Extending the current coho state-space model

Ken B. Newman, May 26, 1998

1 More complex spatial frameworks

The use of a line segment to describe the spatial territory of coho salmon is an optimistic oversimplification. The Humptulips coho salmon stocks do tend to stay in the 'outside' off-shore areas, but many coho and chinook stocks reside and are caught in 'inside' marine areas such as Puget Sound and between Vancouver Island and mainland British Columbia. A more complex migration module will need to be formulated to estimate transition probabilities for moving to and from these inside areas. The initial distribution will also need to include such inside areas.

The next level of complexity for the spatial framework is to include Puget Sound and the waters between Vancouver Island and mainland BC. To describe this non-linear framework the position of an individual fish at time t is defined using a 2 dimensional random variable. The random variable is labeled $[x_t, p_t]$ where x is an label for a region the fish is in at time t and p_t is a particular location within x_t[1].

For example, suppose a fish is located at time t in the Cape Flattery region, and the natal area is somewhere in the Quillayute region. Assume that the fish is allowed to move anywhere along the coast north of its natal area or to the inside waters, but it can move inside only at the Straits of Juan de Fuca, not going around the north side of Vancouver Island. Define x_t to have one of three possible values:

- OS for outside south (from natal area to Straits of Juan de Fuca),
- ON for outside north (from Straits of Juan de Fuca to northern extent of range),
- I for inside waters.

Then p_t is defined conditional on the value of x_t:

- $0 \leq \ (p_t|x_t = OS) \ \leq B$
- $B \leq \ (p_t|x_t = ON) \ \leq UBO$
- $0 \leq \ (p_t|x_t = I) \ \leq UBI$

where B is the location of the Straits of Juan de Fuca, UBO is the upper bound on the outside 'line', and UBI is the upper bound on the inside 'line'.

[1]Formally a random variable must be a real number, so this is technically incorrect. To make this correct x could consist of numbers matched with region.

The next step is to define a probability distribution for (x_t, p_t). Let r_1, r_2, and r_3 be the probability of being in OS, ON, or I and $f_1(p_t)$, $f_2(p_t)$, and $f_3(p_t)$ be corresponding probability functions for locations within each region. So $\sum_{i=1}^{3} r_i = 1$ and each of the f_i integrate to 1 over the domains defined above.

For example, as in the current coho SSM suppose that the outside distribution followed a beta distribution if no inside movement occured. Let f_O be this outside beta distribution. Assuming the fish is located in the Cape Flattery region at time $t-1$, let r_1 be the integral of f_O over the range 0 to B; i.e., r_1 is the probability of staying south of the Straits of Juan de Fuca. The probability of not staying south is then 1-r_1. Conditional on not staying south, let p_I be the probability of moving inside. Then the three r_i can be defined as:

$$\begin{aligned}
r_1 &= \int_0^B f_O(x)\mathrm{d}x \\
r_2 &= (1-r_1)(1-p_I) \\
r_3 &= (1-r_1)p_I
\end{aligned}$$

Further assume that if a fish goes inside (x_t=I), then its distribution is a different beta distribution, denoted f_I. Now the complete probability distribution for (x_t, p_t) can be defined:

$$\begin{aligned}
\Pr(x_t = OS, p_t) &= r_1 \frac{f_O(p_t)}{\int_0^B f_{OS}(x)\mathrm{d}x} \\
\Pr(x_t = ON, p_t) &= r_2 \frac{f_O(p_t)}{1-\int_0^B f_{OS}(x)\mathrm{d}x} \\
\Pr(x_t = I, p_t) &= r_3 f_I(p_t)
\end{aligned}$$

This has been presented formally to hopefully make clear the construction, but considerable simplification can be done by some cancellations and re-expression:

$$\begin{aligned}
\Pr(x_t = OS, p_t) &= f_O(p_t) && (1) \\
\Pr(x_t = ON, p_t) &= (1-p_I)f_O(p_t) && (2) \\
\Pr(x_t = I, p_t) &= p_I f_I(p_t) && (3)
\end{aligned}$$

Given this probability distribution, putting together the movement matrix for time t, M_t, is in theory simple and will be done as with the current coho SSM. To determine the probability of moving from area a to area b, a double integration is done. The outer integral is over the length of the current area a and has function value 1/(length of a). The inner integral is the probability of being anywhere in b given the particular location in a that is specified by the outer integral. The inner integral uses one of the three components of the probability distribution (equations (1), (2), or (3)) depending on which region area b is in. As in the current SSM both r_i and f_i should depend upon previous location and time, p_{t-1} and t.

An example of determining the movement matrix components: the current area a is Cape Flattery and let b_1, b_2 and b_3 be areas in OS, ON, and I, with 'lower' and 'upper' boundaries denoted by $b_{i,L}$ and $b_{i,U}$ (i=1,2,3). Assume that f_O and f_I have parameters that depend upon the location in a and time and denote this by $f_O(p_t|a,t)$ and $f_I(p_t|a,t)$.

$$\Pr(a \to b_1) = \frac{1}{a_U - a_L} \int_{a_L}^{a_U} \left[\int_{b_{1,L}}^{b_{1,U}} f_O(p_t|a,t) \mathrm{d}p_t \right] \mathrm{d}a \tag{4}$$

$$\Pr(a \to b_2) = (1-p_I) \frac{1}{a_U - a_L} \int_{a_L}^{a_U} \left[\int_{b_{2,L}}^{b_{2,U}} f_O(p_t|a,t) \mathrm{d}p_t \right] \mathrm{d}a \tag{5}$$

$$\Pr(a \to b_3) = p_I \frac{1}{a_U - a_L} \int_{a_L}^{a_U} \left[\int_{b_{3,L}}^{b_{3,U}} f_I(p_t|a,t) \mathrm{d}p_t \right] \mathrm{d}a \tag{6}$$

2 Competing fisheries and multiple effort measures

A stock of fish is often harvested simultaneously in time and area by different types of fisheries and mixtures of gear; e.g., recreational troll fisheries and purse seines. Assume for the moment that the information on the effort expended by each type of harvester is available for the same temporal and spatial units. The standard Baranov catch equation (Ricker 1975) can be extended to include effort from competing gear types. For three gear types, for example, and ignoring the U.S.-Canada distinction, the survival and harvest rates in area k

$$\begin{aligned} S[k,k] &= \exp[-F_1 - F_2 - F_3 - N_t] \\ H_i[k,k] &= \frac{F_i}{F_1 + F_2 + F_3 + N_t} (1 - S[k,k]) \end{aligned}$$

where i=1, 2, 3 and F_i is a function of the effort by gear type i and a catchability coefficient particular to the gear.

One difficulty is that data on effort levels for different types of fisheries are sometimes recorded at different temporal resolutions, e.g., monthly versus weekly. Another difficulty is that the catch areas for recreational and commercial fisheries may fail to coincide; e.g., one recreational catch area may include portions of two different commercial fishery areas. These differences will be dealt with separately and then together. In each case I assume that there are only two competing fisheries, commercial and recreational troll, but presumably the approaches will extend to more gear types.

First cut recommendations

1. Given two fisheries with differing temporal resolution: partition the coarser data to the finer resolution, using smoothing techniques (like nonparametric density estimation).

2. Given two fisheries with differing spatial partitioning, one of which is nested in the other: partition the coarser resolution fishery data to the finer scale, using smoothing techniques.

3. Given two fisheries with differing spatial partitioning, neither of which are nested in the other: assuming one is somewhat coarser than the other, first smooth the coarser

fishery data and then partition the smoothed values according to the finer spatial resolution.

4. Given two fisheries with differing temporal resolution and non-nested spatial partitioning: first temporally partition the fishery with coarser temporal data, then spatially partition the fishery with coarser spatial data, in both cases using smoothing techniques.

3 Dealing with chinook

Chinook salmon from the same stock and cohort will mature at differing ages, 2, 3, 4, 5, and occasionally 6 years. Within a given year call an immature fish one that is not heading homeward to spawn, while a mature fish is heading home to spawn. I propose several changes to the SSM to account for this more complex life history.

- Double the length of the SSM state vector to include the immature and mature numbers in each area, stacking the immature K by 1 vector above the mature K by 1 vector.
- Insert a 2K by 2K maturation probability matrix (P) in the state equation, to the left of the survival matrix (maturation occurring instantaneously after exposure to mortality factors, but before movement). The matrix would be logically partitioned into 4 square K by K submatrices. The upper left submatrix is diagonal with elements being the probability of an immature fish staying immature. The lower left submatrix is diagonal with the complements of the upper left values. The upper right submatrix is 0 and the lower right submatrix is an identity matrix; the latter based on the assumption that mature fish stay mature.
- Make the survival and movement matrices block diagonal 2K by 2K matrices. The upper left K by K submatrices give the survival and movement probabilities for immature fish and the lower right submatrices are for the mature fish.
- Make the harvest matrix in the observation equation a K by 2K matrix, with the left half a K by K diagonal matrix with harvest rates for immature fish, and the right half is for mature fish. The assumption here is that recovered fish are not (or cannot) be distinguished by degree of maturation.

More concisely, the SSM:

$$\begin{bmatrix} \mathbf{X_{I,t}} \\ \mathbf{X_{M,t}} \end{bmatrix} = \begin{bmatrix} \mathbf{M_{I,t}} & \mathbf{0} \\ \mathbf{0} & \mathbf{M_{M,t}} \end{bmatrix} \begin{bmatrix} \mathbf{I - P_t} & 0 \\ \mathbf{P_t} & I \end{bmatrix} \begin{bmatrix} \mathbf{S_{I,t}} & \mathbf{0} \\ \mathbf{0} & \mathbf{S_{M,t}} \end{bmatrix} \begin{bmatrix} \mathbf{X_{I,t-1}} \\ \mathbf{X_{M,t-1}} \end{bmatrix} + \begin{bmatrix} \mathbf{w_{I,t-1}} \\ \mathbf{w_{M,t-1}} \end{bmatrix} \tag{7}$$

$$\begin{bmatrix} \mathbf{Y_t} \end{bmatrix} = \begin{bmatrix} \mathbf{H_{I,t}} & \mathbf{H_{M,t}} \end{bmatrix} \begin{bmatrix} \mathbf{X_{I,t}} \\ \mathbf{X_{M,t}} \end{bmatrix} + \begin{bmatrix} \mathbf{w_{I,t-1}} \\ \mathbf{w_{M,t-1}} \end{bmatrix} \tag{8}$$

The particular formulations for each component are flexible, but here are a few possibilities:

1. The harvest submatrices still depend upon fishing effort, but the catchability coefficients for $H_{I,t}$ and $H_{M,t}$ would be allowed to differ to account for differences in vulnerability or size limits.

2. The survival submatrices depend upon the previous period harvest, and thus are affected by the changes to the harvest submatrices, but they are also a function of natural mortality. The natural mortality parameter could be allowed to differ for immature and mature substocks. More importantly, however, the natural mortality parameter should be a function of time to account for the decreased natural mortality rates for aging fish. This is more critical for chinook than coho because of the many more age classes for chinook. The overwinter natural mortality must be accounted for with chinook, whereas for coho only the final age class is harvested (in general).

   One possibility is to assume a logistic model for natural mortality with time as the independent variable. The curve should of course be decreasing with time, but the intercept and slope coefficients should likely vary between age classes. For example,

   $$\text{Pr(Age i fish dies at time t)} = \frac{\exp(\beta_0^i + \beta_1^i t)}{1 + \exp(\beta_0^i + \beta_1^i t)}$$

   for ages i= 2, 3, 4, 5.

3. The movement submatrices must differ, of course, between immature and mature substocks. Different theories about movement need to be evaluated and quality of fit could be a basis for selecting one theory over another (or some variations on hypothesis testing, perhaps). For example, assume that immature fish up to age 5 continue to move north, in general, while the movement matrix for mature fish has zero probabilities for moving away from the natal area. Another theory could assume a cyclic pattern in the ocean with both mature and immature fish tending to head to the natal area late in the summer, but immature fish not moving inland.

   One of the difficult decisions will be specifying the nature of the initial distribution following each winter with no harvest recoveries.