A note on state-space models for salmon migration

Ken Newman
Division of Statistics
University of Idaho, Moscow, ID
newman@uidaho.edu

August 21, 1996

The purpose of this note is to provide a brief, minimally mathematical, overview of a general approach to:

1. modeling the abundance of salmon stocks over a set of time-area cells;

2. estimating or predicting the harvest per cell given a specified level of fishing effort.

A more technically detailed report is available at this web site.

For simplicity, the space occupied by the salmon is modeled as a line segment partitioned into k catch areas; e.g., 13 ocean catch areas running from the coast of northern California to the coast of northern British Columbia; and time is evenly partitioned, $t = 0, 1, \ldots, T$.

The data used to estimate the parameters of this model are CWT release and recovery data by time-area cell along with associated measures of effort. For each release group, there must be a matrix of time-area CWT recoveries; e.g., 16 weeks of fishing by 13 catch areas.

The remainder of this note begins with a description of a model for individual animals, then a modification for groups of animals (spatially aggregated), and then a further modification of the grouped animal model to allow for imperfect information, this last formulation being a state-space model.

1   Individual animal model

Where an animal is now, time t, is where it was at a previous time, $t-1$, plus a spatial translation, assuming that the animal is still alive. More formally, the location of an animal at time t, p_t, can be written as

$$p_t \quad = \quad p_{t-1} + s_t m_t$$

where, s_t=1 if animal is alive at time t and s_t=0 if the animal is dead. m_t is the spatial movement at time t, conditional on the animal still being alive. A more natural representation of movement may be to partition m_t into two pieces: an animal 'chooses' a direction, θ_t, and then distance, r_t.

Animal movement and mortality are thus a function of 3 components:

1. *Initial location*: p_0

2. *Survival* at time t: s_t

3. *Movement* at time t: m_t, or (θ_t, r_t)

Each component may be modeled stochastically with possibly parametrized probability distributions. For example, the initial location, p_0, could be modeled as a uniform random variable along the coast, i.e., any location on the coast is equally likely. One can postulate alternative formulations corresponding to competing theories about fish vulnerability and behavior. Given a time series of observations on an animal's location, one can estimate unknown parameters as well as carry out statistical tests for comparing competing theories.

2 Spatially aggregated animals

With most stocks of salmon, individual fish information is quite limited. At best one knows the general area and time interval where 'some' of the released fish were recaptured. I.e., recoveries are spatially (and temporally) aggregated. The individual fish model can be modified to a group of fish level by doing the appropriate spatial (and temporal) integration.

Begin by partitioning the space occupied by the population of fish into k disjoint areas. Let the number of fish present in a space-time cell $\{a, t\}$ be $n_{a,t}$. There are two cases to consider

1. the number of animals now in one area, say a, that were in another area, b, previously is known;

2. this is not known.

2.1 Previous locations known

If one assumes between animal independence, the stochastic models for each component for a group of fish turn out to be binomial and multinomial models. E.g., for the initial location component, the probability of a fish being found in a particular area is the 'sum' (integral) of the probabilities of being located at any given location in that area; the probability distribution for the initial location of the release group is then multinomial. With n_0 being the total abundance, the probability distribution for the initial number in each of the k areas is

$$\text{Multinomial}(n_0, q_1, \ldots, q_k)$$

where q_a is the probability of a fish being in area a at the beginning of the modeling period.

Likewise the survival and migration components can be shown to be binomial and multinomial random variables.

2.2 Previous locations unknown

For the initial location and survival modules, the stochastic models are no different than for the case of known prevous locations. The difficulty arises in the movement module. For example, assume zero mortality and at time period 1 there is one fish in area a and three fish in area b, then in the next time period there are two fish in area a and two in area b:

Time	Area	
	a	b
1	1	3
2	2	2

To calculate the probability of this result, all possible combinations of movements that could lead to two fish in each area must be calculated. In this situation, there are only two combinations, one from area a stays in a (thus requiring one move from b) or one from a moves to b. The enumeration problem quickly expands for when more areas, and more fish, are considered.

To avoid this problem, an approximation can be used, namely, a multivariate normal distribution. The mean of the distribution by area is simply the sum of the expected numbers moving from other areas; the variance is calculated similarly.

3 Uncertainty and state-space models

Following the aggregated fish model, a more realistic situation is one for which the numbers in each time-area cell are not known precisely, they are either estimates or indices of the actual numbers. The state-space model (SSM) framework is often used to such situations, where two time series exist, the actual numbers and the estimates, for instance. A SSM consist of two time series:

- an unobservable *state* process, n_t that is a function of past values, n_{t-1},

- a known *observation* process, c_t, that is a function of the current state value, $c_t = g(n_t)$.

A special case is the normal dynamic linear model:

$$
\begin{aligned}
n_t &= A_t n_{t-1} + v_t \\
c_t &= B_t n_t + w_t
\end{aligned}
$$

In this special case, the Kalman algorithms (Kalman 1960) can be used to both estimate parameters of the model as well as estimate values of the state process given the observable process values. The estimates of the state process can be made for various points in time relative to the observation process. If s is the length of the observation time series, and t is time period of the state to be estimated, there are three situations:

1. $s < t$: Prediction

2. $s = t$: Filtering

3. $s > t$: Smoothing

4   Remarks

The Kalman algorithms might be particularly useful for the management of exploited fish populations. Historical parameter estimates might be used to simply run the SSM forward in time and do pre-season planning. During the season, given current observations of catch, for example, current abundance and future abundance might be estimated via the filtering and prediction algorithms.

The estimates of parameters (resulting from using the Kalman algorithms to calculate the likelihood function and arrive at maximum likelihood estimates) can have considerable scientific value. Some examples:

1. Estimating the relationship between initial distribution and ocean conditions.

2. Estimating the relationship between harvest effort and harvest mortality.

3. Estimating movement rates per time period.

4. Relating movement parameters to environmental conditions.