State-Space Stock Assessment Models

Tim Miller

Northeast Fisheries Science Center

Hidden process models

- Separate stochastic nature of hidden variables of interest from that of observations.
- Generalizes time series models and models for independent observations.
- Includes all types of mixed effects models, time series, and state-space models as special cases.
- Newman et al. (2006) is a good introduction with a focus on population models.

Hidden process models

- The hiden process $\theta_1, \ldots, \theta_T$ evolves stochastically over time.
- The probability model for the hidden process θ,

$$P\left(\theta_t\right) = f\left(\theta_1, \ldots, \theta_{t-1}\right)$$

  is equivalent to a statistical time series model (e.g., ARIMA), that may or may not be linear and Gaussian.
- The probability model for observations

$$P\left(x_t\right) = f\left(\theta_t\right)$$

  generally treat the observations at time t, conditional on the state of the hidden process, as independent.

State-space models

- For state-space models, probability of current state can be written as a function of the state at the previous time step

$$P(\theta_t) = f(\theta_{t-1})$$

- The Kalman Filter (Kalman 1960) to estimate linear Guassian state-space models was a major breakthrough.
- A number of previous applications of state-space models in fisheries spanning decades.
    - Biomass Dynamics/Production models (Millar and Meyer 2000)
    - Delay-difference models (Meyer and Millar 1999)
    - Length-based models (Sullivan et al. 1990)
    - Age-structured models (Mendelssohn 1988)
- But applications in management are relatively recent (Nielsen and Berg 2014)

Traditional Population models

We model population at time t as a function of population at time $t-1$.

- For a (difference equation) Shaefer model

$$B_t = B_{t-1} r \left(1 - \frac{B_{t-1}}{K}\right) - C_{t-1}$$

- For an age-structured model

$$N_{t,a} = N_{t-1,a-1} e^{-Z_{t,a}}$$

where $Z_{t,a} = F_{t,a} + M_{t,a}$

Traditional Population models

All stochasticity is in the observation equations

$$P(\log I_t) = N\left(\log \widehat{I}_t, \sigma_t^2\right)$$

- For a Shaefer model

$$\widehat{I}_t = qB_t$$

- For an age-structured model

$$\widehat{I}_{t,a} = q_a N_{t,a} e^{-Z_{t,a} f}, \quad \widehat{I}_t = \sum_{a=1}^{A} \widehat{I}_{t,a}$$

There are other probability models for age compostion observations and catch for statistical catch at age models.

State-space versions of population models

Include stochasticity in the evolution of the population over time so population size is the hidden state variable θ

- For a Shaefer model, the expected biomass is the same

$$\widehat{B}_t = B_{t-1} r \left(1 - \frac{B_{t-1}}{K}\right) - C_{t-1}$$

  The realized biomass is a stochastic outcome. For example,

$$P\left(\theta_t | \theta_{t-1}\right) = P\left(B_t | B_{t-1}\right) = \mathrm{N}\left(\log \widehat{B}_t, \sigma_B^2\right)$$

- For an age-structured model, the expected abundance at age is the same

$$\widehat{N}_{t,a} = N_{t-1,a-1} e^{-Z_{t-1,a-1}}$$

  and $\theta_t = N_{t,1}, \ldots, N_{t,A}$ is a vector. For example,

$$P\left(\theta_{t,a} | \theta_{t-1,a-1}\right) = P\left(N_{t,a} | N_{t-1,a-1}\right) = \mathrm{N}\left(\log \widehat{N}_{t,a}, \sigma_N^2\right)$$

State-space probability models

Let D_t be the observations at time t, and modeling process from time $t = 1$ to $t = T$

- For a Shaefer model, the joint probability of the process and observations is

$$P(\boldsymbol{\theta}, \boldsymbol{D}) = P(B_1, \ldots, B_T, D_1, \ldots, D_T) = \prod_{t=1}^{T} P(D_t | B_t) P(B_t | B_{t-1})$$

- For an age-structured model

$$\begin{aligned} P(\boldsymbol{\theta}, \boldsymbol{D}) &= P\left(N_{1,a}, \ldots, N_{T,A}, D_1, \ldots, D_T\right) \\ &= \prod_{t=1}^{T} P\left(D_t | N_{t,1}, \ldots, N_{t,A}\right) \prod_{a=1}^{A} P\left(N_{t,a} | N_{t-1,a-1}\right) \end{aligned}$$

- There are various ways of dealing with the initial state.

Estimating state-space models

- Maximum likelihood
  - The joint likelihood is integrated over (random) state variables yielding the marginal likelihood.
  - Maximize the marginal loglikelihood

$$\mathcal{L}\left(\boldsymbol{\alpha}|\boldsymbol{D}\right) = P\left(\boldsymbol{D}|\boldsymbol{\alpha}\right) = \int P\left(\boldsymbol{\theta}, \boldsymbol{D}|\boldsymbol{\alpha}\right) d\boldsymbol{\theta}$$

    with respect to parameters $\boldsymbol{\alpha}$ (e.g., K, r, fishing mortality, catchability, variance parameters)
  - Different methods can be used to maximize the log-likelihood. E.g., Laplace approximation, importance sampling.
  - Various filters (e.g., Kalman) can also be used likelihood calculations, depending on linearity and probability distributions.
  - States estimates are still Bayesian because they were random variables to begin with.

$$P(\boldsymbol{\theta}|\boldsymbol{D}, \widehat{\boldsymbol{\alpha}}) = \frac{P\left(\boldsymbol{\theta}, \boldsymbol{D}|\widehat{\boldsymbol{\alpha}}\right)}{P\left(\boldsymbol{D}|\widehat{\boldsymbol{\alpha}}\right)}$$

Estimating state-space models

- Full Bayesian
  - Specify appropriate priors for all parameters $\boldsymbol{\alpha}$

$$P\left(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{D}\right) = \frac{P\left(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{D}\right)}{P\left(\boldsymbol{D}\right)} \propto P\left(\boldsymbol{\theta}, \boldsymbol{D} | \boldsymbol{\alpha}\right) P\left(\boldsymbol{\alpha}\right)$$

  - Use your favorite Markov Chain Monte Carlo method to obtain pseudo-independent samples from the posterior distribution.
  - Pretty simple in principle, but results can be sensitive to choice of prior distributions for parameters.
  - Ensuring estimates are appropriate for a given model requires attention by the analyst.
  - Care must be taken in evaluating independence of samples from the posterior.
  - Can be more computationally intensive and take longer to obtain results.

Recent work on incorporating environmental data

State-space age-structured model

More specifics on state-space implimentation of age-structured models

- Recruitment is a stochastic outcome

$$\log N_{t,1} | \mathrm{SSB}_{t-1} \sim \mathrm{N}\left(g(\theta), \sigma^2_{N_1}\right)$$

  where expected log-recruitment might be just an average over the time series

$$g(\theta) = \log(\overline{R})$$

  or related to SSB (e.g., Beverton-Holt)

$$g(\theta) = \log\left[\frac{a\mathrm{SSB}_{t-1}}{1 + b\mathrm{SSB}_{t-1}}\right].$$

State-space age-structured model

- Log-numbers surviving is normally distributed with expectations defined by Baranov equations.

$$\log N_{t+1,a} | N_{t,a-1} \sim \mathrm{N}\left(\log N_{t,a-1} - Z_{t,a-1}, \sigma_{N_a}\right)$$

  because $\widehat{N}_{t+1,a} = N_{t,a-1} e^{-Z_{t,a-1}}$.

- For the plus group $a = A$, The expected abundance is the sum of those surviving in the plus group and those of the younger age class surviving

$$\widehat{N}_{t,A} = N_{t,a-1} e^{-Z_{t,a-1}} + N_{t,A} e^{-Z_{t,A}}$$

  so,

$$\log N_{t+1,A} | N_{t,a-1}, N_{t,A} \sim \mathrm{N}\left(\log\left(N_{t,a-1} e^{-Z_{t,a-1}} + N_{t,A} e^{-Z_{t,A}}\right), \sigma_{N_A}\right).$$

- Initial numbers-at-age $N_{1,1}, \ldots, N_{1,A}$ are estimated parameters.

State-space assessment model with environmental effects

The environmental covariate(s) can be viewed as a component of the state-space

$$x_{t+1}|x_t \sim \mathrm{N}\left(f(x_t), \sigma_x^2\right)$$

(a simple random walk in this case) and demographic parameters can be reparameterized as functions of the environment state. For recruitment we considered two types of effects,

$$g(\theta, x_{t-1}) = \log(\overline{R}) + \beta_R x_{t-1}$$

and

$$g(\theta, x_{t-1}) = \log\left[\frac{a\mathrm{SSB}_{t-1}}{1 + \left(e^{b_0 + \beta_R x_{t-1}}\right)\mathrm{SSB}_{t-1}}\right].$$

State-space assessment model with environmental effects

- Observations of the environmental covariates are normally distributed
$$y_t | X_t \sim \mathrm{N}\left(X_t, \sigma^2_{y,t}\right)$$
- aggregate log-catch observations are normally distributed

$$\log C_t | \boldsymbol{N}_t \sim \mathrm{N}\left[f(\boldsymbol{N}_t, \boldsymbol{F}_t, \boldsymbol{M}_t), \sigma^2_{C,t}\right]$$

where

$$f(\boldsymbol{N}_t, \boldsymbol{F}_t, \boldsymbol{M}_t) = \log\left[\sum_{a=1}^{A} N_{a,t} W_{a,t} \frac{F_{a,t}}{F_{a,t} + M_{a,t}}\left(1 - e^{-F_{a,t} - M_{a,t}}\right)\right]$$

State-space assessment model with environmental effects

- aggregate log-survey observations are normally distributed

$$\log I_t | \boldsymbol{N}_t \sim \mathrm{N}\left[\log\left(\sum_{a=1}^{A} Q_a N_{a,t} e^{-(F_{a,t}+M_{a,t})\delta_{I,t}}\right), \sigma^2_{I,t}\right]$$

- catch and survey age composition are logistic normal distributed

Maximum Likelihood Estimation

The parameters $\boldsymbol{\theta}$ estimated by maximum likelihood include

- variance parameters for state transitions and age composition observations,
- annual fishing mortalities and selectivities
- survey catchability and selectivities, and
- mean recruitment parameters.

We used AD Model Builder to maximize the Laplace approximation of the marginal likelihood.

- MSY-based reference points also estimated as functions of environmental covariates with uncertainty.
- Projection years also included in likelihood to estimate future state of environment and stock and propagate uncertainties.

Southern New England yellowtail flounder and Mid-Atlantic Cold Pool

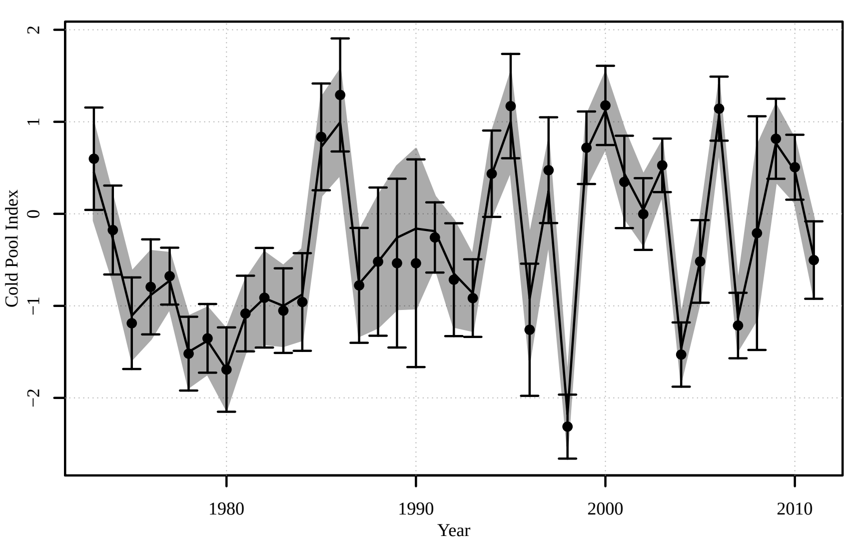
- Use the same data inputs from most recent benchmark assessment.
- Some simplifying assumptions about fishery and survey selectivity to reduce parameters.
- Mid-Atlantic Bight cold pool is thought to affect recruitment.
- Fit 4 models with the different recruitment assumptions with and without environmental effects and compared performance using AIC and Mohn's ρ.

Model Performance

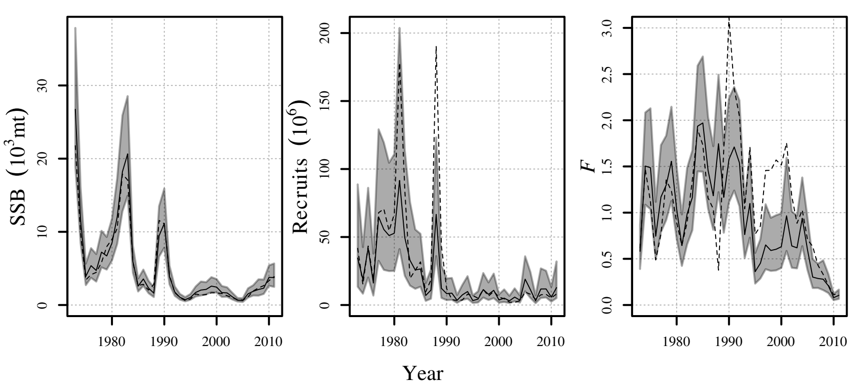
Model	$-\log(L)$	n_p	AIC	$\rho\,(SSB)$	$\rho\,(F)$	$\rho\,(R)$
R	-798.748	65	-1467.496	0.112	-0.145	0.266
R(CP)	-805.714	66	-1479.428	0.115	-0.146	0.236
R(SSB)	-804.080	66	-1476.160	0.106	-0.136	0.238
R(SSB, CP)	-809.327	67	-1484.654	0.109	-0.139	0.219

Peel	R	R(CP)	R(SSB)	R(SSB, CP)
1	16.83	4.80	8.69	0
2	15.94	4.65	7.96	0
3	15.79	4.22	8.29	0
4	14.83	3.74	7.73	0
5	12.42	2.84	6.79	0
6	13.34	4.04	5.45	0
7	15.16	6.61	3.51	0

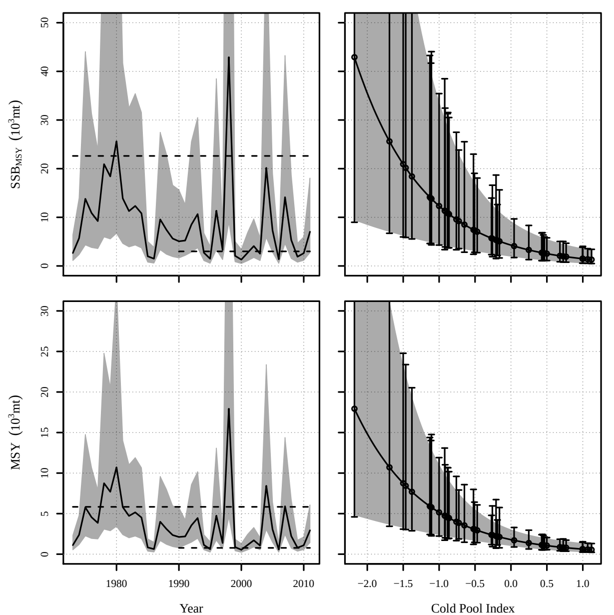
Estimated Cold Pool



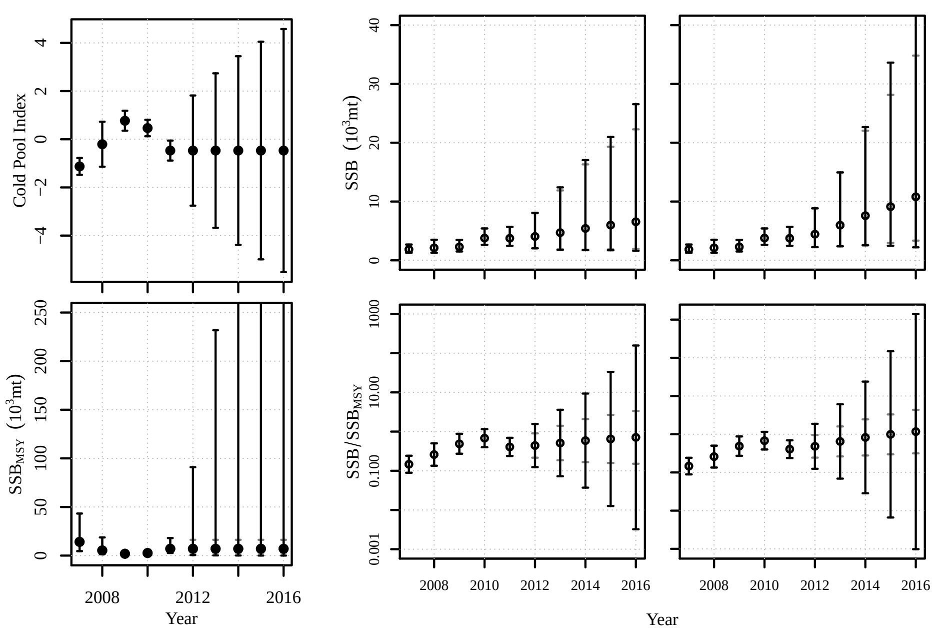
Spawning Stock Biomass, Recruitment, and F



Cold Pool effects on Biological Reference Points



Projections



Last Slide

- Ability to statistically compare models with different recruitment assumptions is an important feature of state space approach.
- Treating environmental covariate as a component of the state space
  - accounts for sampling error of environmental observations,
  - draws attention to the need to better understand error in these observations,
  - propogates uncertainty in environment into population effects, and
  - allows predictions of missing observations based on time-series model.
- The state-space approach also provides a convenient tool for error propogation in forecasting population.

References

Kalman, R. E. 1960. A new approach to linear filtering and prediction problems. Transactions of the ASME–Journal of Basic Engineering **82**(Series D): 35–45.

Mendelssohn, R. 1988. Some problems in estimating population sizes from catch-at-age data. Fishery Bulletin **86**(4): 617–630.

Meyer, R. and Millar, R. B. 1999. Bayesian stock assessment using a statespace implementation of the delay difference model. Canadian Journal of Fisheries and Aquatic Sciences **56**(1): 37–52.

Millar, R. B. and Meyer, R. 2000. Non-linear state space modelling of fisheries biomass dynamics by using Metropolis-Hastings within-Gibbs sampling. Applied Statistics **49**: 327–342.

Newman, K. B., Buckland, S. T., Lindley, S. T., Thomas, L., and Fernandez, C. 2006. Hidden process models for animal population dynamics. Ecological Applications **16**(1): 74–86.

Nielsen, A. and Berg, C. W. 2014. Estimation of time-varying selectivity in stock assessments using state–space models. Fisheries Research **158**(1): 96–101.

Sullivan, P. J., Lai, H.-L., and Gallucci, V. 1990. A catch-at-length analysis that incorporates a stochastic model of growth. Canadian Journal of Fisheries and Aquatic Sciences **47**(1): 184–198.