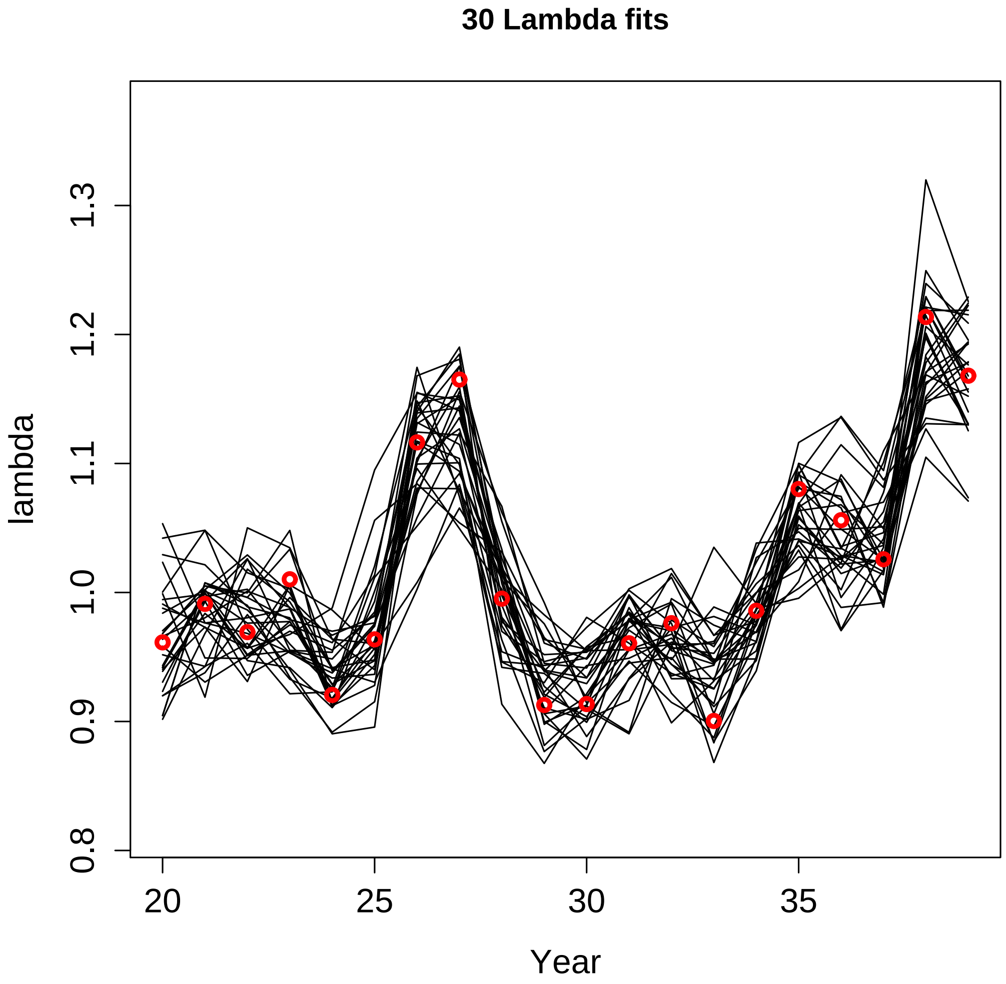


Figure 2: Histograms summarizing the posterior probability density function for key stock-recruit parameters ($R_0$, $h$, and $\sigma_R$) from the analysis of simulated data. Posterior plots were centered at true values (dashed lines).

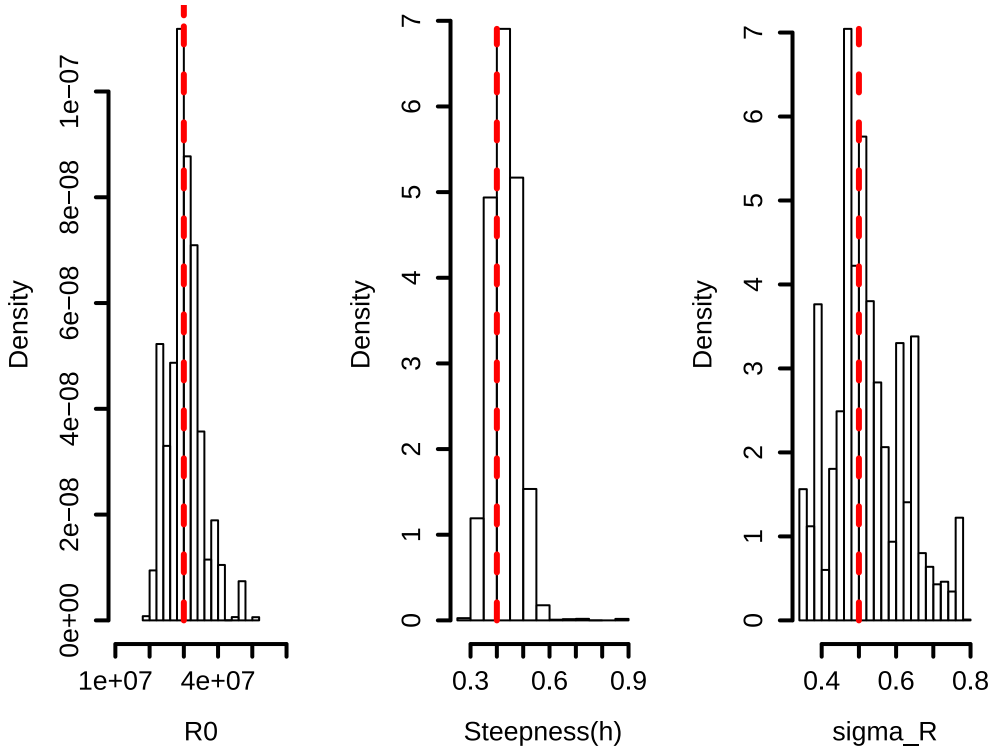
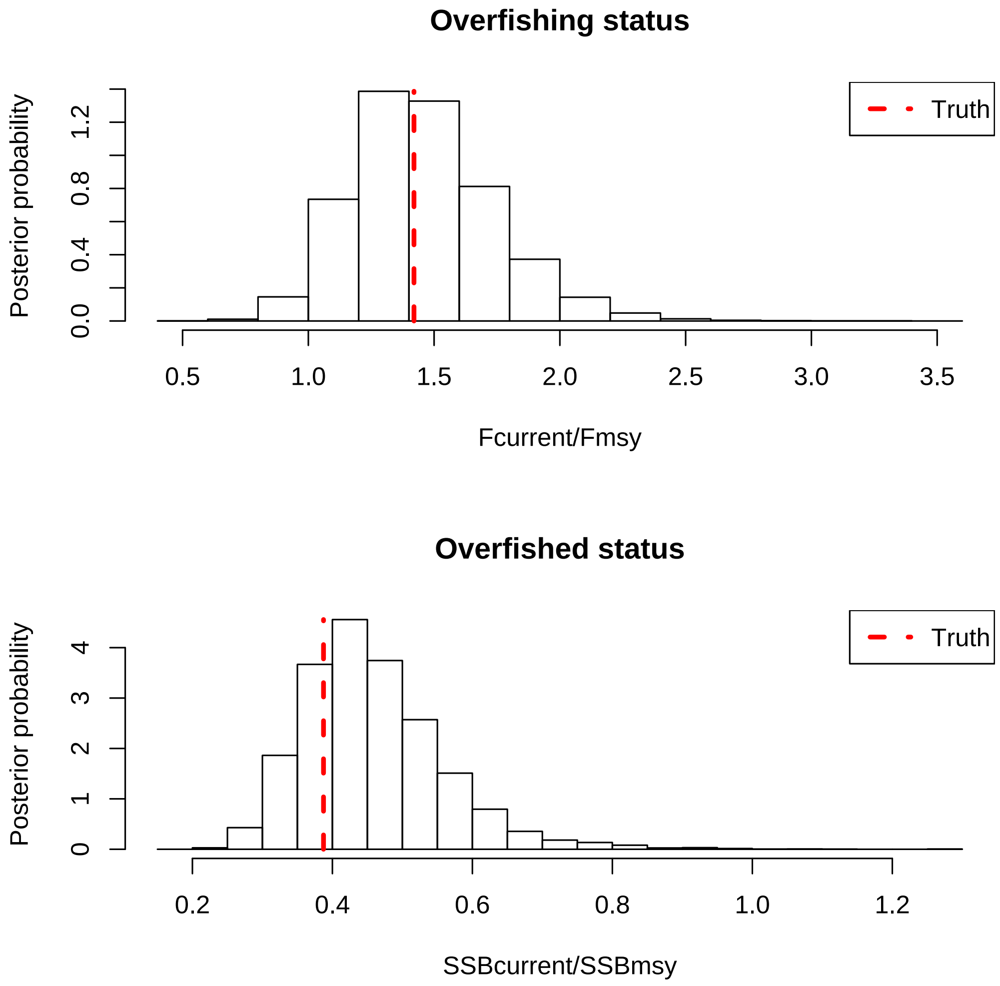


Figure 3: Posterior distributions for stock status in relation to relative management benchmarks from analysis of simulated data. True stock status (dashed lines) is located close to posterior modes.



Overfishing status

Overfished status

Figure 4: Modeled landings for the stochastic stock reduction analysis of U.S. Atlantic Spanish mackerel. The period 1901-1949 was treated as an initialization period where landings were linearly interpolated back to zero. Recruitment was constrained to follow a Beverton-Holt model exactly during the initialization period (i.e., no recruitment deviations).

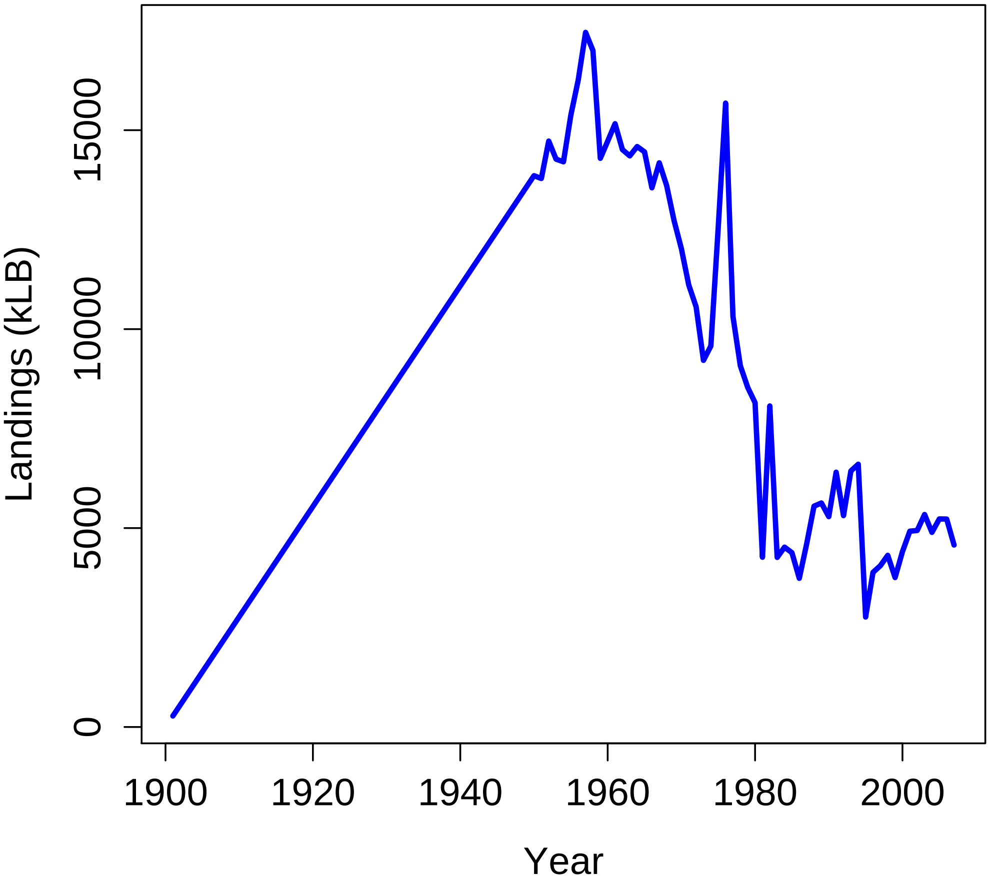


Figure 5: Fitted trajectory of 30 randomly selected particles to $\lambda_t$. Error bars give the estimates of $\lambda_t \pm 2SE$ that were used in the fitting process

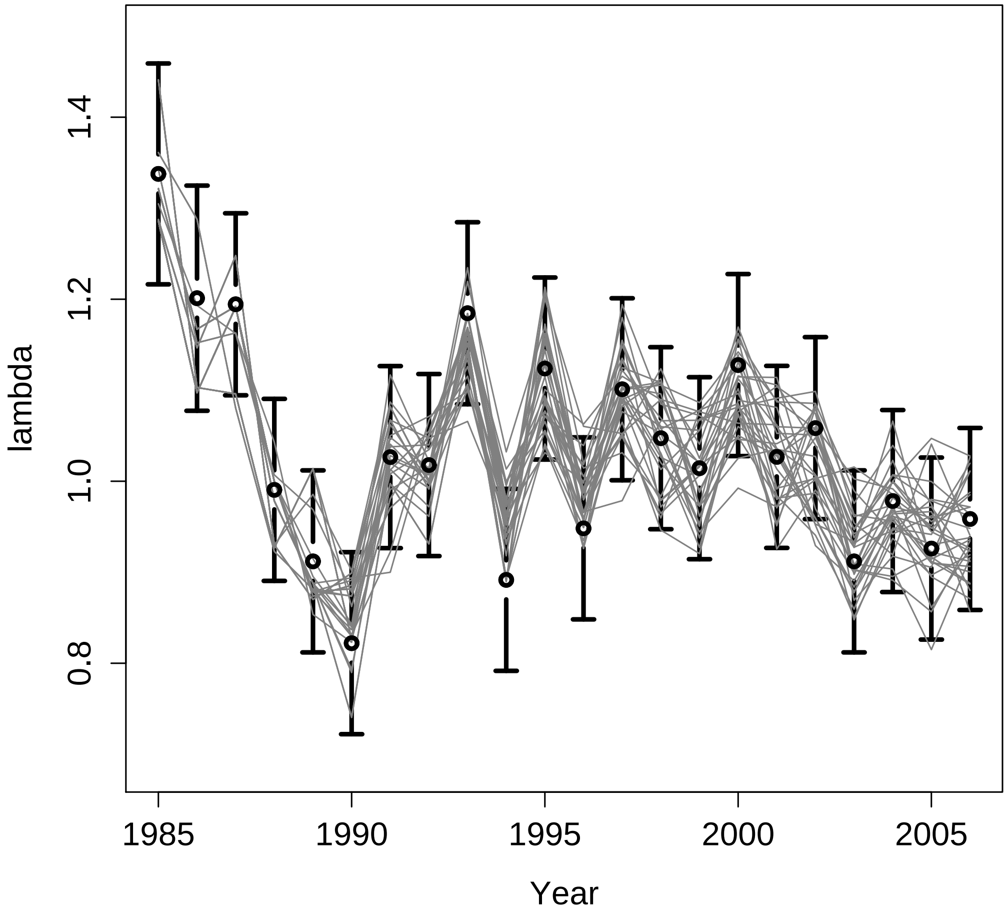


Figure 6: Catch-curve estimates of $Z_t$ used to derive a prior distributions for $F_t$. Included are results from four fisheries: Commercial gillnet ("Comm gillnet"); recreational ("MRFSS"); commercial castnet ("Comm castnet"); and commercial handlines ("Comm HL").

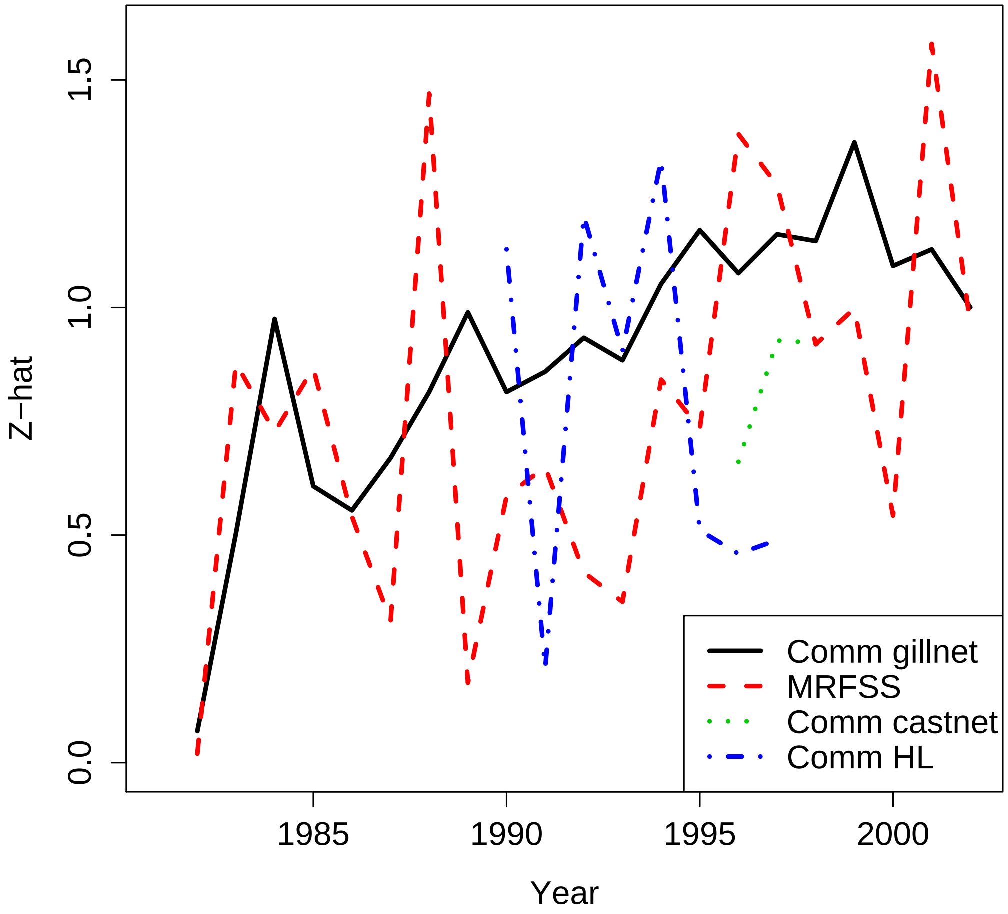


Figure 7: Histograms representing posterior probabilities for model parameters. A larger number of particles would be needed to generate a smooth distribution

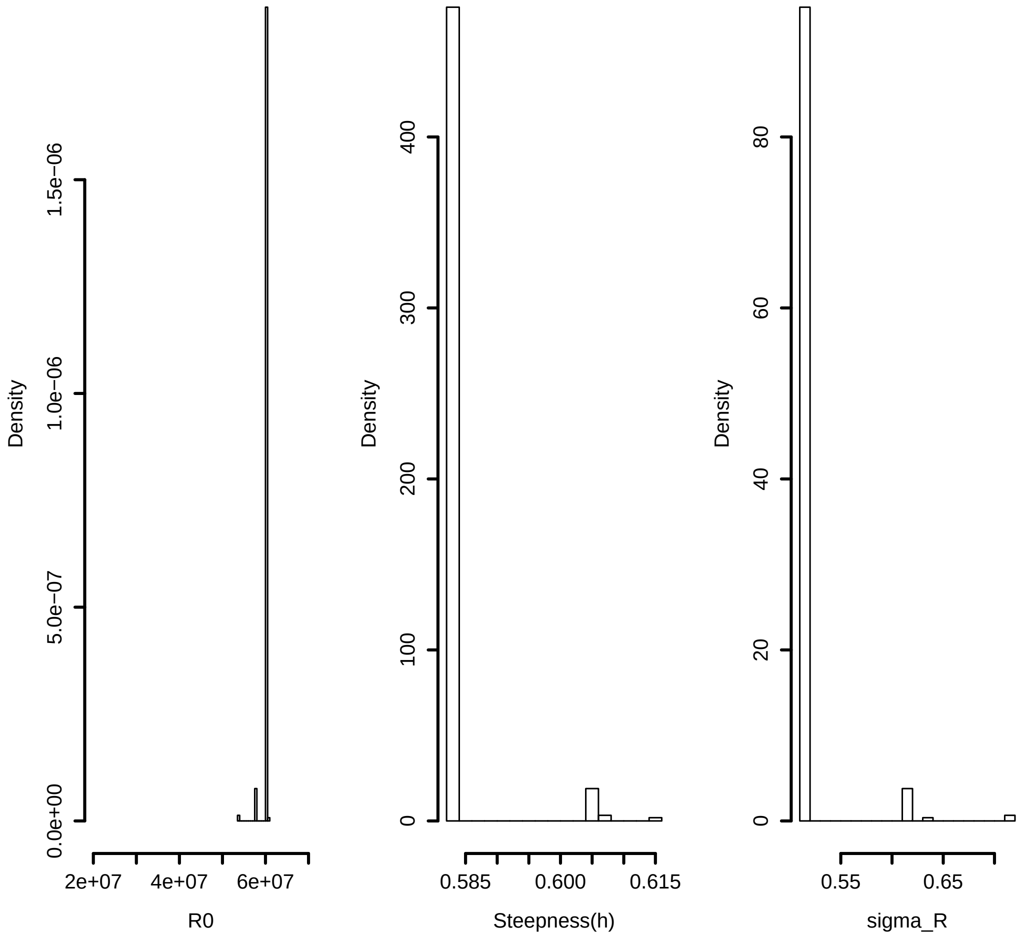
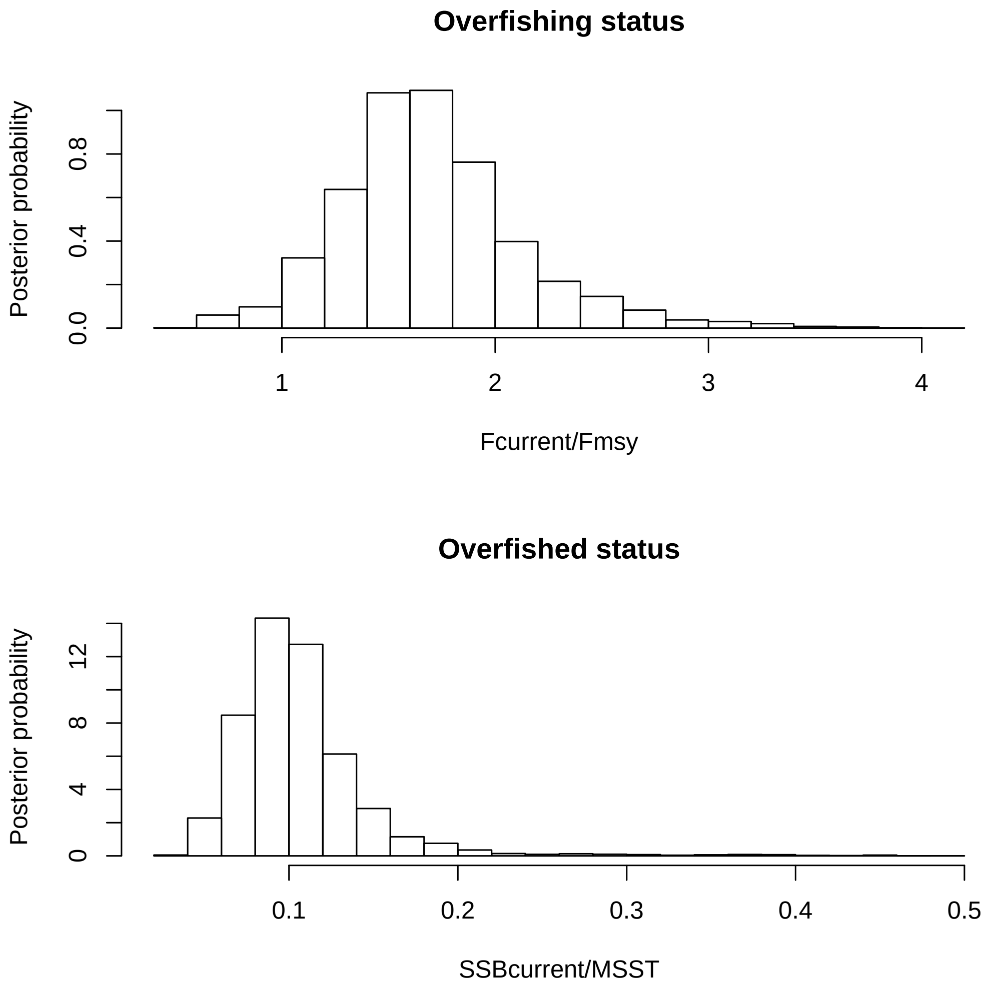


Figure 8: Histograms representing posterior predictions for stock status in relation to management benchmarks.



Overfishing status

Overfished status